UNIVERSITY OF RWANDA

COLLEGE OF SCIENCE AND TECHNOLOGY

AFRICAN CENTRE OF EXCELLENCE IN INTERNET OF THINGS (ACEIoT)

AN IOT-BASED TEA FERMENTATION DETECTION MODEL BASED ON DEEP LEARNING AND MAJORITY VOTING TECHNIQUES

PhD. Thesis submitted in the fulfilment of requirements of award of PhD Degree in Internet of Things – Embedded Computing Systems

Gibson Kimutai

JUNE, 2022

UNIVERSITY OF RWANDA

COLLEGE OF SCIENCE AND TECHNOLOGY

AFRICAN CENTRE OF EXCELLENCE IN INTERNET OF THINGS (ACEIoT)

AN IOT-BASED TEA FERMENTATION DETECTION MODEL BASED ON DEEP LEARNING AND MAJORITY VOTING TECHNIQUES

PhD. Thesis submitted in the fulfilment of requirements of award of PhD Degree in Internet of Things – Embedded Computing Systems

Gibson Kimutai
Regsitration number: 219008262

Thesis Supervisor: Prof. Dr. Anna Förster, PhD
Thesis Co-Supervisor: Dr. Said Rutabayiro Ngoga, PhD
Thesis Resident Co-Supervisor: Dr. Alexander Ngenzi, PhD

JUNE, 2022

DECLARATION

I KIMUTAI Gibson do hereby declare that this thesis entitled **"An IoT-based Tea Fermentation Detection Model based on Deep Learning and Majority voting Techniques",** submitted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in the Internet of Things (Embedded Computing Systems) at the University of Rwanda/College of Science and Technology is my original work and has not been submitted to any other university or higher learning institution for any other awards. I also declare that, all source of information used was acknowledged by a complete list of references and was well cited.

Signature:
KIMUTAI Gibson

Gibson Kimutai, a Ph.D. student of UR-ACEIoT student ID 219008262, successfully defended the thesis entitled "**AN IOT-BASED TEA FERMENTATION DETECTION MODEL BASED ON DEEP LEARNING AND MAJORITY VOTING TECHNIQUES",** which he prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Supervisor : Prof. Dr. Anna Förster
University of Bremen, Germany

Co-Supervisor: Dr. Said Rutabayiro Ngoga
University of Rwanda, Rwanda

Resident Co-Supervisor: Dr. Alexander Ngenzi
University of Rwanda, Rwanda

Viva Voce Members : Prof. Umaru Garba Wali
University of Rwanda, Rwanda

Dr. Michael George Masangala Zimba
Mzuzu University, Malawi

Prof. Manzoni Pietro
Universitat Politecnica Valencia (UPV), Spain

Date of Defense : 16/06/2022

To Gimson Kipkoech,

FOREWORD

Before you lies the PhD thesis "An IoT-based Tea Fermentation Detection Model based on Deep Learning and Majority voting Techniques." It has been written to fulfill the graduation requirements of the PhD in Internet of Things (Embedded Computing Systems) program at the University of Rwanda, in African Center of Excellence in Internet of Things. This thesis is a culmination of research which commenced in July 2019.

I would like to thank my supervisor, Prof. Dr. Anna Förster, for the excellent guidance and support during the process. I deliberately chose you to be my supervisor, because I knew you would provide me with challenges. This has maximized the learning opportunities, for which I am grateful. Special mention to my co-supervisors Dr. Alexander Ngenzi and Dr. Said Ngoga, who helped shed new light on many of my ideas; I am forever grateful. Additionally, I would like to express gratitude to Prof. Dr. Ambrose Kiprop and Dr. Rose Ramkat for their treasured support, which influenced my work.

My gratitude extends to the African Center of Excellence faculty in the Internet of Things for the funding opportunity to undertake studies at the College of Science and Technology, University of Rwanda. I also thank Prof. Damien Hanyurwimfura, and Dr. Omar Gatera, for their mentorship. My appreciation also goes to my friends, laboratory mates, colleagues and research team – Calorine Katushabe, George Odongo, Omar Kombo, and Farian Ishankomba for a cherished time spent in the Internet of Things laboratory.

I would sincerely want to thank the staff of the Sisibo tea factory, Kenya, for the necessary support during the data collection, tagging of tea fermentation images and deployment of the solution in the factory. My sincere appreciation goes to the University of Bremen for the Bremen International Student Internship Program (BSIP) scholarship, which enabled me to work more closely with my main supervisor: Prof. Dr. Anna Förster, at her laboratory.

Last but not least, my appreciation goes out to my family and friends for their encouragement and supports all throughout my studies.

JUNE, 2022 Gibson Kimutai

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

GDP	**:** Gross Domestic Product
ANN	**:** Artificial Neural Network
CNN	**:** Convolutional Neural Networks
IoT	**:** Internet of Things
SDGs	**:** Sustainable Development Goals
DL	**:** Deep Learning
ML	**:** Machine Learning
RF	**:** Random Forest
KNN	**:** K-Nearest Neighbor
DT	**:** Decision Tree
SVM	**:** Support Vector Machine
NB	**:** Naïve Bayes
LDA	**:** Linear Discriminant Analysis
TF	:Theaflavins
TR	:Thearubins
MV	:Machine Vision
ET	:Electronic Tongue
EN	:Electronic Nose

SUMMARY

This thesis aims to present efforts to solve tea fermentation monitoring using Machine Learning (ML) and the Internet of Things (IoT). This is a thesis by a publication containing five articles. Each of the articles presents a contribution to achieving the thesis's objectives. The design science approach guided the research process carried out in this study.

The first article analyzed the proposed computing models for monitoring tea fermentation. It was evident that most of the proposed models were in the form of feasibility studies; there was no existing dataset on tea fermentation. Furthermore, most of the work applied machine learning and vision because of the fair cost of using these technologies and their ease of implementation.

In the second article, we collected and released a tea fermentation dataset since there was none existing as reported in the first article. We describe the article and release it for use by the community. We expect that it will contribute to the progress in the field as researchers have a dataset to train, validate and test their models.

In the third article, we performed a feasibility study on the applicability of machine learners in classifying the tea fermentation dataset and LabelMe dataset, where the performance of a deep learner dubbed ”TeaNet”, a simplified version of AlexNet, was compared to the standard machine learning models: Random Forest (RF), K-Nearest Neighbor (KNN), Decision Tree (DT), Support Vector Machine (SVM), Linear Discriminant Analysis (LDA) and Naive Bayes (NB). The Alexnet model was chosen for inspiration as it has shown promising performances in classification tasks across the field and is one of the most adopted deep learners for transfer learning. Further, AlexNet is one of the earliest developed deep learners; thus, we expected it to be stable. The deep learner outperformed the other classifiers.

In the fourth article, we adopted the TeaNet model that reported promising classification results in the third article for real-time detection of tea fermentation in the Sisibo tea factory, Kenya. The majority voting techniques were adopted to improve the model's performance as it did not show good performance compared to results reported in article three. The model was

made up of raspberry pi 3 models with a Pi camera to take real-time tea images as fermentation progresses. The incorporation of the majority voting technique to aid in the decision greatly improved the model and made the model usable for the task.

During the deployment of the model, high latency intermittent internet and electricity challenges were encountered. Thus we offloaded the solution from the cloud to the edge and fog and powered the solution using a photo-voltaic energy source. Further, we applied duty-cycling where idle components were allowed to sleep, which saved 50.6559Wh during the deployment and reported the experiments in the fifth article.

Chapter 1

Background of the study

This chapter presents the background of the study where the problem area is defined together with the aims and objectives. Further, it discusses the contributions of the research and the structure of the thesis and the organization of the rest of the chapters.

1.1 Introduction

Agriculture is the main focus of Goal 2 of the Sustainable development goals (SDGs) [1], which aims to achieve ”zero hunger” . Healthy, sustainable and inclusive food systems are critical to reaching the world’s development goals [2]. Agricultural development is one of the most powerful tools to end extreme poverty, boost shared prosperity, and feed a projected 9.7 billion people by 2050 [3]. Growth in the agriculture sector is two to four times more effective in raising incomes among the poorest than other sectors [4]. Analyses in 2016 found that 65% of poor working adults made a living through agriculture [5]. Furthermore, agriculture plays an essential role in the economy of many countries and specifically developing countries [6]. In Kenya, agriculture is the backbone of the economy, and it provides income to more than 80% of her population of around 51 million people [7]. Additionally, it contributes to more than 36% of its income through its linkages with manufacturing, logistics, and other sectors. Hence it contributes to more than 65% of the total revenue to the country [8].

Tea is one of the most popular and lowest cost beverages in the world [9]. Currently, more than 3 billion cups of tea are consumed every day worldwide. Tea popularity could be attributed to its health benefits which include prevention of breast cancer [10], skin cancer [11], colon cancer [12], neurodegenerative complication [13], and prostate cancer, many others. Tea is also attributed to the prevention of diabetes and boosting metabolism [14]. Depending on the manufacturing technique,tea may be described as green, black, oolong, white, yellow and compressed tea, among other categories [15]. Black tea accounts for approximately 78% of tea produced worldwide. The top four tea producing countries are China, Sri Lanka, Kenya, and India (Table 1.1).

Table 1.1: Top tea producing countries globally

Rank	Country	Percentage
1	China	20.6%
2	Sri-Lanka	19.3%
3	Kenya	18.2%
4	India	7.5%

Kenya is the largest producer of black tea globally [14] due to its low altitude, rich loamy soil conditions, ample rainfall, and a unique climate [7]. Despite the importance of tea to the country, the sector is facing a myriad of challenges which include high production costs, mismanagement, bad agricultural practices, climate change, market competition from other countries, low prices, and lack of automation, among others [16].

There are 5 steps in the production of black tea (Figure 1.1). The process starts with the plucking of tea leaves, where two leaves and a bud is the standard practice [17]. The next step is withering, where tea leaves are spread on a withering bed for them to lose moisture. There is then the cut, tear and curl step, where tea leaves are cut and torn to open them up for oxidation[18]. The fermentation stage follows where tea reacts with oxygen to produce compounds theaflavins and thearubins which are responsible for the quality of tea. Heat is passed through tea in the drying stage to remove moisture[19]–[21]. The last step is sorting, where tea is put into various categories based on its quality. Out of these steps, fermentation is the most important in determining the quality of tea produced [22].

The fermentation process begins when cells of ruptured tea leaves react with oxygen to produce two compounds: Theaflavins (TF) and Thearubins (TR) [23], [24]. Theaflavins are respon-

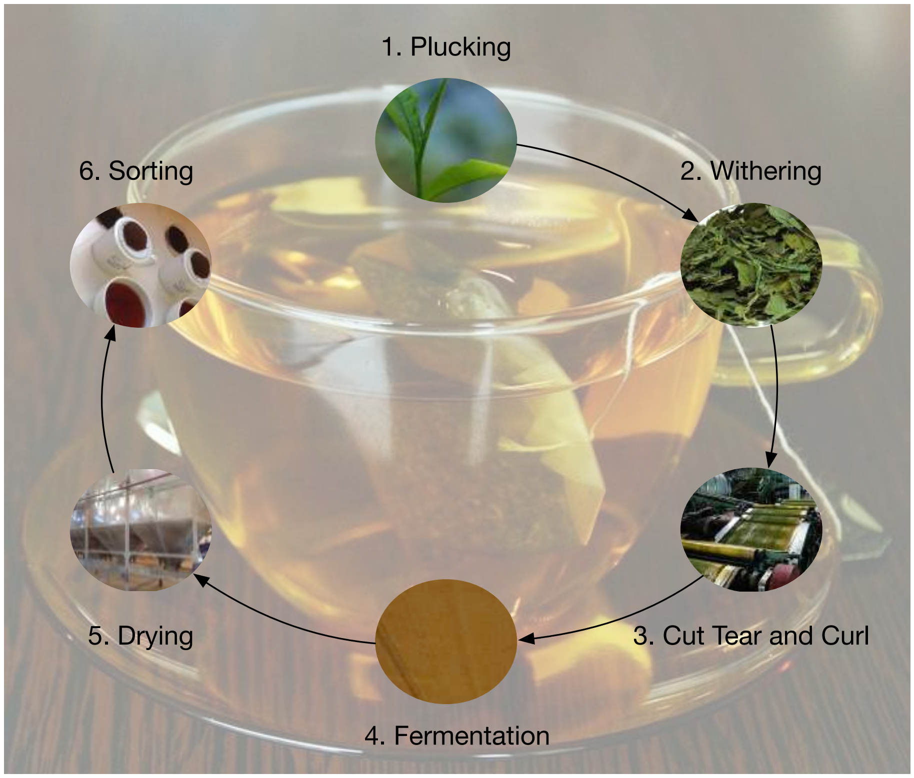


Figure 1.1: The 5 basic steps of processing black tea

sible for the brightness and briskness of the tea liquor, while TR is responsible for the colour, taste, and body of tea [24]. During fermentation, the following parameters must be maintained: temperature, and relative humidity [23]. The optimum temperature under which fermentation should take place should be approximately 25 °C. The ideal humidity should be about 42 % [25]. Fermentation is a time-bounded process (Figure 1.2); at the beginning, the liquor is raw and with a green infusion. The formation of TF and TR increases with time until optimum fermentation is achieved. At the optimum fermentation time, the liquor is mellow and with a bright infusion. This is the desired point in fermentation. After optimum fermentation time, the formation of TR reduces, and degradation of TF begins. This stage is over-fermentation, where the liquor is soft and with a dark infusion (Figure 1.2). Thus there is a need for automation which will ensure that timing is well done so as to maintain the quality of tea by avoiding overfermentation and underfermentation.

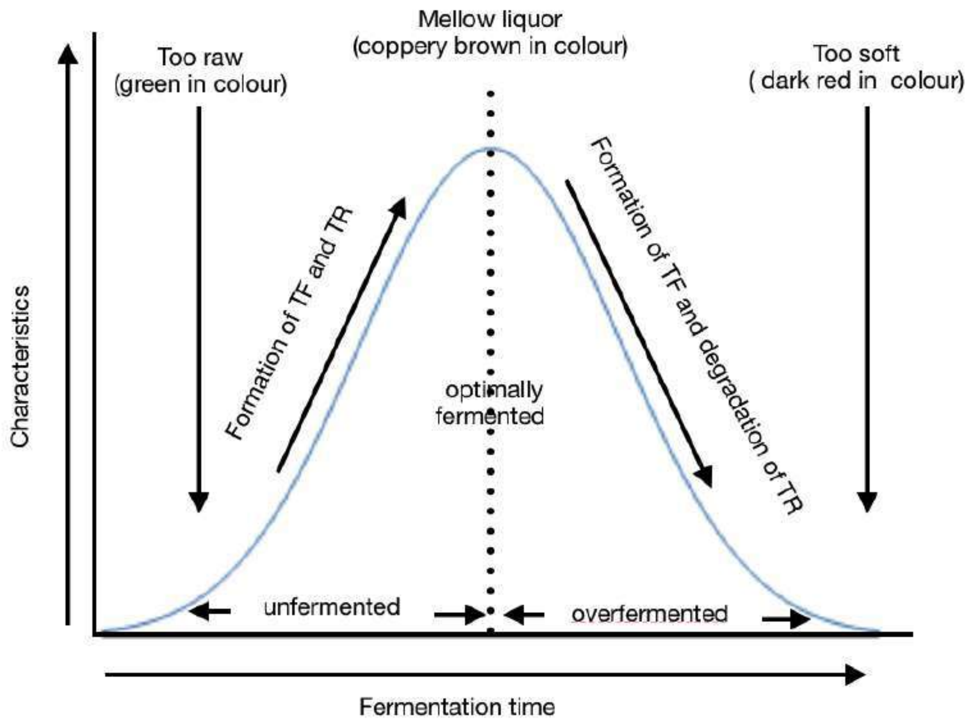


Figure 1.2: Illustration of Tea fermentation process with time. Adapted from [26]

1.2 Introduction to the Internet of Things

The internet of things (IoT) has established itself as one of the greatest smart ideas of the modern-day [27], and its effects have been seen in each feature of human ventures, with huge possibilities for smarter living [28]. A tremendous number of physical objects are now being connected to each other thus giving rise to a network of "things" that can "collaborate" by "talking" with each other and sharing data. These traditional objects gain intelligence through their underlying technologies such as ubiquitous and pervasive computing, embedded devices, communication technologies, sensor networks, Internet protocols and applications. Additionally, the ability of the IoT to work with Machine Learning is a game-changer in their applications. The IoT has shown huge potential in many fields, including agriculture, medicine, manufacturing, sports, and governance, among others [29], [30]. The IoT is also projected to have huge impacts with healthcare and manufacturing applications contributing to great economic impacts

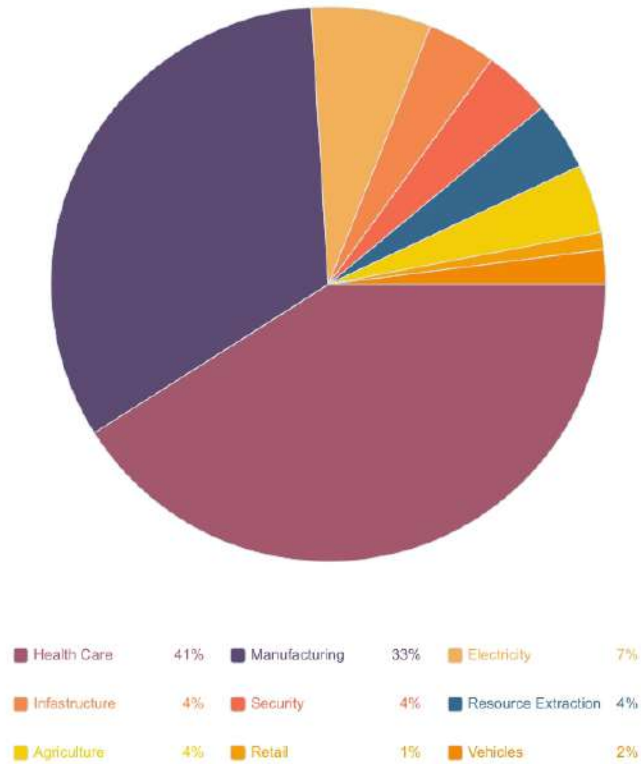


Figure 1.3: Projected Market share of IoT-based systems. Adapted from [31]

[31]. Figure 1.3 shows the projected market share of dominant IoT applications [31].

1.3 Introduction to Machine Learning

Machine learning is a branch of computer science that deals with understanding intelligence to design and develop algorithms that can learn from data, gain knowledge from experience, and improve their learning behaviour over time [32]. The challenge is to discover relevant structural and temporal patterns ("knowledge") in data, which is often hidden in arbitrarily high dimensional spaces, thus not accessible to a human. Deep Learning (DL) is a branch of Machine Learning (ML) that is currently considered as one of the enabling technologies for the industrial 4.0 revolution [33]. DL aims at simulating the mechanism of learning that is done by biological counterparts [34] by implementing neurons in humans through computational units connected with weights. The learning process in DL takes place when weights are updated based on the runs made by the model by sensing pattern in data [35]. Back-propagation technique is used in deep learning to show how a computer should alter internal parameters that are required to compute the representation in each layer from the weight in the previous layer, allowing it to find complicated structures in large data sets [36]. A typical working of DL network is depicted in Figure 1.4.

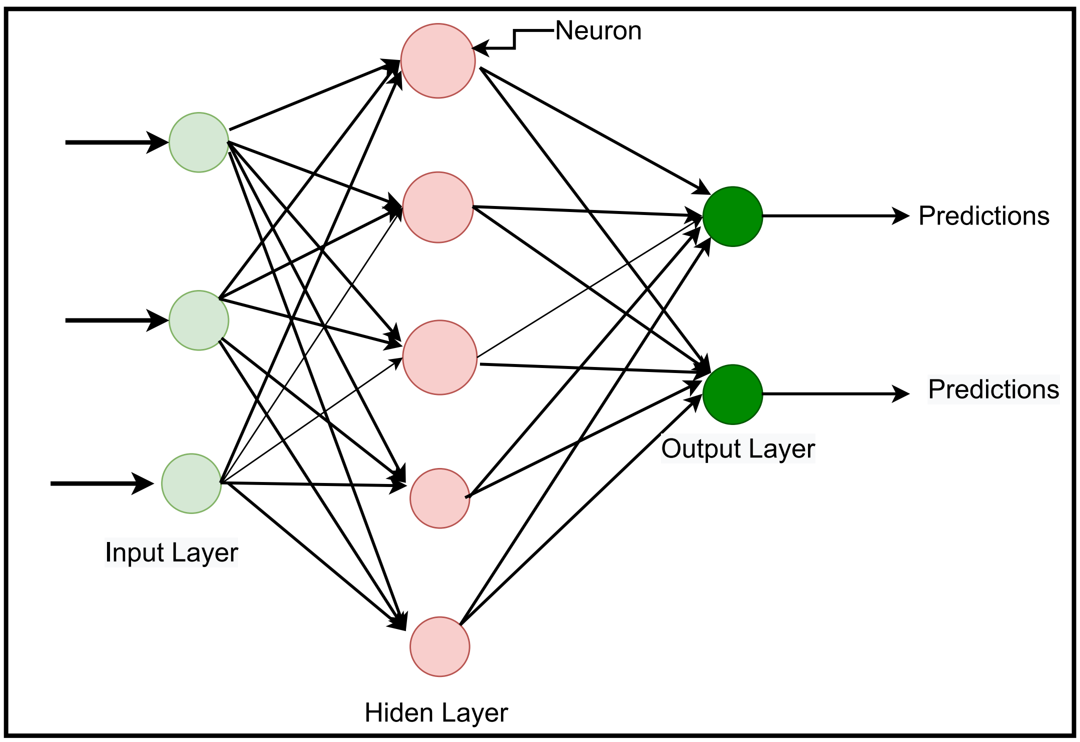


Figure 1.4: Illustration of the working of a typical DL network

The input layer receives data while convolution layers apply filters to the input for features to be effectively detected [37]. The pooling layer reduces feature sensitivity, resulting in a more generalized structure. The activation layer develops the ability to learn something complex and intriguing. The dropout layer removes specific random input values, resulting in a broad dataset that avoids over-fitting. Finally, the output layer produces results devised by the neural network [38].

1.4 Problem Statement

Tea tasters currently determine optimum fermentation manually by either of the following methods: smell peaks, colour change, infusion and tasting of tea (Figure 1.5) [39]. These methods require humans to visit tea fermentation beds to take tea samples constantly. The constant intervention of humans in a fermentation room disturbs the conditions required for high-quality fermentation, is unhygienic and not appropriate, especially in this era of Covid-19. Moreover, humans are subjective and prone to error [14]. These manual methods result in reduced quality of produced tea and translate to low tea prices. Therefore, there is a need for alternative means of monitoring tea fermentation process , which was the focus of this research.

a)Tea taster tasting tea in Sisibo tea factory, Kenya

b)Researcher tasting tea in Sisibo tea factory, Kenya

Figure 1.5: Tasting tea to determine its fermentation levels

1.5 Motivation and Aims of the Study

The livelihoods of many people depend on tea for income generation. The quality of tea determined by the level of accuracy in ending the fermentation process contributes to its market value [40]. Further, the level of accuracy in ending the fermentation process determines the quality of the produced tea [41]. Additionally, fermentation of black tea must take place within a given range of temperature, and humidity [42], [43]. As previously mentioned in Section 1.4, currently, tea tasters determine optimum fermentation by smelling tea, tasting tea and by looking at the color change. These methods are inaccurate and compromise the quality of the made tea, thus reducing their market value. Therefore, there is a need for alternative means of monitoring the process of black tea fermentation, which was the focus of this research. To solve this challenge, this study was to develop and deploy an IoT based tea fermentation monitoring model based on machine learning and image processing techniques for improved quality of black tea. Briefly, the specific objectives of this study were to:

1. Analyse selected existing methods for monitoring fermentation of black tea.

2. Develop a model for monitoring fermentation of black tea using image processing and machine learning techniques.

3. Deploy the developed energy-efficient model to monitor black tea processing for improved quality of the tea.

1.6 Contributions of the Study

This thesis reports on the application of deep learning, IoT, image processing and majority voting techniques in the detection of fermentation levels of tea. The following are a recap of some of the research contributions of the study:

1. A detailed state of the art review on methods of monitoring tea processing has been made and presented to the community.

2. The study has developed a deep learner dubbed “TeaNet”. The methodology of building the network including its structure has been explained.

3. The study deployed the developed model for real-time monitoring of tea as they undergo fermentation.

4. The study applied duty cycling or sleeping, where idle components sleep when not performing tasks. This significantly reduced the power consumption of the model.

5. This study has explored incorporating a majority voting technique to help improve the decisions of TeaNet during real-time monitoring of fermentation of black tea.

6. The study offloaded the developed model from the cloud to the Edge and Fog environments. This was to reduce the latency characterising the cloud, reduce cost and make it usable in areas not connected to the electricity grid.

7. Finally, this study suggests future research directions that researchers in the field can focus on.

1.7 Research Approach

Design science research is the design and investigation of artifacts in context [44]. It involves creation of missing knowledge, design of novel or innovative artifacts, analysis of the use and performance of artifacts and a problem solving paradigm [45]. Design science approach shall be used in this research to maintain control over the development and research results. Figure 1.6 shows a diagrammatic representation of the approach.

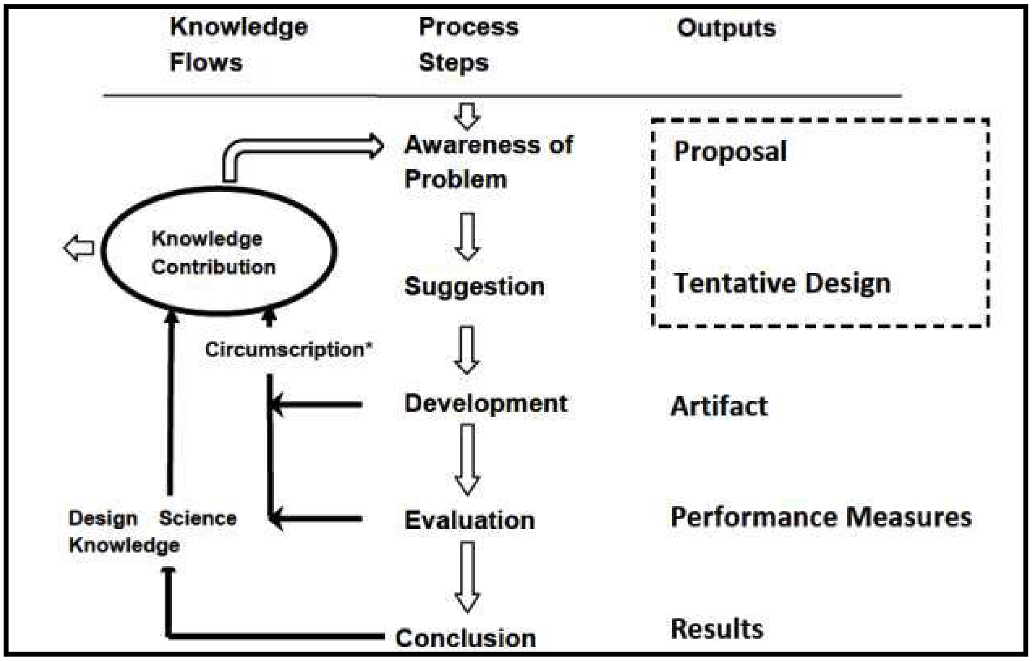


Figure 1.6: Design science approach that was followed in this study. Adapted from [46]

1.7.1 Awareness of the Problem

This stage involved analyzing existing methods of monitoring and controlling tea fermentation. This phase helped the researcher to achieve the first objective by clearly understanding the magnitude of the problem, recognition and articulation. To achieve this objective, document analysis was adopted. The literature search focused on identifying the existing methods of monitoring tea fermentation. The strengths and weaknesses of each current approach were identified to determine the knowledge gap. The main output of this phase was a review paper which is presented in Chapter 3 of this thesis.

1.7.2 Development

This stage involved developing a tea fermentation detection model based on machine learning, internet of things and image processing techniques. The model that was developed in this research was presented as a model with a focus on the functional and non-function requirements.

1.7.3 Artifact Evaluation

This involved evaluating the effectiveness of the developed model by maintaining the quality of tea produced and saving energy. The metrics that were used to evaluate the model included: precision, accuracy, f1-score, and recall [47]. The predictions of the designed model was tested and compared with the experts' decisions.

1.8 Structure and Organization of the Thesis

In an effort to achieve the objectives of the study, this thesis is divided into 12 chapters. Chapter 1 gives the background of the study. Chapter 3 presents a publication addressing the first objective of the study, which was to analyse selected existing methods for monitoring fermentation of black tea. Chapter 5 and Chapter 7 discusses publications that addresses objective 2 of the study; to develop a model for monitoring fermentation of black tea using image processing and machine learning techniques. We address Objective 3; where we deploy the developed model to monitor black tea processing for improved quality and efficient use of energy in Chapter 9 and Chapter 11 (Figure 1.7). We conclude and suggest future research directions in Chapter 12.

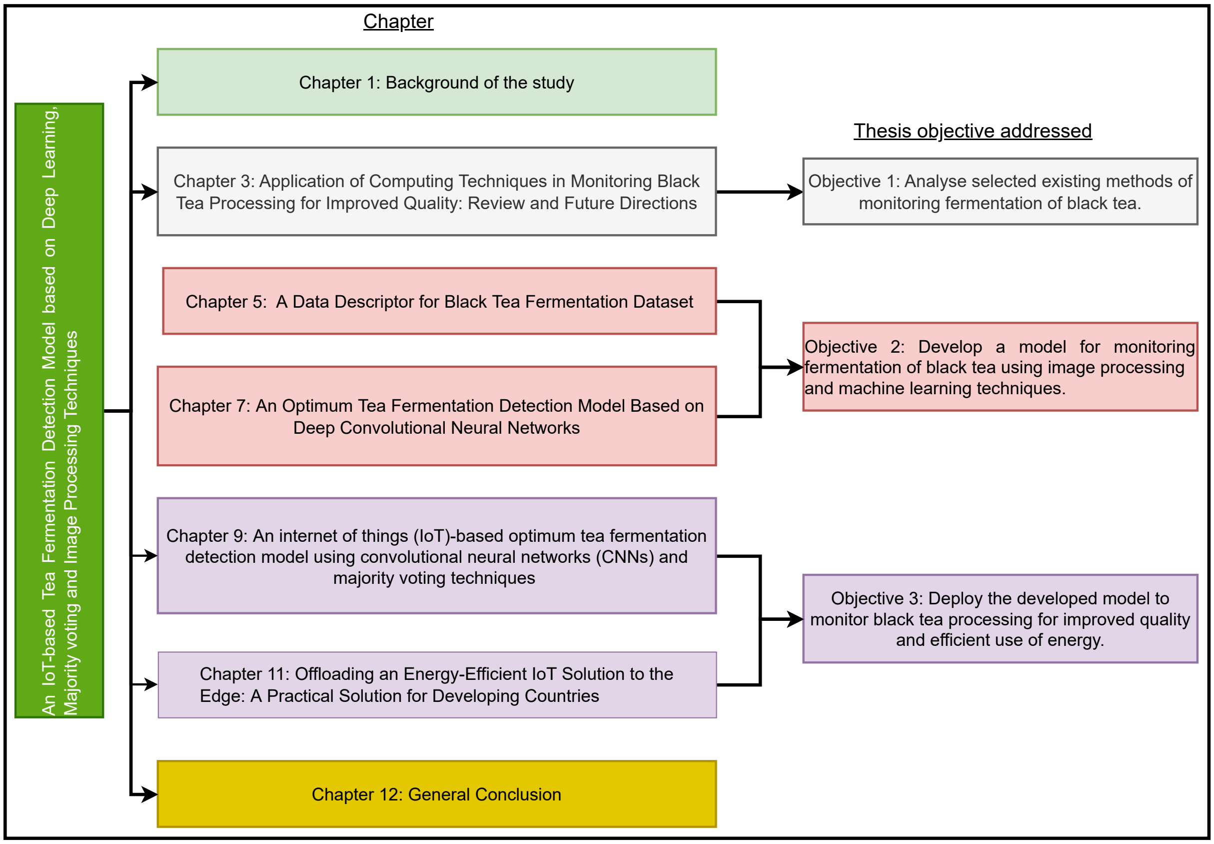


Figure 1.7: The structure and organization of the thesis

Chapter 2

Prologue to First Article

2.1 Article Details

G. Kimutai, A. Ngenzi, R. Ngoga Said, R. C. Ramkat, and A. Förster, Application of Computing Techniques in Monitoring Black Tea Processing for Improved Quality: Review and Future Directions," in 9th International Workshop on Soft Computing Applications, Virtual:Nov. 2020

Personal Contribution. The idea of writing a review paper on application of computing techniques on tea processing was my own idea. Anna Förster suggested focussing on tea fermentation with four main computing technologies: Electronic Nose, Electronic Tongue, Internet of Things and Machine Vision. Rose C. Ramkat, Alexender Ngenzi, Ngoga Said and Anna Förster proofread the paper while giving suggestions in each version. I wrote the majority of the sections of the paper, with significant contributions to the writing from Rose C. Ramkat, Alexender Ngenzi, Ngoga Said and Anna Förster. Anna Förster gave ideas on writing the Future research directions section. I produced all of the figures and tables.

2.2 Context

At the time that we wrote this article, many researchers were writing review papers on tea fermentation while focussing on individual technologies only: Electronic Nose(EN), Electronic Nose(ET), Machine Vision (MV), and Internet of Things (IoT). Some of these works include [48] which reviewed EN-based techniques for monitoring tea processing, researchers in [49] reviewed the techniques for monitoring withering of tea. Authors in [50] reviewed techniques for monitoring and grading of tea by computer vision. This motivated us to explore writing a state of the art review focusing on the following major computing technologies of monitoring tea processing: EN, ET, MV, and IoT. In this article, we present a state of the art review on computing techniques for processing tea processing. We present the original paper as it was presented in the 9th International Workshop on Soft Computing Applications, Virtual: IEEE, Nov. 2020

2.3 Contributions

The contribution of this paper is the review of state of the art literature on tea processing focusing on the following computing techniques: EN, ET, MV, and IoT. To the best of our knowledge, this is the only review paper reviewing major computing technologies for tea processing. Further, we suggest future research directions that can be taken.

2.4 Recent Developments

Following the promising classification capabilities of machine learning techniques, most of the authors presently apply machine learning simulation models in monitoring processing of tea. The ability to integrate machine learning in the IoT is providing researchers with ability to move their simulation models to actual deployments.

Chapter 3

Application of Computing Techniques in Monitoring Black Tea Processing for Improved Quality: Review and Future Directions

Tea (Camellia Sinensis) is among the most popular crops worldwide. The crop has been cultivated for more than 300 years and gives various types of produced tea based on the processing technique: green, black, yellow, white, oolong, ilex, among others [51]. Among these categories of tea, black tea is the most popular with an estimated consumption of 78% of the total tea consumption [52]. The steps of processing black tea are: plucking, withering, cutting, tearing and curling, fermentation, drying and sorting and packaging. Monitoring tea processing has been an area of interest to researchers for more than 50 years with proposals being put forward on various ways of professionalizing one or more of the steps of tea processing [53]. Some of the proposals being put forward are in the use of: Machine Vision, Internet of Things (IoT), Electronic Nose and Electronic Tongue, among others. This Chapter discusses existing methods proposed for detecting optimum fermentation of tea. Furthermore, it identifies research gaps. .

3.1 Introduction

As mentioned in Chapter 1, the tea fermentation process is the most important in determining the quality of processed tea. The optimum detection of tea fermentation is realised by smelling and tasting tea and monitoring its colour change. Some of the techniques proposed for monitoring tea fermentation include Electronic Nose, Electronic Tongue and Machine Vision (Figure 3.1).

We describe these technologies in the following paragraphs.

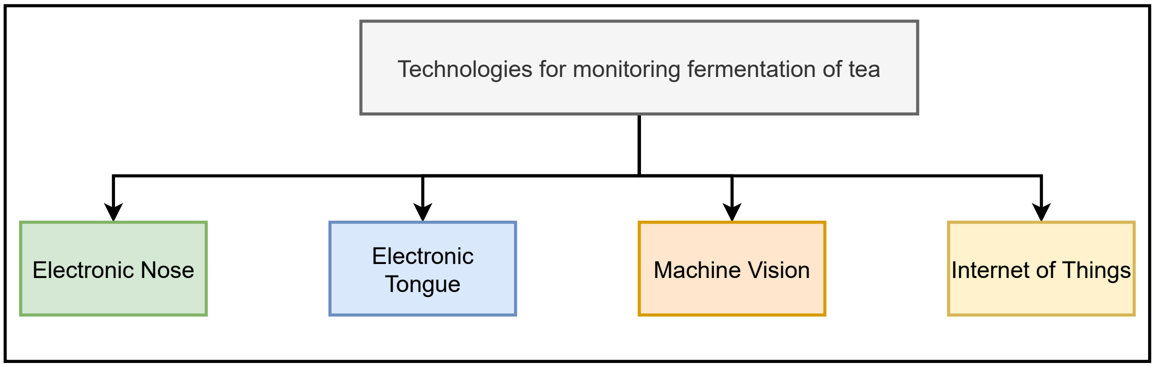


Figure 3.1: The proposed taxonomy for the techniques of monitoring tea fermentation

3.2 Electronic Nose based Models for Monitoring Tea Fermentation

An Electronic Nose (EN) is a collection of electronic chemical sensors that can recognize patterns in smell [54], [55]. The basic components of an EN are interaction system, oduor acquisition, data processing, pattern matching and the output of the result [48], [56] (Figure 3.2).

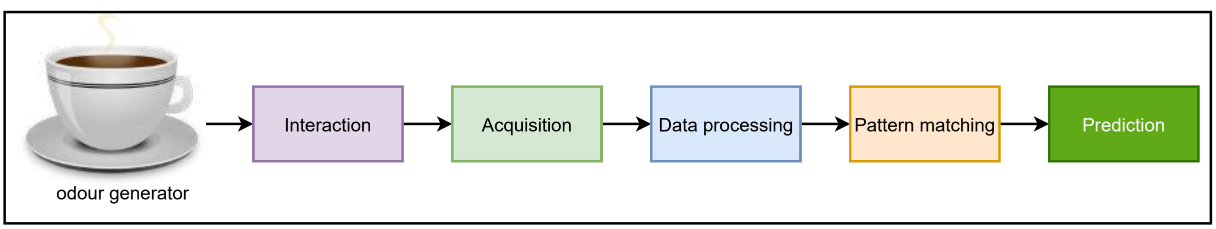


Figure 3.2: A basic workflow of E-Nose

Getting inspiration from successes registered by E-Nose in other fields, researchers in [57] developed a CNN-based prototype to determine aroma levels of tea using E-Nose sensors. Further, they monitored temperature and humidity during tea fermentation. The model predicted well the fermentation levels of tea. However, the stability of the model cannot be ascertained until its tested in real deployment. Another exciting research was presented by researchers in [58] with a proposal of monitoring tea fermentation using E-Nose where principal component analysis (PCA) was adopted for data analysis. Their prototype showed a lot of correlation with the decisions of the tea tasters, thus confirming that E-Nose can be implemented in the detection of fermentation levels of tea. Authors in [59] proposed the use of an EN-based prototype in the detection of fermentation levels of oolong tea. The authors proposed CNN for data analysis, and their approach correlated well with the expert judgements. The feasibility study presented in [58] was improved by the authors in [60] by providing it as a portable solution and improving its performance by using CNN to classify odours captured the E-Nose. The CNN model was trained using back-propagation techniques where ground thruths of collected aroma of tea was fed into the CNN network to update its weights. However, the researchers did not report the perform of the model during its evaluation. Furthermore, authors in [61] performed a feasibility study on fusing E-Nose and image processing in detecting fermentation levels of tea. They trained a CNN model using the odors collected from the E-Nose. They compared the prediction values of the CNN model with the use of Red Green and Blue(RGB) values and the decisions of the human tasters. They showed a good correlation and confirmed that E-Nose could build up a prototype for automatic tea fermentation detection. A study in [14] performed a feasibility study on the use of E-Nose and PCA in the detection of fermentation levels of tea. The model recorded an average accuracy of 75% during the evaluation which was low considering that the model was being trained and evaluated using in offline mode with tea images that had been collected and cleaned. It will be interesting to see how the model performs in real-deployment where noise in images in inevitable. Authors in [21] conducted a feasibility study on the applicability of E-Nose in the detection of optimum fermentation levels of tea, where CNN was adopted for data analysis. The model was able to detect with confidence the level of fermentation of black tea that was fed to it. However, the model seemed to have been evaluated based on 31 samples of tea which was limited. A proper conclusion of the viability of the approach will be concluded when the model is tested using more dataset. Furthermore, authors in [62] performed a feasibility study on the use of E-Nose and PCA for the optimal detection of black tea fermentation. The feasibility study showed alot of promise as the model recorded an accuracy of

100% in the classification tasks during simulation. Authors in [63] proposed the use of E-Nose in the detection of fermentation levels of tea. They adopted local discriminant analysis (LDA) to analyse the collected odours and presented their solution in a prototype.The model was able to distinguish first shaking (BS1), before the shaking group, and after the shaking group of tea and its decision correlated well with the experts' decisions. The ability of the model to detect the optimum fermentation levels of the tea was not adequately addressed in the study.

Table 3.1 presents selected work on the application of E-nose in the detection of fermentation levels of tea. There are only two studies where authors showed their proposed solution as a prototype, while the rest are feasibility studies. One of the studies fused MV and E-Nose and they both correlated well with the decisions from the experts which implied that any of the technique could be applied in the detection of fermentation levels of tea. This could be due to the large time consumption in the development of E-Nose. Development of prototype requires more resources in terms of cost compared to running simulation model. Furthermore, from the literature, more researchers prefer performing feasibility studies so as to inform themselves whether to proceed to prototype building. Generally, all the proposals showed good correlations with the decisions from the experts. Although the abilities to use EN in monitoring fermentation continue to improve, their current abilities are more minor in comparison to the abilities of the human nose [64]–[66].

Table 3.1: Selected studies on application of E-Nose in the detection of tea fermentation

Paper	**Technology readiness**	**sensors**	**data processing**	**YOP**
[57]	feasibility study	E-Nose	CNN	2007
[58]	feasibility study	E-Nose	PCA	2007
[59]	prototype	E-Nose	CNN	2007
[60]	feasibility study	E-Nose	CNN	2007
[61]	feasibility study	E-Nose+camera	traditional	2012
[14]	feasibility study	E-Nose	PCA	2018
[21]	prototype	E-Nose	CNN	2019
[62]	feasibility study	E-Nose	PCA	2019
[63]	prototype	E-Nose	LDA	2021

YOP: Year of publication   CNN: Convolution Neural Network   E-Nose: Electronic Nose
LDA: Local Discriminant Analysis   PCA: principal component analysis

3.3 Electronic Tongue Models for Monitoring Tea Fermentation

An Electronic Tongue (E-Tongue) is a sensor system for analyzing liquid based on chemical sensor arrays and a pattern recognition [67], [68]. Electronic Tongue has four basic elements: sensor acquisition hardware, feature extraction, pattern recognition and predictions (Figure 3.3).

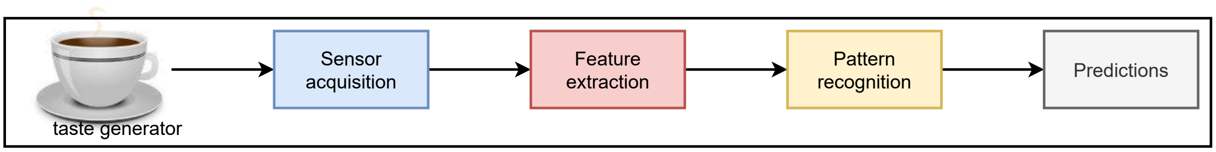


Figure 3.3: A basic workflow of E-Tongue

Some studies exist on the application of an electronic tongue to monitor tea fermentation. In [69], a handheld E-Tongue-based model prototype for detecting optimum fermentation levels of tea has been proposed. The prototype used the level of taste of tea to determine its level of fermentation. The model recorded a correlation of 0.995 with a commercially available E-tongue. However, the prototype was not implemented in real-deployment environment thus its stability has not been confirmed. Researchers in [70] performed a feasibility study on adopting SVM in classifying tea taste obtained from E-Tongue in order to detect the fermentation levels of tea. The model obtained good correlation with the experts' decision. In [71], an E-tongue to monitor theaflavins levels of tea with a an aim of detecting optimum levels of fermentation has been proposed. The PCA was adopted for data analysis and the model recorded good correlation with the expert judgement. The proposal was delivered as a feasibility study. Research in [43] proposes an electronic tongue with the use of a KNN algorithm and adaptive boosting for development of an E-Tongue for monitoring fermentation levels of tea. The model recorded good results in the classification tasks where an ensemble of KNN and CNN produced the highest accuracy of 100%. This is a promising approach and it will be interesting to establish how the model will perform in real-deployment.

In [42], authors performed a feasibility of applying CNN in the classification of taste from 5 major categories of tea in China. The approach was promising as it recorded an accuracy of 99.9% in the classification tasks. This is very encouraging and its very interesting to see how the approach perform in real-deployment. Work in [72] performed a feasibility study on the application of ET and support vector regression in classifying Cut-Tear-Curl (CTC) tea taste during fermentation of tea. The model showed promising results although it was evaluated on a limited dataset Further, a study in [73] proposed an ensemble of E-Tongue and E-Nose in the detection of optimum tea fermentation levels. They adopted PCA for data analysis and the approach recorded good classification performance and showed the advantage of using both smell and taste in the detection of tea fermentation levels. Additionally, study in [74] proposed an approach of classifying fermentation levels of tea using E-Tongue and SVM for processing of the taste patterns. The stability of the approach will be ascertained when implemented in real-deployment.

Table 3.2 shows some of the selected work on the application of E-Nose in the detection of tea fermentation. Most of the studies are in the form of feasibility studies while a study in [69] proposes their solution in the form of a prototype. However, machine learning models are consistently being adopted for data processing with a shift towards CNN models in the recent years.

Table 3.2: Selected studies on application of E-Nose in the detection of tea fermentation

Paper	**Technology readiness**	**sensors**	**data processing**	**Publication year**
[69]	prototype	E-Tongue	traditional	2010
[70]	feasibility study	E-Togue	SVM	2011
[73]	feasibility study	E-Tongue+E-Nose	PCA	2013
[71]	prototype	E-Tongue	PCA	2015
[72]	feasibility study	E-Togue	SVM	2017
[74]	feasibility study	E-Tongue	SVM	2017
[42]	feasibility study	E-Tongue	DL	2019

E-Nose: Electronic Nose E-Tongue: Electronic Nose
DL: Deep Learning PCA: Principal Component Analysis

3.4 Machine Vision based Models for Tea Fermentation

The rapid development of computer vision technology in recent years has led to an increased usage of computational image processing and recognition methods. Perhaps a pioneering work in the application of machine vision in the detection of optimum fermentation levels of tea was presented in [26] in the form of a feasibility study while adopting traditional methods for data processing. The approach involved a camera continuously taking pictures of tea and based on traditional data processing techniques, predict the classes of tea. This pioneering feasibility study was improved and presented as a prototype in [39] where a digital camera was used to capture colour images during the fermentation process. The RGB colour model was used so that histogram dissimilarity measurements matched the colour of test images to a standard image colour [26]. The approach showed good correlation with the expert judgements. Further, authors in [75] performed a feasibility study on the application of MV in the optimum detection of black tea fermentation where SVM was used for data processing. The SVM was tasked with classifying the tea fermentation images that had been captured. The average accuracy in one of the classes was 73% while the highest recorded average was 94% which were very encouraging. Research in [76] proposes a quality indexing model for black tea during fermentation using image processing techniques. The analysis of the images was done using LDA and PCA where they correlated well with the expert judgement.

Authors in [77] undertakes a feasibility study on applying SVM in the classification of tea fermentation images. The evaluation of the approach was limited but further conlusion can be made when it is applied in real-time monitoring of tea fermentation. In [78], artificial neural networks (ANN) and image processing techniques are applied to detect color changes of tea during fermentation The approach recorded good correlation with the decisions of the experts. However, it was a complex approach and computationally intensive. Additionally, a study in [79] adopted machine vision for the detection of optimum fermentation levels of tea where SVM was adopted for data processing. The approach recorded promising results on the limited dataset that was used for evaluation Recently, researchers in [80] proposed the detection of fermentation levels of tea using MV and traditional regression methods for data processing. They compared their prediction to the classification from biochemistry analysis where both approached recorded a correlation of 95%. Table 3.3 presents some of the selected studies on application of M-vision in the detection of tea fermentation. There is an accelerated acceptance of MV in the detection of optimum tea fermentation because it is cheap and easier to implement [81]. However, the same phenomena witnessed in Table 3.1 and Table 3.2 where most of the studies are in the form of feasibility studies is witnessed in Table 3.3. This could be due to the cost involved in moving a feasibility study to real-deployment, and aso it time to develop prototypes.

Table 3.3: Selected studies on application of M-vision in the detection of tea fermentation

Paper	**Technology readiness**	**sensors**	**data processing**	**publication year**
[26]	feasibility study	camera	traditional	2002
[39]	prototype	camera	traditional	2005
[81]	feasibility study	camera	traditional	2014
[75]	prototype	camera	SVM	2016
[77]	feasibility study	camera	SVM	2018
[76]	feasibility study	camera	traditional	2020
[78]	feasibility study	camera	CNN	2020
[79]	feasibility study	camera	SVM	2021
[80]	feasibility study	camera	traditional	2021

CNN: Convolution Neural Network SVM: Support Vector Machine MV: Machine vision

3.5 Internet of Things based techniques

The internet of things is one of the current major revolution methods in computing. The IoT is being applied in many fields, including agriculture [82]. The IoT connects several devices and enables them to communicate for data sharing and for coordinating activities as shown in Figure 3.4 where several heterogeneous devices have been connected.

The tea sector is attracting attention from researchers, and the authors have proposed the application of IoT to monitor temperature and humidity during the fermentation of tea. Research

Figure 3.4: The concept of the Internet of Things according to [83]

in [84] proposed a sensor network to monitor the relative humidity and temperature of tea during fermentation. Also, [85] developed an IoT-based system for monitoring the temperature and humidity of tea during processing, while [86] proposed an IoT-based system to monitor the temperature and humidity of tea during fermentation. The proposed works have been deployed in a tea factory to monitor temperature and humidity during tea processing. The models are, thus, a step in the right direction, but their scope was only on monitoring of physical conditions of tea during its fermentation. An opportunity exist to deploy the smart tea fermentation detection models using the IoT for real-time detection of tea fermentation.

3.6 Research Gaps and Summary

This Chapter has reviewed the proposals to the detection of fermentation levels of tea in the areas of: Electronic Nose, Electronic Tongue, Machine Vision and the internet of things. From the studies, it is evident that automating the detection of tea fermentation levels is an active research area as how fermentation is done directly affects the quality of the made tea and consequently its market value [87], [88]. However, most of the proposals are in the form of simulation models. The implication is that there are promising results being reported by these studies but the real test of a model should be in their operation environment. However, some of the technological proposals have deployed their models in tea factories. There are a handful proposals which are in the form of an ensemble of more than one of these techniques which is an encouraging phenomena. Perturbing, there is no reported dataset of tea fermentation images which is a limiting phenomenon as researchers have no dataset to train and evaluate their machine learning models for the detection of optimum fermentation levels of tea.

The IoT is presently gaining momentum in its application to monitoring temperature and humidity during tea fermentation. However, most of the proposals are simulated models and have not been deployed in real tea fermentation environments.

3.7 Future Research Directions

The following are some of our proposed research directions: applying image processing with machine learning techniques to decide the part of a leaf to pluck is recommended. The following machine learning classification models can be applied together with image processing: Random Forest [89], decision tree [90], Convolutional Neural Network [91], K-Nearest Neighbor [92], Support Vector Machine [22], Linear Discriminant Analysis [93] and Naive Bayes [94]. We suggest that researchers consider the following in choosing the ML model to adopt: computing resources available, performance requirements, energy consumption and time available for training them.

There is also a need to utilize renewable energy [95] in deploying the IoT devices to help in overcoming the challenge of need for high amount of energy. It was estimated that by the end of 2020 [83], there were more than 20 billion IoT devices deployed therefore the demand for energy will is very high. There is a need to also deploy underground sensor networks [96] in tea plantation to monitor physical conditions for improved tea yields. The sensor networks would ensure that tea grows under optimum conditions thus ensuring that the quality of tea produced is high.

From this chapter,there is no existing dataset on tea fermentation images. This is a serious gap as there will be a challenge of lack of data to train and evaluate machine learning models. Consequently, it is recommended that in the future, efforts should be made to release more open source image datasets in tea processing.

Finally, there is a need to explore on deploying more deep learning techniques [117] in monitoring tea processing. This will make many Convolutional Neural Network (CNN) networks available and will make the application of transfer learning [117] in monitoring tea processing a viable approach.

Chapter 4

Prologue to Second Article

4.1 Article Details

G. Kimutai, A. Ngenzi, R. Ngoga Said, R. C. Ramkat, and A. Förster, “A Data Descriptor for Black Tea Fermentation Dataset,” Data, vol. 6, no. 3, p. 34, Mar. 2021, issn: 2306-5729. doi: 10.3390/data6030034.

Personal Contribution. The idea of writing a data descriptor arose out of a meeting with Anna Förster. Formal analysis was done by A. Förster; I did methodology together with A. Förster; I implemented the software for data collection with supervision, from A. Ngenzi, R. Ngoga Said, R. C. Ramkat, and A. Förster. I did the validation and visualization of the results. I also wrote the original draft while A.Ngenzi, R. Ngoga Said, R. C. Ramkat, and A. Förster performed writing review and editing. I produced all of the figures and tables. All authors read and agreed to the published version of the manuscript.

4.2 Context

At the time that we wrote this article, there was no existing data descriptor on black tea fermentation images. This motivated us to explore collecting a tea fermentation dataset and releasing it to the community. The data descriptor underwent a peer review, published and is available in [97]

4.3 Contributions

The contribution of this paper is to release and a describe a black tea fermentation dataset. To the best of our knowledge, this is the only existing black tea fermentation dataset thus we expect to witness researchers using the dataset to train, validate and evaluate their machine learners.

4.4 Recent Developments

After publishing the data descriptor, we have also collected a tea sickness dataset from a tea farm in Koiwa location, Bomet county, Kenya. The dataset contains the following classes of tea diseases: red leaf spot, algal leaf spot, bird's eye spot, gray blight, White spot, anthracnose, and brown blight. Further, it also has a healthy tea leaves to act as a baseline. The dataset can be found in [98]. This will add richness to the tea sickness dataset in [99] which contains the following classes: tea leaf blight, tea red leaf spot and tea red scab.

Chapter 5

Data Descriptor for a Black Tea Fermentation Dataset

Tea is currently the most popular beverage after water. Tea contributes to the livelihood of more than 10 million people globally [100]. There are several categories of tea, but black tea is the most popular, accounting for about 78% of total tea consumption [101]. Processing of black tea involves the following steps: plucking, withering, crushing, tearing and curling, fermentation, drying, sorting, and packaging [102]. Fermentation is the most important step in determining the final quality of the processed tea [103]. Fermentation is a time-bounded process and it must take place under certain temperature and humidity conditions [104]. During fermentation, tea color changes from green to coppery brown to signify the attainment of optimum fermentation levels [105]. These color changes are currently manually monitored. At present, there is no existing dataset on tea fermentation images which is a limiting phenomena as datasets are required for training machine learners. Thus this study provides a tea fermentation dataset available, composed of tea fermentation conditions and tea fermentation images.

5.1 Background and Rationale

Tea is currently among the most popularly consumed beverage across the world [11] and is responsible for the economic growth of many countries, including India, Sri-Lanka, Kenya, China among other countries [15]. These top tea-producing countries produce several varieties of tea, which include: yellow tea, illex tea, oolong tea, black tea, white tea, among others [106]. Among these categories of tea, black tea is the most consumed, accounting for approximately 78% of the total daily consumption of tea [107]. Kenya is the leading exporter of black tea worldwide, with her major tea-producing counties being Kericho, Bomet, Nandi, and Nyeri [16]. The crop is a source of livelihood for more than 10 million of the total countries' estimated population of 47 million people [108]. Although the crop is still the leading exchange earner for the county, the sector is ailing due to ever-reducing tea prices [109]. This is attributed to increased competition from other countries, poor management, and the low quality of tea produced, among others [40]. The steps of processing black tea are plucking, withering, crushing, tearing and curling, fermentation, drying, sorting, and packaging [21]. Among these processes, fermentation is the key determinant of the final quality of the processed tea [49]. The process should take place at 22-26 $^{\circ}$ C for 60-90 minutes depending on the clone of the tea [78], [110]. During fermentation processes, tea changes color from green to coppery brown and finally to dark red [84]. The optimally fermented tea is coppery brown in color, and has a fruity smell and a sweet taste, while the underfermented tea is green in color, and has a grassy smell and a strong taste [111]. The overfermented tea is dark red and is characterized by a bitter taste [24]. 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 [112]. 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 [113], education, E-Commerce [17] tourism and banking, among others [114]. 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 [115], 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 chapter is arranged as follows. Section 5.2 presents the materials and methods while section 5.3 provides the description of the dataset. We do provide data validation in section and conclude the chapter in section 5.5.

5.2 Materials and Methods

This section describes the resources and the approach followed in the collection of the dataset. Section 5.2.1 discusses on the resources while the collection of the dataset is discussed in Section 5.2.2.

5.2.1 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 [116]. The Raspbian operating system for raspberry pi [117] 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) [118] 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 [119] was used in writing programs to capture the images and using the Pi camera. The block diagram of the system for capturing the dataset is shown in Figure 5.1.

5.2.2 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 as shown in (Figure 5.2). After collection of each image, the tea experts provided ground truths on their correct classes.

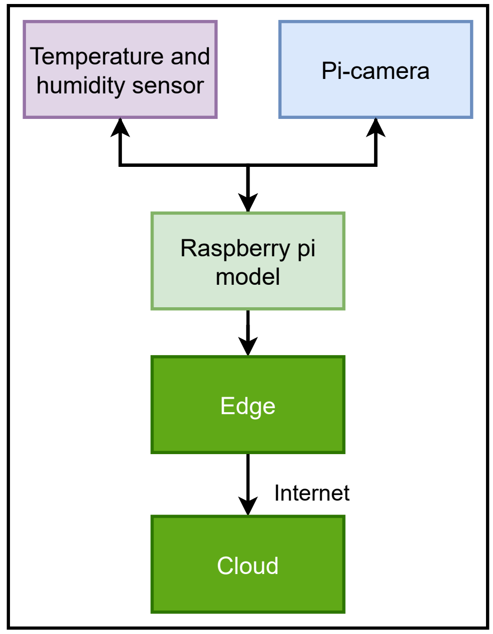


Figure 5.1: Block diagram of the data collection system

5.3 Data Description

This study releases a black tea fermentation dataset [120] 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 file structure of the dataset is shown in Figure 5.3.

Black tea fermentation conditions are contained in a comma-separated values (CSV) file and contain the fermentation time, temperature, humidity, reference to image and the category of tea (Figure 5.4).

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

Figure 5.2: Collection of the dataset in Sisibo tea factory, Kenya using Raspberry pi and Pi camera.

underfermented, fermented, and overfermented (Figure 5.5).

The underfermented tea is in a folder labeled “underfermented tea images” in the “Categorized black tea fermentation images” folder. Tea in this category is partially fermented. 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 5.5 (a). The optimally fermented tea is correctly fermented and is usually coppery brown. The fermented tea is in a folder labeled “optimally fermented tea images” in the images folder. The images are labeled from “fermented _00” to “fermented _1999”. A sample of the images is presented in Figure 5.5 (b). The fermentation level of overfermented tea is beyond the optimum and is dark red. Overfermented tea is in a folder labeled “overfermeted tea images”. The images are labeled from “Overfermented _00” to “Overfermented _1999”. A sample of the images is presented in Figure 5.5 (c).

5.4 Data Validation

The fermentation of tea must take place within a given range of temperature and humidity [121]. These ranges vary in different countries but, in Kenya, the acceptable range is between 18 and 30 $^{\circ}$ celcius while the acceptable humidity values are 35 % and 95%. Figure 5.6 shows temperature and humidity values during a fermentation cycle in Sisibo Tea Factory on 10 August

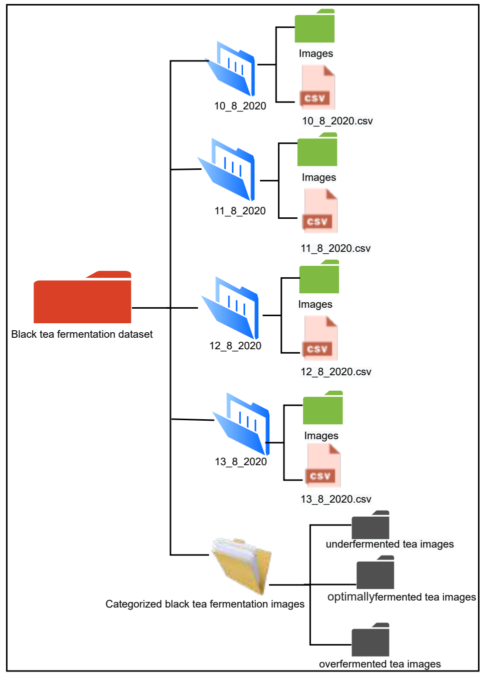


Figure 5.3: File structure of the black tea fermentation dataset.

	A	B	C	D	E
1	Time	Temperature (°C)	Humidity (%)	Reference to image	Category of tea
2	12:31:13	24	61.6	underfermented_1	underfermented
3	12:32:12	23.9	62.1	underfermented_2	underfermented
4	12:35:58	23.7	62.8	underfermented_3	underfermented
5	12:36:52	23.6	64	underfermented_4	underfermented
6	12:37:55	23.4	63.9	underfermented_5	underfermented

Figure 5.4: File structure of the black tea fermentation dataset.

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 fermentation curve was smooth and it took 68 min before the tea was fully fermented.

Figure 5.7 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

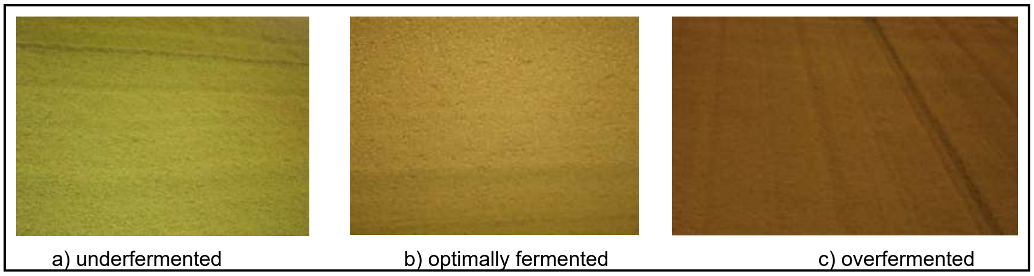
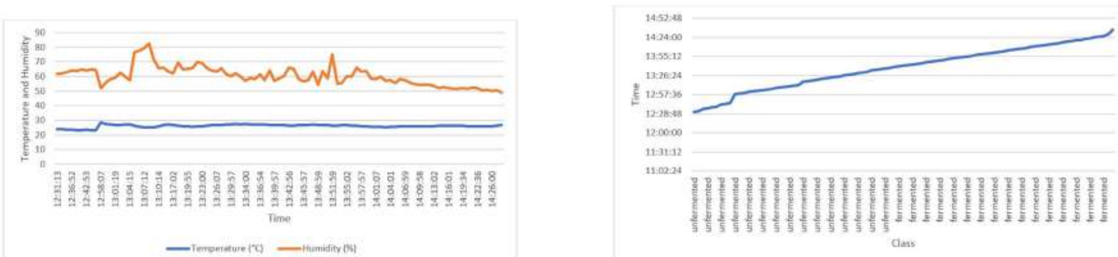


Figure 5.5: Classes of tea fermentation dataset a)underfermented b)optimally fermented c)overfermented.

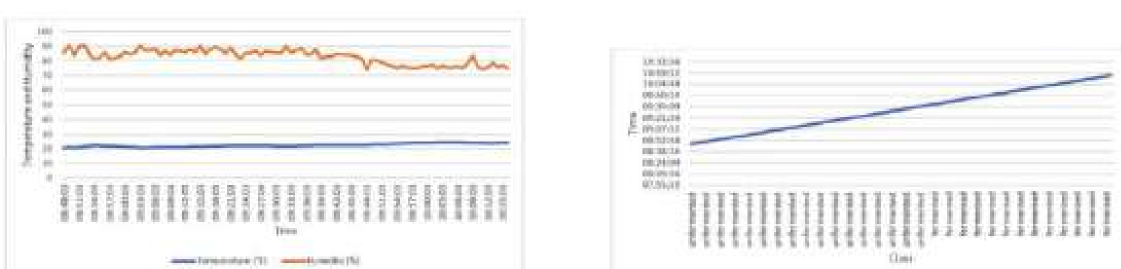


(a) Temperature and humidity values

(b) Fermentation classes of tea with time

Figure 5.6: Fermentation conditions of tea in Sisibo tea factory on 10 August 2020.

the area. The highest humidity value recorded was at 90.8% while the least humidity was 38 ° celcius which was within the accepted humidity values. 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.

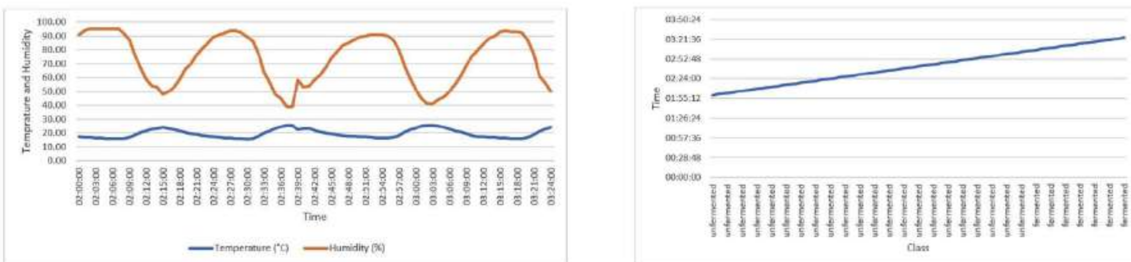


(a) Temperature and humidity values

(b) Fermentation classes of tea with time

Figure 5.7: Fermentation conditions of tea in Sisibo tea factory on 10 August 2020.

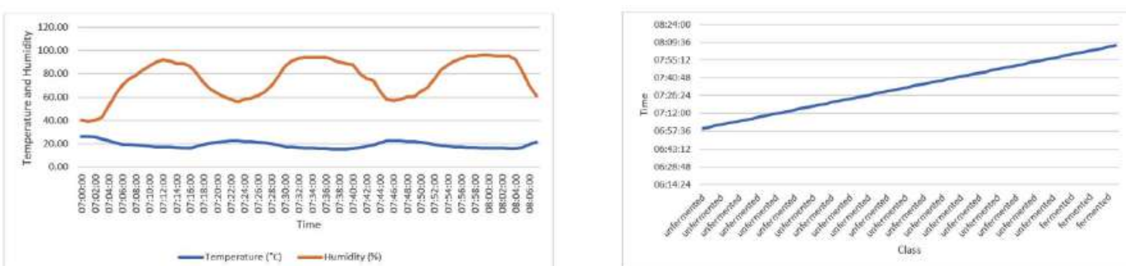
Figure 5.8 shows temperature and humidity values during a fermentation cycle in Sisibo tea factory on 12 August 2020. The fermentation process started at 02:00:13 h. The temperature recorded ranged from 15 °C to 25.53 °C. Some of the temperatures values fell below the optimum, unlike on 10 August 2020. This is due to the fermentation occurring in the early morning when the region is naturally colder than during midday hours, however, these values increased steadily with time. The humidity values recorded ranged from 37% to 95%. The fermentation curve was smooth and it took 64 min before the tea was fully fermented.



(a) Temperature and humidity values (b) Fermentation classes of tea with time

Figure 5.8: Fermentation conditions of tea in Sisibo tea factory on 12 August 2020.

The average temperature and humidity values recorded on 13 August 2020 is presented in Figure 5.9. The temperature values ranged from 17 ° c to 23 °c. The humidity values ranged from 40% to 98% which was beyond the range of 95%.



(a) Temperature and humidity values (b) Fermentation classes of tea with time

Figure 5.9: Fermentation conditions of tea in Sisibo tea factory on 13 August 2020.

5.5 Conclusion

In this chapter, a tea fermentation dataset has been described. 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 this is the only existing dataset for training these models.

Chapter 6

Prologue to Third Article

6.1 Article Details

Kimutai G, Ngenzi A, Said RN, Kiprop A, Förster A. An Optimum Tea Fermentation Detection Model Based on Deep Convolutional Neural Networks. Data. 2020; 5(2):44.

Personal Contribution. The idea of writing this feasibility study arose out of a meeting with Anna Förster,A. Ngenzi and R. Ngoga Said. Formal analysis was done by A. Förster; I did methodology together with A. Förster; I implemented the software for data collection with supervision, from A. Ngenzi, R. Ngoga Said, Kiprop A, and A. Förster. I did the validation and visualization of the results. I also wrote the original draft while A. Ngenzi, R. Ngoga Said, Kiprop A, and A. Förster performed writing review and editing. I produced all of the figures and tables. All authors read and agreed to the published version of the manuscript.

6.2 Context

At the time that we wrote this article, machine learning was showing a lot of promise in various domains. However, from the first article, we realised that the same success had not be realised in the detection of optimum tea fermentation. This motivated us to explore machine learning for the detection of optimal tea fermentation. The work has been published and is available in [107].

6.3 Contributions

The contribution of this paper was the development of a deep learner for the detection of optimum fermentation levels of tea. The methodology of its development was given and its available for adoption in transfer learning for other comparable tasks.

6.4 Recent Developments

Since the development of this model, [122] developed a deep learner for recognising chinese tea. Authors in [123] adopted machine learning to classification odor from Electronic Nose(EN) for the detection of optimal fermentation levels of tea. More recently, authors in [124] performed a feasibility study on the use of deep learning in identifying tea shoots for plucking.

Chapter 7

An Optimum Tea Fermentation Detection Model Based on Deep Convolutional Neural Networks

Tea is one of the most popular beverages in the world and its processing involves a number of steps which includes fermentation. Tea fermentation is the most important step in determining the quality of tea. Currently, optimum fermentation of tea is detected by tasters using any of the following methods: monitoring change in color of the tea, tasting and smelling the tea as fermentation progresses. These manual methods are not accurate. They lead to a compromise in the quality of tea. This study proposes a deep learning model dubbed TeaNet based on Convolution Neural Networks (CNN). The input data to TeaNet are images from fermentation and LabelMe datasets. We compared the performance of TeaNet with other standard machine learning techniques: Random Forest (RF), K-Nearest Neighbor (KNN), Decision Tree (DT), Support Vector Machine (SVM), Linear Discriminant Analysis (LDA) and Naive Bayes (NB). TeaNet was more superior in the classification tasks compared to the other machine learning techniques. However, we will confirm the stability of TeaNet in the classification tasks in our future studies when we deploy it in a tea factory in Kenya.

7.1 Introduction

As mentioned in Chapter 1, tea is one of the most popular and lowest cost beverages in the world [9]. Currently, more than 3 billion cups of tea are consumed every day worldwide. This popularity is attributed to its health benefits which include prevention of breast cancer [10], skin cancer [11], colon cancer [12], neurodegenerative complication [13], prostate cancer, many others. Tea is also attributed to the prevention of diabetes and boosting metabolism [14]. Depending on the manufacturing technique, it may be described as green, black, oolong, white, yellow and compressed tea [125]. Black tea accounts for approximately 70% of tea produced worldwide. The top four tea producing countries are China, Sri Lanka, Kenya, and India.

Perturbing, tea tasters determine optimum fermentation manually by either of the following methods: smell peaks, color change, infusion and tasting of tea. The constant intervention of humans in a fermentation room disturbs the environment created for fermentation and is also unhygienic. Moreover, humans are subjective and prone to error [88]. These manual methods lead to a compromise in the quality of produced tea and translate to low prices of tea. Therefore, there is a need for alternative means of monitoring the process of fermentation which is the focus of this research.

7.1.1 Introduction to Machine Learning

Currently, machine learning has been applied to many different fields: engineering, science, education, medicine, business, accounting, finance, marketing, economics, stock market, and law, among others [27], [126]–[129]. Machine learning is a branch of artificial intelligence (AI) that enables a system to learn from concepts and knowledge [36]. Deep learning is a collection of machine learning algorithms which models high-level abstractions in data with non-linear transformations [130]. Deep learning works with the principle of the Artificial Neural Networks (ANN) system and its fundamental computation unit is a neuron [91], [127], [130]. In ML, feature extraction and classification are in different steps, while in deep learning they are in a single step and are done concurrently.

The contribution of chapter paper is twofold: First, this research proposes a deep learning model based on CNN for monitoring black tea during fermentation. Secondly, this research releases a tea fermentation dataset [115]. The rest of the Chapter is arranged as follows: presentation of some of the studies aimed at digitizing fermentation is done in Section 7.2 and a discussion of materials and methods used in this research is presented in Section 7.3. We discuss the implementation of the algorithms in Section 7.4. Section 7.6 provides the results and their discussion and the Chapter is concluded in Section 7.7.

7.2 Related Work

With advancements in computing, digitization across many fields is being witnessed [131]. In agriculture, tea processing has been receiving attention from researchers. Research in [76] proposes a quality indexing model for black tea during fermentation using image processing techniques.They adopted component analysis (PCA) and linear discriminant analysis (LDA) for data processing and they provided the proposal as a feasibility study. Research in [77] proposed a model that could detect fermentation levels of tea based by classification fermentation images. In [78], artificial neural networks (ANN) and image processing techniques are applied to detect color changes of tea during fermentation. Research in [132] applied SVM with image processing to detect optimum tea fermentation. In [133], the authors used image processing to detect the color change of tea during fermentation where is showed good correlation with the experts' decisions. Authors in [71] implemented an ensemble of electronic tongue with machine vision to predict the optimum fermentation of black tea. They applied CNN model to predict the classes based on images and also taste of tea. From the literature, tea fermentation is an active research area with authors suggesting different approaches. However, the tea fermentation dataset has not to be used. The use of image processing is the most viable approach due to the low cost of imaging devices. Additionally, a color change is easy to detect compared to taste and odor.

7.3 Materials and Methods

After acquiring data, the next phase was data preprocessing where activities discussed in section 7.3.2 were done. The cleaned data was fed to the ML classifiers for training.The ML classifiers adopted were: decision tree, random forest, k-nearest neighbor, TeaNet, support vector machine, linear discriminant analysis, and Naive Bayes. The training involved hyperparameter tuning until when the models were fully trained. Some of the hyperparameters were the learning rate, number of the epoch, regularization coefficient and batch size. Some of the optimization strategies available included grid search, random search, hill-climbing, and bayesian optimization, among others [134]. In this study, we adopted the grid search and random search methods. The models were then validated and evaluated using the data discussed in section 7.3.1. In Model validation, models which did not pass the validation tests were taken back to training phase. The evaluation results are presented in Section 7.6. The models can be deployed to a tea fermentation environment after the aforementioned steps.

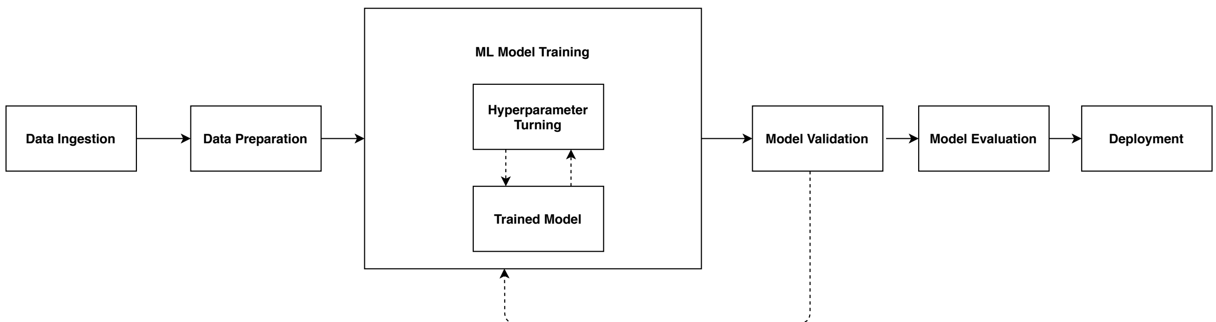


Figure 7.1: Implementation of machine learning techniques

7.3.1 Datasets

In this chapter, two datasets were used: tea fermentation and LabelMe datasets. Since there was no existing standard dataset on tea fermentation images existing in the community, we used LabelMe dataset to validate our results for it is widely used by researchers in image classification to report their results, the dataset is available at no cost, and there was no available dataset on images of tea fermentation images. Further, two distinct datasets were used so as to produce a general model that could solve be adopted by transfer learning for use across domains. We discuss each of the datasets in the following paragraphs.

The images in the tea fermentation dataset [97] were taken in a black tea fermentation environment in a tea factory in Kenya as we have discussed in Chapter 5. The tea fermentation dataset contained 6000 images that were captured during the fermentation of black tea. Figure

7.2 shows an image of each of the classes of the tea fermentation dataset.The classes of the images in this dataset are: underfermented, fermented and overfermented. The underfermented tea is green in color while the fermented tea is coppery brown. The overfemented tea is dark red in color. Ideally, fermentation should be ended when tea is coppery brown in color.

Figure 7.2: Examples of classes of the tea fermentation dataset

Table 7.1 shows the number of images for every class that was used as training, validation, and testing datasets for the classification algorithms. The 80/20 ratio of training/test data is the most commonly most commonly used ratio in neural network applications and was adopted in this research. Besides, 10% subset of the test dataset was used to validate the results. A total of 4,800 images distributed equally to the 3 classes of images were used for training of the models. To perform validation, 40 images were used in each of the classes while 360 images were used to test the model in each of the 3 classes.

Table 7.1: The image dataset comprising of three classes of images of tea fermentation

Class	**training**	**validation**	**testing**
Underfermented	1600	40	360
Fermented	1600	40	360
Overfermented	1600	40	360
Total	4800	120	1080

The other dataset that we adopted in this study is LabelMe dataset [135] The dataset is one of the standard datasets which researchers in the field of image classification use to report their results. The dataset contains 2688 images from 3 classes of outdoor scenes. The classes are forest, coast, and highway. Examples of image from each of the classes is shown in Figure 7.3.

Figure 7.3: Examples of categories of LabelMe dataset. Source [135]

Table 7.2 shows the number of images used for training, validation, and testing in each of the categories. As with the case in LabelMe dataset, we adopted the 80/20 ratio for training and testing and 10% for validation.

Table 7.2: Number of images used for training, validation and testing in the LabelMe dataset

Class	training	validation	testing
Coast	717	18	161
Forest	717	18	161
Highway	717	18	161
Total	2151	54	483

7.3.2 Data Preprocessing and Augmentation

After collecting the images they were resized to 150×150. Resizing images to 150×150 before inputting them into different networks was done to adapt different pre-training CNN structures. We adopted semantic segmentation annotation method discussed in [136] to annotate the images. There are numerous types of noise in images but the most common are photon noise, readout noise, and dark noise [137], [138]. To perform denoising, we adopted the linear filtering method.

The models were evaluated by the following metrics: Precision, Recall, F1-Score, and Confusion matrix. These metrics have been discussed in Section 7.5. We briefly describe the performances in the following paragraphs.

7.3.3 Feature Extraction

Feature extraction in image processing is the process of extracting image features. It is the most crucial step in image classification as it directly affects perfomance of the classifiers [139]. There are various techniques of feature extraction but in this Chapter, we adopted color histogram for color feature extraction and Local Binary Patterns (LBP) algorithm for texture extraction.

7.3.3.1 Color Feature Extraction

Color is an important feature descriptor of an image. During tea fermentation, the color change is evident as the process continues. Relative color histograms in different color spaces can be used to describe tea fermentation images. There are several color spaces which include Red-Green Blue (RGB), Hue Saturation Value (HSV), Hue Saturation, Brightness (HSB), and among others[140]–[142]. RGB color space represents a mixture of Red, Green, and Blue. This is the color space that was used to represent the images in this paper. We used color histogram [143] to extract color features of the images that are then fed to the classifiers for training, evaluation, and testing. To construct a feature vector from the color histogram, we used OpenCV [144]. The input was an image of RGB color space. The RGB color space was converted to HSV so as to separate image luminance from color information and represented by 3 channels (the Hue, the Saturation and the Value). We used 8 bins to represent the three channels. Finally, the range of the channels was between 0-150 since the images had been resized to 150 by 150 pixels. Figure 7.4 (a) shows an image of underfermented tea while Figure 7.4 (b) shows the corresponding color histogram.

7.3.3.2 Texture feature extraction

Textures are characteristic intensity variations that originate from the roughness of an object surface. The texture of an image is classified into first-order, second-order and higher-order

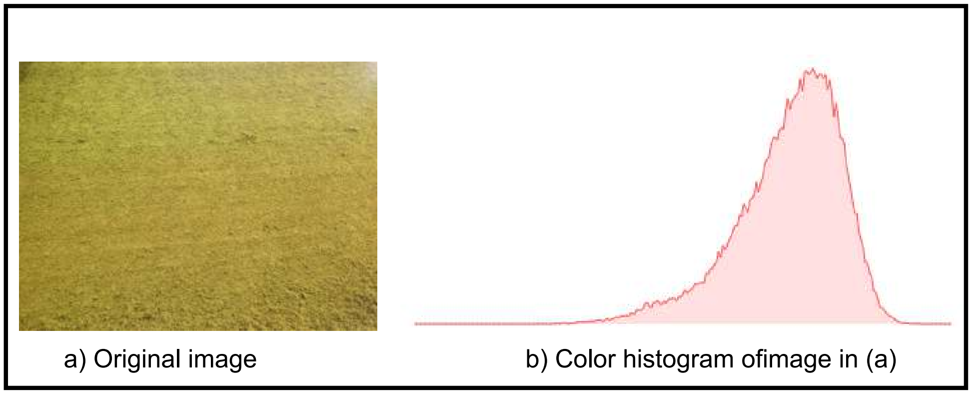


Figure 7.4: Generation of color features of an image using color histogram

statistics [145]. There are a variety of methods of extracting texture features including Local Binary Patterns (LBP), the Canny edge detection, discrete wavelet transform, gray level occurrence matrix and among others [146]–[148]. In this paper, we adopted LBP to extract the texture features of the images. LBP has many advantages which include: reduced histograms, considers the center pixels point effect [149] , and among others. The LBP algorithm is represented by equation 7.1 defined on a 3 × 3 lattice :

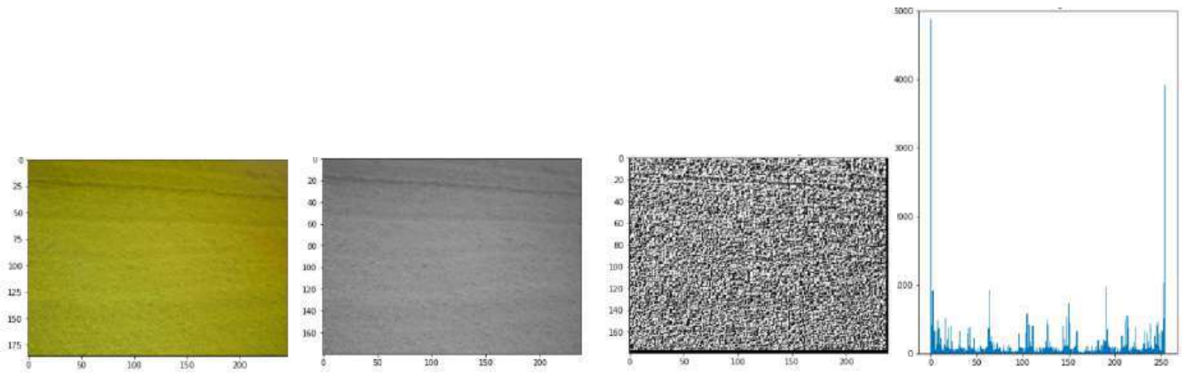
$$LBP_{\mathrm{x_c,y_c}} = \sum_{n=0}^{7} 2^{\mathrm{n}}(I_{\mathrm{n}} - I(x_{\mathrm{c}}, y_{\mathrm{c}})) \tag{7.1}$$

Where LBP $_{\mathrm{x_c,y_c}}$ is the value at the center pixel $\mathrm{x_c,y_c}$, $\mathrm{I_n}$is the values of neighbor pixel. I($\mathrm{x_c}$, $\mathrm{y_c}$) is the intensity at the center pixel.

The steps of the Texture Feature extraction were as follows:

1. The original image was converted into a grayscale image using the approach discussed in [150]. The color grayscale image generated is shown in Figure 7.5 (b)
2. LBP algorithm was then used to calculate each of the pixel in grayscale image as shown in Figure 7.5. Both LBP $_{\mathrm{x_c,y_c}}$ value and texture image are generated. The generated texture image is shown in Figure 7.5 (c)
3. Finally, the texture image obtained was converted into gray- scale histogram as shown in

Figure 7.5 (d)



(a) Original image (b) Gray scale of image in (a) (c) Texture image extracted by LBP (d) Gray-scale histogram

Figure 7.5: Conversion of image to grayscale histogram using LBP

7.3.4 Classification Models

We perform the classification of the images in the datasets discussed in Section 7.3.1 using the following classifiers: The next paragraphs discuss each of the classifiers.

7.3.4.1 Decision Tree

Decision Tree is a machine learning technique that employs a tree structure to specify the order of the decisions and the consequences [151]. During training, it generates rules and decision trees. The generated trees are followed in the classification of the new data [152]. It has the following constituents: root node, internal node and leaf node (Figure 7.6). Branches and leaves point to the factors that concern a particular situation [153].

It is one of the most used machine learning algorithms in classification [154], [155], because of its advantages which include: high tolerance to multicollinearity [156], flexibility, and exclusion of factors which are not important automatically [151], [157], among others. Some of the weaknesses of the classifiers are that: it is unstable and a small change in data pattern result in a completely different tree and are often inaccurate among others [90].

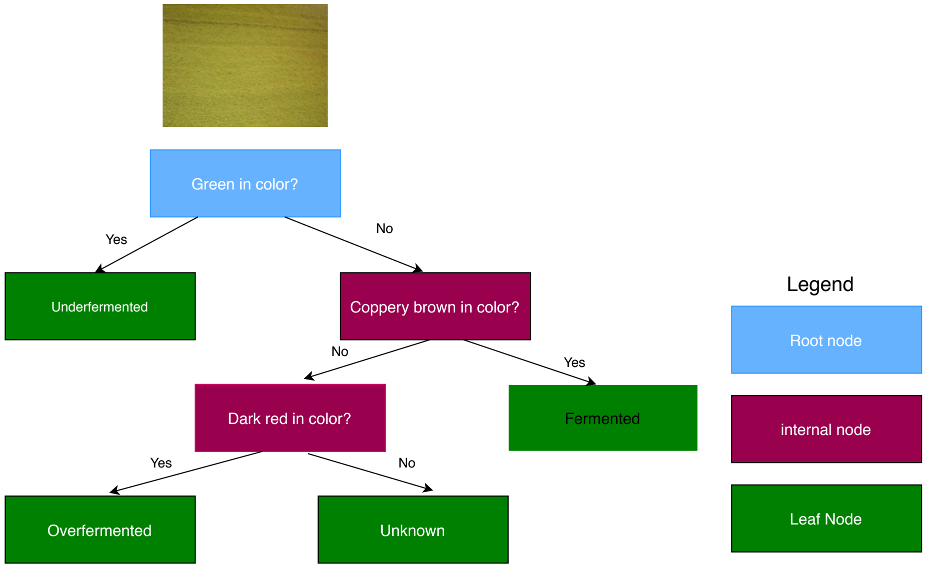


Figure 7.6: Example of classification by a Decision Tree

7.3.4.2 Random Forest

Random forest is a machine learning model that operates by constructing multiple decision trees during training [132], [158]. The constructed multiple trees are then used for prediction during classification. Each individual tree in the random forest outputs a class prediction and the class with most votes becomes the model's prediction [159]. Figure 7.7 shows an example of classifying an image using Random Forest. The image was classified as belonging to Class A since the majority of the trees (2) classified it as belonging to Class A. The classifier can estimate missing data, it can balance errors in datasets where classes are imbalanced and can be used for both classification and regression [160]–[163]. Additionally, it has better results compared to decision tree algorithm, the random forest has a better classification result [132], [164]. However, Random Forest is not as effective in regression tasks as it is in classification and is a black box model [89], [165], [166].

7.3.4.3 K-Nearest Neighbor

K-Nearest Neighbor (KNN) is a non-parametric machine learning model used for classification and regression [92], [167], [168]. In classification tasks, KNN determines the class of a new

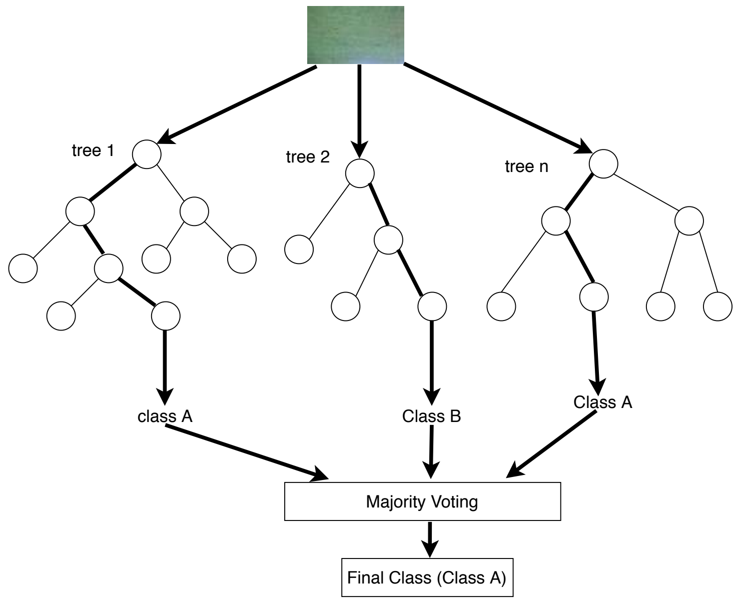


Figure 7.7: Example of a Random Forest operation

sample based on the class of its nearest neighbors. During decision making in a classification task, it finds k training instances that are closest to the unknown instance. It then picks the most occurring classification for the k instances [169]. It determines a dominant category to the target object in which k is the number of training samples. This algorithm assumes that samples are close to each other belong to the same category in classification [92]. Figure 7.8 illustrates an example of a classification using KNN. The task is to find a class that the triangle belongs to. It can either belong to the blue ball class or the green rectangle. The k is the algorithm we wish to take a vote from. In this case, let us say k=4. Hence we will make a cycle with the triangle as the center just to enclose only three data points on the plane. Clearly, the triangle belongs to the blue ball class since all of its nearest neighbors belong to that class.

The algorithm is simple to implement and has a robust search space [170]–[172]. The main challenge of the model is the expense incurred in terms of large computations in identifying neighbors in a large amount of data[173], [174].

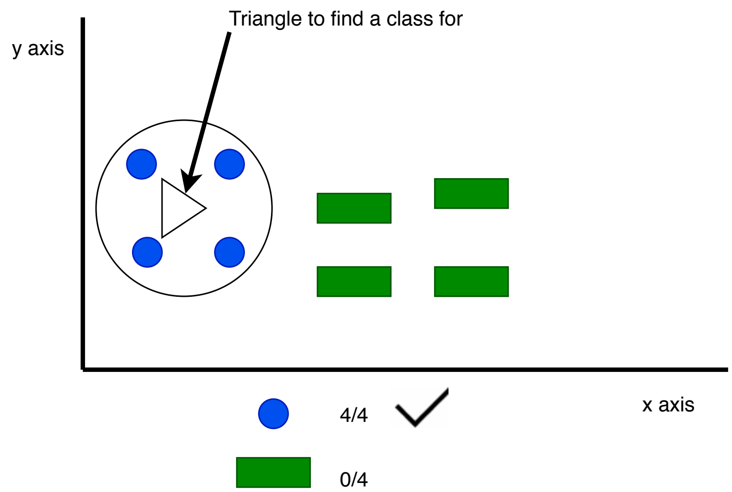


Figure 7.8: K-NN proximity algorithm map

7.3.4.4 Convolutional Neural Network

Convolutional Neural Network (CNN) is a class of deep learning technique that is currently emerging in solving computer vision challenges which includes detection of objects [175], segmentation [176], image classification [177], among others. CNN emerged in the mid-2000s due to the development in computing power of hardware of the computer [178]. A CNN is composed of the following layers (Figure 7.9): an input layer, convolutional layer, pooling layer, dense layer, and output layer. An input layer of a CNN is the layer where the input is passed to the network. In Figure 7.9, the input layer contains an image which needs to be classified [179]. Convolutional layers are a set of filters need to learn. The filters are used to calculate output feature maps, with all units in a feature map sharing the same weights [179]–[181]. A pooling layer will then sum up the activities and selects the maximum values in the neighborhood of each feature map [182]. A dense layer consists of neurons in a neural network which receive inputs from all the neurons in the previous layer [183]. Convolutional has shown high accuracy in image recognition tasks however they have high computation tasks [184].

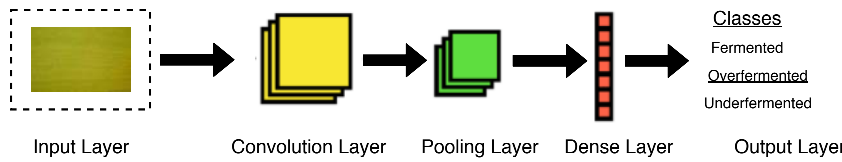


Figure 7.9: A typical CNN architecture

7.3.4.5 Support Vector Machine

Support Vector Machine (SVM) is a non-probabilistic binary classifier that aims at finding a hyperplane with a maximum margin to separate high dimension classes by focusing on the training samples located at the edge of the class distribution [185]. The model is based on statistical learning theory and the structural risk minimization theory [186]. The model chooses extreme vectors which help in creating the hyperplane. These extreme points are referred to as support vectors. In binary classification problem with linearly separable (Figure 7.10), has a goal to find the optimum hyperplane, through maximizing the margin and minimizing the classification error between each class.

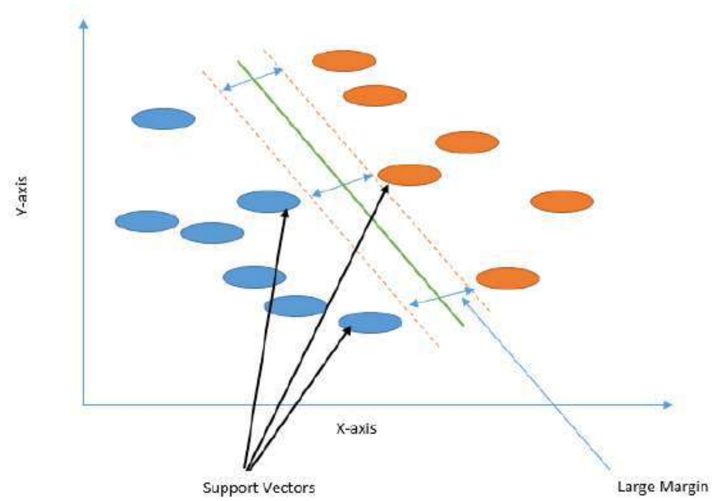


Figure 7.10: An Example of a classification task using SVM

Some of the advantages of SVM is its ability to rely on its own memory efficiency and ability to work well with classes having distinct margins [187], [188]. However, SVM tends to take large training time for a large dataset and is not effective for overlapping classes [177],

[189].

7.3.4.6 Naive Bayes

Naive Bayes is a probabilistic model based on Bayes' theorem. Bayes' theorem provides the relationship between the probabilities of two events and their conditional probabilities [94], [190], [191]. A Naive Bayes classifier assumes that the presence or absence of a particular feature of a class is unrelated to the presence or absence of other features[192], [193]. In classification tasks, NB constructs a probabilistic model of the features and applies the model in prediction of the new instances. Figure 7.11 shows a sample of balls belonging to two classes, yellow and green. The task is to estimate the class which the ball with a question mark belongs to. There is a very high probability that the ball belongs to class green since most of the balls belong to that class.

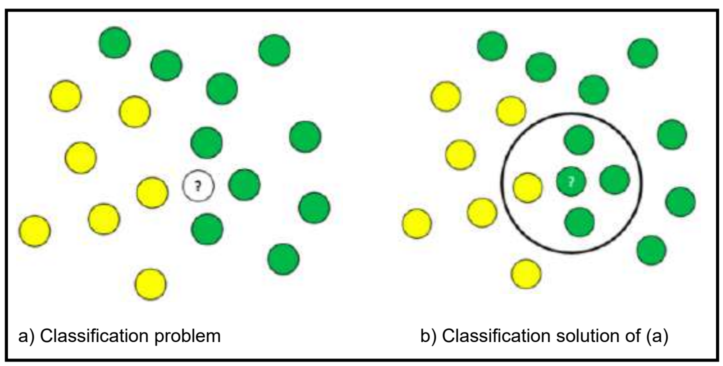


Figure 7.11: Example of classification using Naive Bayes

7.3.4.7 Linear discriminant analysis

Linear discriminant analysis is an approach developed by the famous statistician R.A. Fisher, who arrived at linear discriminants from a different perspective [194]. He was interested in finding a linear projection for data that maximizes the variance between classes relative to the variance for data from the same class [195]. LDA combines features of a class and builds on separating the classes. It models the differences between classes and builds a vector for differentiating the classes based on the difference in the classes [21], [196]. LDA is popular because of its low-cost implementation, its ease of adaptation for discriminating non-linearly separable classes, through the kernel trick method [197], and among others. Some of the weaknesses of LDA includes its challenge in handling large dataset among others [198]. Figure 7.12 (a) shows a classification problem while Figure 7.12 (b) shows the solution to the classification problem using LDA.

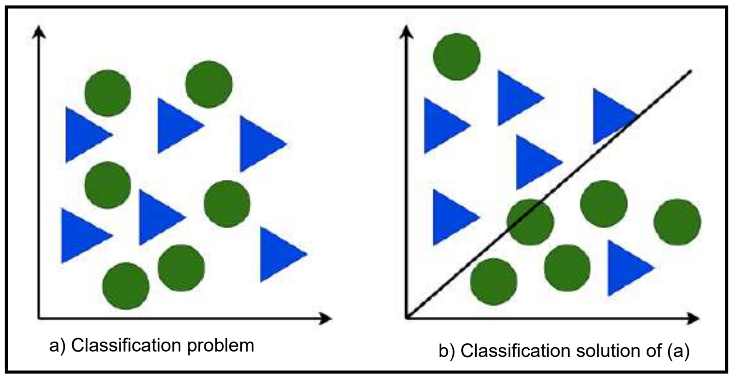


Figure 7.12: Example of classification using Local Discriminant Analysis

7.3.5 TeaNet

TeaNet is a deep learning model based on Convolutional Neural Network (CNN). The network architecture of TeaNet is an improvement upon the standard AlexNet model [199]. We chose AlexNet as inspiration because: it was among the earliest CNN architectures thus stable [200], [201], it has proven to show outstanding performance in an agricultural domain which is comparable to our current challenge [202]. We designed an optimum tea fermentation detection model with relatively simple network structure and a small computational needs. To construct TeaNet, we reduced the number of convolutional layer filters and the number of nodes in the fully connected layer. This reduced the number of parameters that require training thus reducing overfitting problem. Further reducing them reduces the cost of running the model and also reduced the training time. The patterns that our CNN was to learn was not too complex like the original challenge that AlexNet had to solve thus we had a liberty to reduce the complexity of the orginal AlexNet mode. The basic architecture of the network is shown in Figure 7.13.

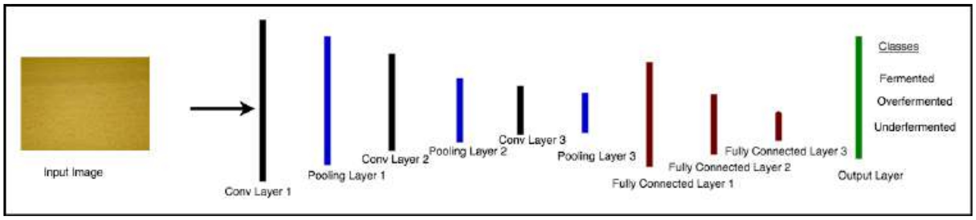


Figure 7.13: The architecture of the TeaNet that we propose for optimum detection of tea fermentation

The Input images were rescaled to 150 × 150 pixels and the three color channels discussed in Section 7.3.3.1 were all processed directly by the network. Table 4 shows the layer parameters of TeaNet.

Table 7.3: Layer parameters for the TeaNet

Layer	Parameter	Activation Function
input	150 × 150 × 3	-
Convolution1 (Conv1)	32 convolution filters (11 × 11), 4 stride	ReLU
Pooling1 (Pool1)	Max pooling (3 × 3) 2 stride	-
Convolution2 (Conv2)	64 convolution filters (3 × 3), 1 stride	ReLU
Pooling2 (Pool2)	Max pooling (2 × 2) 2 stride	-
Convolution3 (Conv3)	128 convolution filters (3 × 3), 3 stride	ReLU
Pooling3 (Pool3)	Max pooling (2 × 2) 2 stride	-
Full Connect4 (fc4)	512 nodes, 1 stride	ReLU
Full Connect5 (fc5)	128 nodes, 1 stride	ReLU
Full Connect5 (fc6)	3 nodes, 1 stride	ReLU
Output	1 node	Softmax

The layers are defined as follows:

1. The first convolutional layer comprises of 32 filters and a kernel size of 11 × 11 pixels. This layer is followed by a rectified linear unit (ReLU) operation. ReLU is an activation function that provides a solution to vanishing gradients [180]. Its pooling layer has a kernel size of 3 ×3 pixels, with two strides.

2. The second convolutional layer comprises of 64 filters and a kernel size of 3 × 3 pixels and is followed by a ReLU operation, its pooling layer has a kernel size of 2 × 2 pixels.

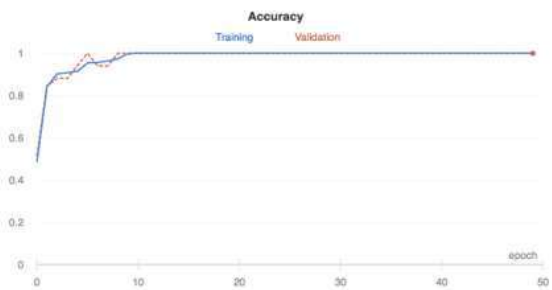
3. Additionally, the third convolutional layer comprises of 128 filters and a kernel size of 3 × 3 pixels and followed by ReLU with a kernel size of 2 × 2 pixels.

4. The first full connection layer was made up of 512 neurons and followed by a ReLu and a dropout operation. The dropout operation is proposed in CNN to solve overfitting as it trains only a randomly selected nodes [203]. In this study, We set the ratio of dropout to 0.5.

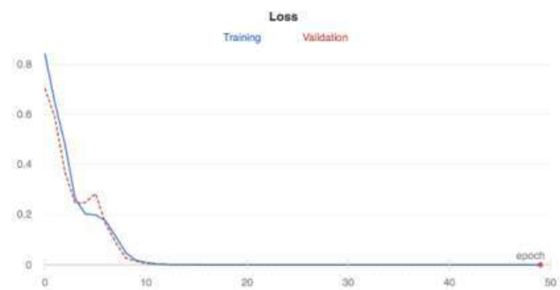
5. The second full convolutional layer had 128 neurons and was followed by a ReLU and dropout operations.

6. The last fully convolutional layer contains three neurons, which represent 3 classes of images in tea fermentation and LabelMe datasets. The output of this layer is transferred to the output layer to determine the class of the input image. A softmax activation function is then implemented to force the sum of the output values to be equal to 1.0. Softmax also limits the individual output values between 0-1.

At the beginning, the weights of the layers were initialized with random values from a Gaussian distribution. To train the network, a stochastic gradient descent (SGD) technique, with a batch size of 16 and a momentum value of 0.9 [204] were adopted. Initially, the learning rate across the network was set to 0.1 and a minumum threshold set at 0.0001. The number of epochs was set as 50 and the weight decay set to 0.0005. The accuracy of TeaNet increased with an increase in epoch and it achieved an accuracy of 1.0 at epoch 10 (Figure 7.14 (a)). At the beginning of the iteration, the accuracy is low since the weights of the neurons are not fully set. After each iteration, the weights are updated. The validation accuracy shows a steady increase and the model had an accuracy of 1.0. The loss of TeaNet during training and validation is illustrated in Figure 7.14 (b). There is a steady reduction in the loss from the first epoch up to epoch 10 where the loss value is at 0 for both training and validation sets. From Figure 7.14, the model has good performance and is not overfitted as it records good results in unseen data.



(a) Accuracy curve



(b) Loss curve of TeaNet

Figure 7.14: Accuracy and loss of TeaNet during training and validation

7.4 Implementation

To implement the classification models discussed in section 7.3.4 and TeaNet model discussed in section 7.3.5. The following are the specifications of computational platform used in performing the experiments: Processor: intel Core i7, Memory size of 64GB, RAM of 8GB, Hard Drives: 1TB, GPU: NVIDIA GeForce RTX 2070 8 GB, and Windows 10 Pro 64-bit operating system. python programming language was adopted. Some of the reasons for adopting python was: it has rich libraries [119], moderate learning curve [205], it is free and open source [206]. Some of the libraries that we adopted alongside python are: Tensorflow [207], [208], Keras [209], Seaborn [210], matplotlib [211], sklearn [212], [213], OpenCV [144], pandas [214] and numpy [215].

7.5 Performance Evaluation Metrics in Machine Learning

In this section, we describe the metrics that were used in evaluation of the ML models and results reported in Section 7.6.

Precision is the ratio of the correct classification to the total number of classifications [216], [217]. A low precision indicates a large number of false positives [197]. It can be represented by equation 7.2.

$$Precision = \frac{TP}{TP+FP} \tag{7.2}$$

Where: TP is an outcome where the model correctly classifies a class, FP is an outcome where the model incorrectly classifies a class.

Recall is the ratio of the number of correctly classified images to the total number of images [197], [216]. It is the actual positives that are correctly classified to the correct classes. Recall can be represented by equation 7.3.

$$Recall = \frac{TP}{TP+FN} \tag{7.3}$$

Where: TP is an outcome where the model correctly classifies the positive class, FN is an outcome where the model incorrectly classifies the negative class.

F1 Score is the harmonic mean between precision and recall. It tells how precise a classifier is in the classification tasks, as well as how robust it is [218]. It is represented by equation 7.4.

$$F1 - Score = 2 \times \frac{P \times R}{P+R} \tag{7.4}$$

Where P is the precision and R is the recall.

Accuracy is the fraction of predictions the model got right. Therefore it is the sum of correct predictions divided by all the predictions. It can be represented by equation 7.5.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{7.5}$$

Where: TP is an outcome where a model correctly classifies the positive class, FP is an outcome where a model incorrectly classifies the positive class, TN is an outcome where the model correctly classifies a negative class, FN is an outcome where a model incorrectly classifies a negative class.

Logarithmic Loss or Log Loss works by penalizing false classifications [216]. In classification tasks, it is the measure of the inaccuracy of classification. An ideal logarithmic loss should be 0. Logarithmic loss can be represented by equation 7.6.

$$Loss = -(g(log(p)) + (1-g)log(1-p)) \tag{7.6}$$

Where: g is the predicted probability, p is the true label.

A confusion matrix is used to summarize the classification performance of a classifier with test data [197]. Sensitivity in a confusion matrix measures the proportion of actual positives that are correctly identified and can be represented by equation 7.7.

$$Sensitivity = \frac{TP}{TP+FN} \tag{7.7}$$

Where: TP is the number of correct classification, while FN is an outcome where the model incorrectly classifies the negative class.

7.6 Evaluation Results

Results of the precision of the classifiers in the two datasets are shown in Figure 7.15. All the classifiers generally categorized majority of the images correctly. TeaNet classifier clearly performed better than the rest of the classifiers. TeaNet achieved an average precision of 1.0 and 0.96 on the tea fermentation and LabelMe datasets respectively. Generally, the majority of the classifiers except Decision Tree produced better precision in the fermentation dataset compared to the LabelMe dataset. This is because there was a distinctive change in color in the 3 categories of fermentation images. The classifiers recorded an average precision of between 0.78-1.00 in the fermentation dataset and 0.65-0.96 for the LabelMe dataset. From these results, any of the classifiers can be adopted for classifying tea fermentation. further, TeaNet recorded good results across the datasets and thus shows a lot of stability in the classification tasks and is a good fit for classifying tea fermentation images. Furthermore, these results are significant because majority of the models classified correctly most of the true positives. Even though NB recorded least results across the datasets, by achieving an average of 0.78 in tea fermentation implied that it was still applicable for this task.

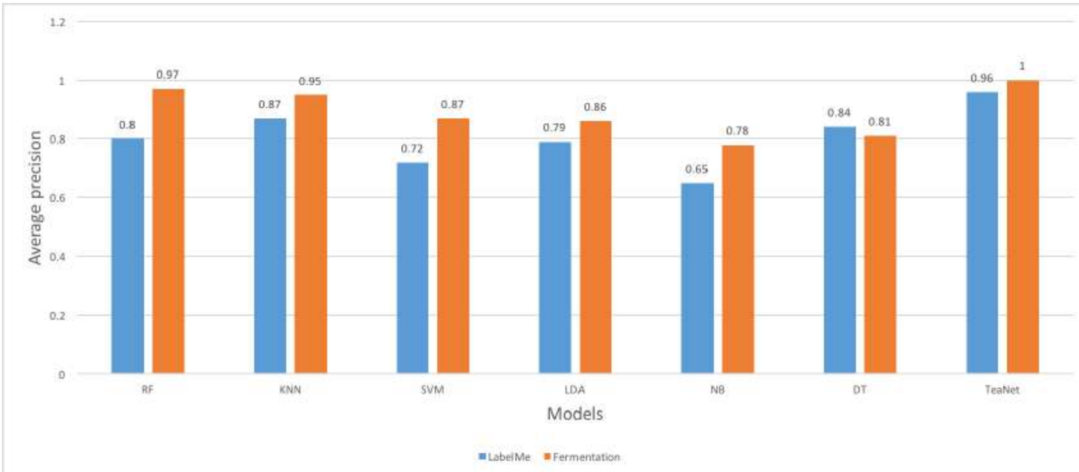


Figure 7.15: Precision of classification for each of the classifiers for the two datasets.

Recall values are illustrated in Figure 7.16. Once again TeaNet outperformed the other classifiers by producing the highest average recall values across the datasets. The majority of the classifiers had better performance in the tea fermentation dataset compared to the LabelMe dataset. The classifiers had an average recall of 0.75-1.0 for the tea fermentation dataset and an average of 0.58-0.96 for LabelMe dataset. KNN also had a good performance by recording an average recall of 0.93 and 0.85 for the tea fermentation and the LabelMe dataset respectively. Naive Bayes recorded the lowest recall values. From these results, it is evident that TeaNet and KNN produced the best recall values.

We compared the F1 score of TeaNet with the other classifiers and presented the results in Figure 7.17. The F1 score values of TeaNet was higher than the other classifiers. The classifiers recorded F1 values of between 0.58-0.9 for the LabelMe dataset and 0.75-1.00 for the tea fermentation dataset. We can note that TeaNet showed alot of effectiveness as it achieved an F1 of 1.00 in the tea Fermentation dataset and 0.9 for the LabelMe dataset(Figure 7.17). KNN also recorded a good performance of 0.93 and 0.85 for tea fermentation and LabelMe datasets respectively.

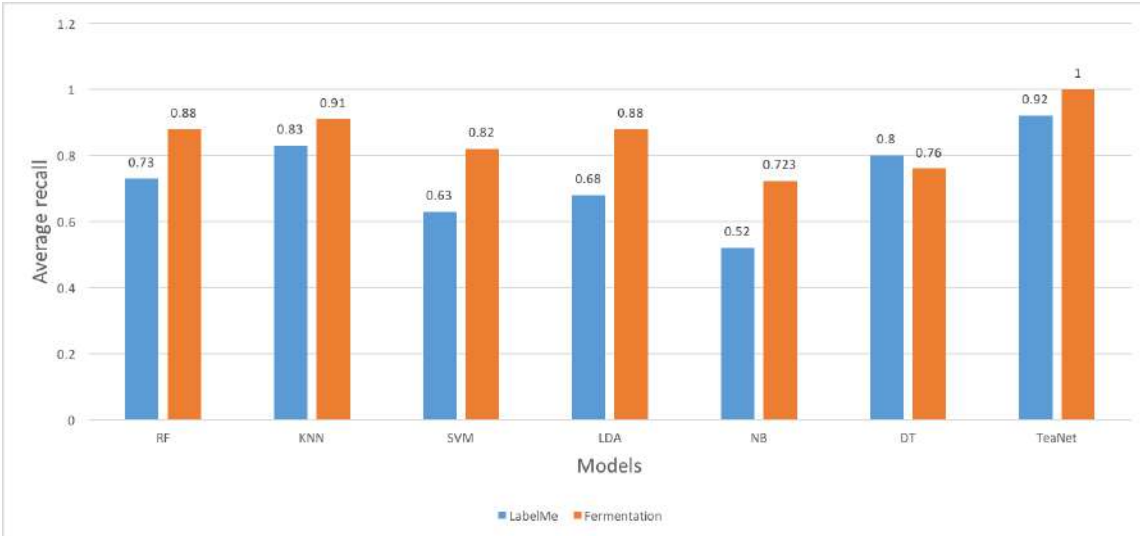


Figure 7.16: Recall of classification for each of the classifiers for the two datasets.

The performance of the classifiers in terms of accuracy is presented in Figure 7.18. The majority of the classifiers had good accuracy results. TeaNet achieved an average accuracy of 1.00 for the tea fermentation dataset and an average accuracy of 0.958 for the LabelMe dataset. This shows that TeaNet once again outperforms the other classifiers. Each of the classifiers produced an accuracy of more than 0.6 across the datasets. It shows that the probability of each of the classifiers in classifying the dataset is more than 60%. Naive Bayes recorded an average accuracy of 0.67 and 0.77 for the LabelMe and tea fermentation dataset respectively. On the other hand, Decision Tree recorded an average accuracy of 0.94 and 0.85 for the tea fermentation and the LabelMe datasets respectively. These results show that the majority of the classifiers can be applied to detect the tea fermentation images.

Finally, a confusion matrix was used to further evaluate the classification models (Table 7.4). The least specificity recorded by the classifiers was 73.5% and the highest was 100.0%. TeaNet recorded an average Sensitivity of 100% for fermented, an average of 100% for overfermented and finally an average of 100% for underfermented. From these experimental results, TeaNet outperformed the other classifiers in the classification tasks. In general, all the classifiers had good performance across the two datasets. Most importantly, none of the classifiers confused overfermented for underfermented tea. Most of them confused overfermented with fermented

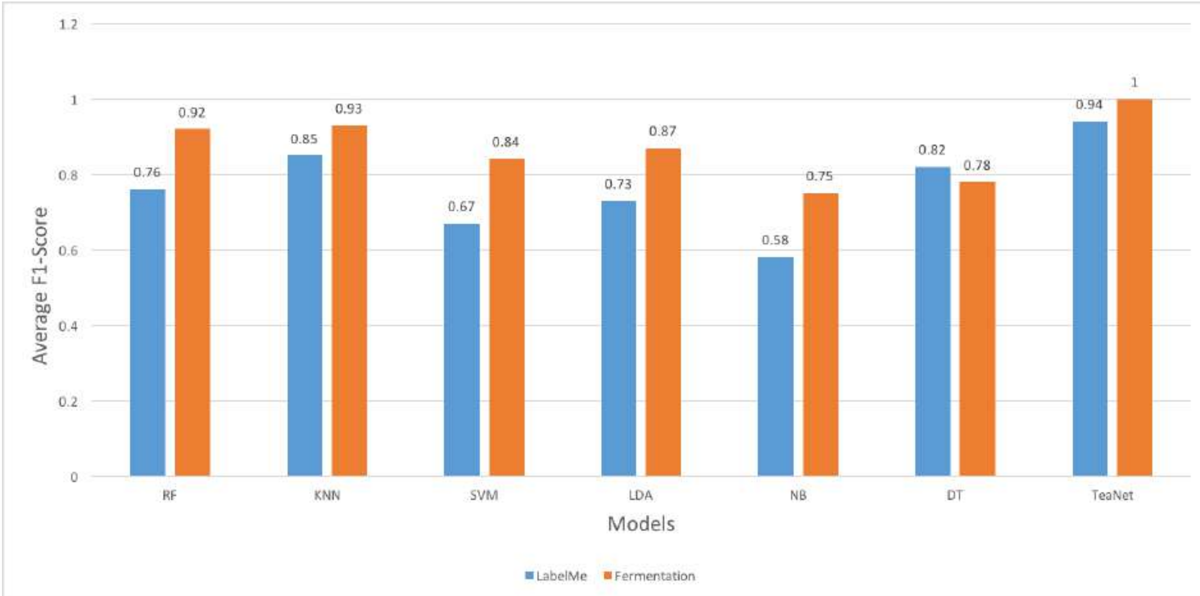


Figure 7.17: F1 Score of classification for each of the classifiers for the two datasets.

classes which is not a serious concern especially during transition between underfermented and fermented, and between fermented and overfermented, the images look very similar. These results show that the majority of the classifiers could be used in real deployments. Importantly, the effectiveness of TeaNet in the tea fermentation dataset was a great achievement.

7.6.1 Discussions of the Results

We have evaluated the performances of the following ML models using tea fermentation dataset and LabelMe datasets: Random Forest (RF), K-Nearest Neighbor (KNN), Decision Tree (DT), Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), TeaNet and Naive Bayes (NB). We evaluated them using the following metrics: precision, recall, f-measure, accuracy and sensitivity. From the evaluation results, TeaNet outperformed the rest of the classifiers across the metrics. This is in agreement with the state of the art [219] where DL have better performances than the traditional ML models because they are more sophisticated. Most importantly, all the classifiers recorded promising results as well. The study also highlighted that the standard ML models performed promisingly and thus in resource constrained environments where the

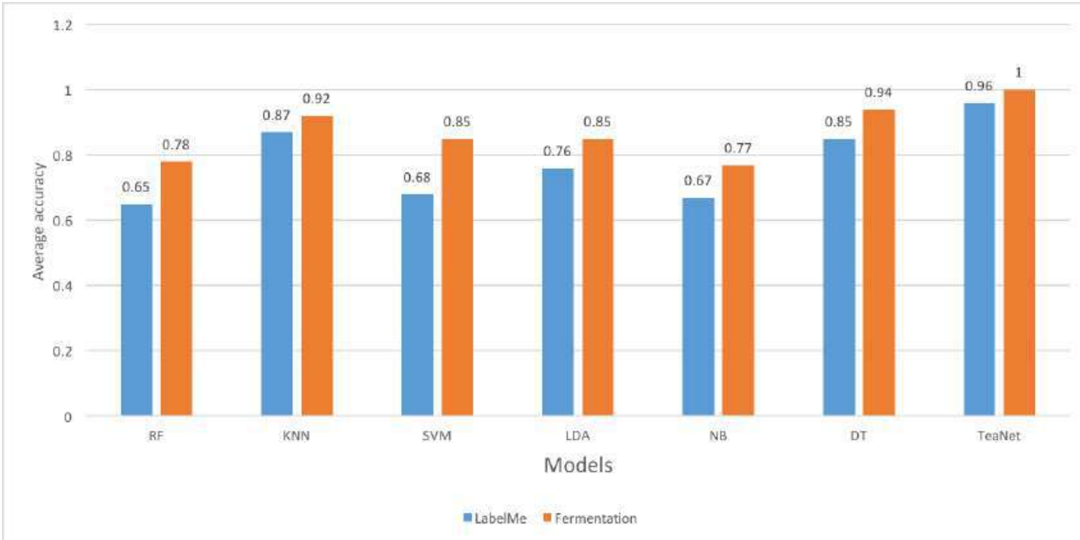


Figure 7.18: Accuracy of classification for each of the classifiers for the two datasets.

correctness of the classification is not critical, picking one of them can be satisfactory.

7.7 conclusion

In this Chapter, a deep learning model dubbed TeaNet was proposed. We have assessed the capabilities of TeaNet and other standard machine learning classifiers in categorizing images. We used tea fermentation and LabelMe datasets for training and evaluating the classifiers. From the experimental results, TeaNet outperformed the other classifiers in the classification tasks. In general, all the classifiers had good performance across the two datasets. These results show that the majority of the classifiers can be used in real deployments. Importantly, the effectiveness of TeaNet in the tea fermentation dataset is a great achievement.

Additionally, the results from this study highlight the feasibility of applying TeaNet in the detection of tea fermentation, which would significantly improve the process. This will, in turn, increase the quality of produced tea and subsequently increase the value of the made tea. This will lead to improved livelihoods of the farmers and the general improvement of the country's GDP. The same technique can be applied to the fermentation of coffee and cocoa. In our future studies, we will implement TeaNet with the Internet of Things in real deployment in a tea factory

Table 7.4: Confusion matrix recorded by the classifiers

Class	fermented	overfermented	underfermented	Sensitivity
DT(fermented)	250	32	78	69.4%
DT(overfermented)	59	301	0	83.6%
DT(underfermented)	271	0	89	75.3%
SVM (fermented)	296	22	39	82.2%
SVM (overfermented)	68	291	1	80.8%
SVM (underfermented)	61	0	299	83.1%
KNN (fermented)	339	14	7	94.2%
KNN (overfermented)	41	300	19	83.3%
KNN (underfermented)	17	0	343	95.3%
LDA (fermented)	331	11	18	92.0 %
LDA (overfermented)	17	335	8	93.3%
LDA (underfermented)	76	0	284	78.9%
RF (fermented)	325	14	21	90.3%
RF (overfermented)	50	310	0	86.1%
RF (underfermented)	45	0	315	87.5%
NB (fermented)	261	19	80	72.5%
NB (overfermented)	89	253	19	70.3%
NB (under fermented)	96	0	264	73.3%
TeaNet (fermented)	**360**	**0**	**0**	**100.0%**
TeaNet (overfermented)	**0**	**360**	**0**	**100.0%**
TeaNet (underfermented)	**0**	**0**	**360**	**100.0%**

to monitor fermentation of black tea.

Chapter 8

Prologue to Fourth Article

8.1 Article Details

Kimutai, G., Ngenzi, A., Rutabayiro Ngoga, S., Ramkat, R. C., and Förster, A.: An internet of things (IoT)-based optimum tea fermentation detection model using convolutional neural networks (CNNs) and majority voting techniques, J. Sens. Sens. Syst., 10, 153–162, https://doi.org/10.5194/jsss-10-153-2021, 2021.

Personal Contribution. After successful feasibility study and results reported in third article, Anna Förster, A. Ngenzi and R. Ngoga Said urged me to go for full deployment of the model. The formal analysis of the article was done by A. Förster; I did methodology together with A. Förster; I implemented the software for data collection with supervision, from A. Ngenzi, R. Ngoga Said, S., Ramkat, and A. Förster. I did the validation and visualization of the results. I also wrote the original draft while Ngenzi, R. Ngoga Said, S., Ramkat, and A. Förster performed writing review and editing. I produced all of the figures and tables. All authors read and agreed to the published version of the manuscript.

8.2 Context

After showing promising results in the feasibility study, we were motivated to implement the TeaNet model for real-deployment in the monitoring of tea fermentation in Sisibo tea factory, Kenya.

8.3 Contributions

The contribution of this paper is twofold. First, this study is the only attempt at implementing the internet of things and deep learning for the monitoring of tea fermentation. Second we share some of the lessons that we learned in the process of deploying the model in the factory.

8.4 Recent Developments

This paper is very recent therefore no more recent developments to report.

Chapter 9

An Internet of Things (IoT)-based Optimum Tea Fermentation Detection model using Convolutional Neural Networks (CNNs) and Majority Voting Techniques

Tea (Camellia sinensis) is one of the most consumed drinks across the world. Based on processing techniques, there are more than 15000 categories of tea, but the main categories include yellow tea, oolong tea, illex tea, black tea, matcha tea, green tea, and sencha tea, among others. Black tea is the most popular among the categories worldwide. During black tea processing, the following stages occur: plucking, withering, cutting, tearing, curling, fermentation, drying, and sorting. Although all these stages affect the quality of the processed tea, fermentation is the most vital stage as it directly defines the quality. Fermentation is a time-bound process, and its optimum is currently manually detected by tea tasters monitoring colour change, smelling the tea, and tasting the tea as fermentation progresses. This chapter explores the use of the internet of things (IoT), deep convolutional neural networks, and image processing with majority voting techniques in detecting the optimum fermentation of black tea. The model was made up of raspberry pi 3 models with a Pi camera to take real-time images of tea as fermentation progresses. We deployed the model in the Sisibo Tea Factory for training, validation, and evaluation. When the deep learner was evaluated on offline images, it had a perfect precision and accuracy of 1.0 each. The deep learner recorded the highest precision and accuracy of 0.9589 and 0.8646, respectively, when evaluated on real-time images. Additionally, the deep learner recorded an average precision and accuracy of 0.9737 and 0.8953, respectively, when a majority voting technique was applied in decision-making. From the results, it is evident that the model can be used to monitor the fermentation of various categories of tea that undergo fermentation, including oolong and black tea, among others. Additionally, the model can also be scaled up by retraining it for use in monitoring the fermentation of other crops, including coffee and cocoa.

9.1 Introduction

Tea (Camellia sinensis) is currently among the most prevalent and extensively consumed drinks across the world, with a daily consumption of more than 2 million cups. The high consumption is credited to its medicinal values, i.e. reducing heart diseases, aiding in weight management, preventing strokes, lowering blood pressure, preventing bone loss, and boosting the immune system, among others. Historical evidence indicates that the tea plant was indigenous to China and Burma, among other countries [18]. Tea is a source of many types of tea, which includes oolong tea, black tea, white tea, matcha tea, sencha tea, green tea, and yellow tea, among others. The processing techniques determine the category of tea produced. Globally, Kenya is the leading producer of black tea. Black tea is the most popular among the categories of tea, and it accounts for an estimate of 79% [7] of the entire global tea consumption. The processing steps of black tea are plucking, withering, cutting, tearing and curling, fermentation, drying, and sorting redas discussed in Chapter 1 and illustrated in Figure 1.1. The fermentation step is the most crucial in deciding the final quality of the black tea [125]. During the process, catechin compounds react with oxygen during oxidation to produce two compounds, namely theaflavins (TF) and thearubigins (TR). These compounds determine the aroma and taste of the tea [14]. Also, fermentation changes the tea colour to coppery brown and causes a fruity smell. Hence, the fermentation process must stop at the optimum point as fermentation beyond the optimum point destroys the quality of the tea [26]. Presently, tea tasters estimate the level of fermentation of tea by monitoring change in colour, smelling the tea, and tasting an infusion of tea [107]. These methods are biased, intrusive, consume a lot of time, and are inaccurate, which compromises the quality of the produced tea [42].

The internet of things (IoT) has established itself as one of the greatest smart ideas of the modern day [27] and its effects have been seen in each feature of human ventures, with huge possibilities for smarter living [220]. IoT has shown huge potential in many fields, including agriculture, medicine, manufacturing, sports, and governance, among others [221].

With the complication of challenges in the 21st century, majority voting is being applied to machine learning to improve performance as it provides an extra layer of decision-making to the model. Majority voting is an ensemble machine learning model that combines the predictions from multiple other models. It is a technique that may be used to improve model performance, ideally achieving better performance than any single model used in the ensemble [222]

Deep learning is currently shaping how machine learning is applied to various areas. Some of the prominent areas where deep learning has shown a lot of promise include machine vision, speech recognition, audio processing, health diagnosis, and fraud detection, among others [223]. In [107], a deep learner dubbed "TeaNet" was developed based on image processing and machine learning techniques. The deep learner was trained, evaluated, and validated based on the dataset in [97]. In this Chapter, the TeaNet model was deployed to classify real-time tea fermentation images in Sisibo Tea Factory. Additionally, a majority voting technique was applied to aid in decision-making by the model. The subsequent sections of this paper are presented as follows: Section 9.2 provides the related work, Section 9.3 provides the materials and methods, while Section 9.4 presents the evaluation results, and Section 9.6 gives the conclusion of the Chapter.

9.2 Related Work

As mentioned in Chapter 1, the fermentation process is the most important step in determining the quality of the produced tea. Consequently, researchers have been proposing various methods of improving the monitoring of the fermentation process of tea. There are proposals to apply image processing, IoT, electronic nose, electronic tongue, and machine learning, among others. With the maturity of image processing, many researchers are presently proposing it for application in the detection of optimum tea fermentation [224]. The application of image processing and a support vector machine algorithm in the detection of the optimum fermentation of tea has been proposed [26], [225]. Additionally, [26] proposed colour matching of tea during fermentation with neural networks and image processing techniques. Convolutional neural networks (CNNs) have shown great promise in image classification tasks across fields. Additionally, [226] and [227] show the great capabilities of CNNs in image classification tasks. They have shown that the data-hungry nature of deep learning has been mitigated by the aspect of transfer learning. Consequently, CNNs are now being applied in monitoring tea processing, including the detection of optimum fermentation of tea [107], [228]. Additionally, CNNs have been adopted to detect diseases and pest-infected leaves [179], [229]–[231]. Furthermore, a neural-network-based model for estimating the basic components of tea theaflavins (TF) and thearubigins (TR) is proposed [18]. A study in [225] fused near-infrared spectroscopy with computer vision to detect optimum fermentation in tea. All these proposals are in the form of simulation models. They have reported promising results, but they are yet to be deployed in real tea-processing environments.

IoT is being applied in many fields, including agriculture. The tea sector is attracting attention from researchers, and the authors have proposed the application of IoT to monitor temperature and humidity during the fermentation of tea. A study in [84] proposed a sensor network to monitor the relative humidity and temperature of tea during fermentation. Also, [85] developed an IoT-based system for monitoring the temperature and humidity of tea during processing. The proposed works have been deployed in a tea factory to monitor temperature and humidity during tea processing. The models are, thus, a step in the right direction, but their scope was only on monitoring those physical parameters during tea processing. From the literature, it is evident that the tea fermentation process is receiving most of the attention from researchers due to its importance in the determination of the quality of tea, with many of the proposals being an ensemble of machine learning and image processing techniques. The IoT is presently gaining momentum in its application to monitoring temperature and humidity during tea processing. However, most of the proposals are in the form of simulation models and have not been deployed in real tea fermentation environments. Additionally, deep learning is gaining more acceptance compared to standard machine learning classifiers in monitoring tea processing due to its intelligence and the ability to use transfer learning to solve challenges across various domains.

9.3 Materials and Methods

This section presents the following: the system architecture, resources, deployment of the model, image database, majority voting-based model, and the evaluation metrics.

9.3.1 System architecture

The architecture of the proposed model to monitor the fermentation of tea in real time is presented in Figure 9.1. The system had a node containing a Pi camera attached to a Raspberry model. The system is connected through Wi-Fi technology to the edge and the cloud environments.

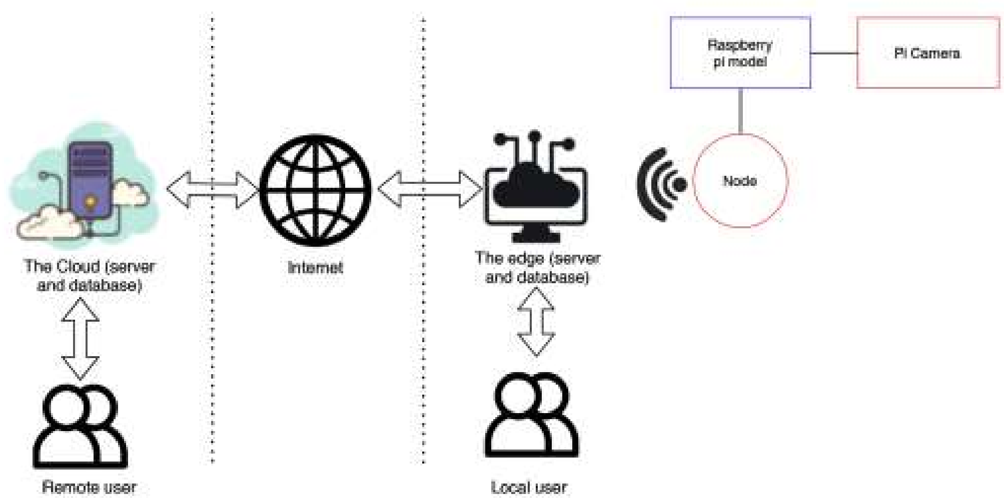


Figure 9.1: System architecture for the IoT-based optimum tea fermentation monitoring system

9.3.2 Resources

The following resources were applied in the implementation: Raspberry Pi 3 Model B+, a Pi camera, an operating system, a server, and programming languages. We discuss each of these in the next paragraphs:

1. Raspberry Pi 3: This study adopted Raspberry Pi 3 Model B+ with the following specifications: quadcore 1.2 GHz, a 64 bit central processing unit (CPU), 2 GB random-access memory (RAM), a 40-pin extended general-purpose input/output (GPIO), camera serial interface port, and micro secure digital (SD) card. The model was powered by 2A power supply [116].

2. Pi camera: In this research, a Raspberry Pi camera of 8 MP was used. The board was chosen since it is tiny and weighs around 3 g, making it perfect for deployment with the Raspberry Pi.

3. Operating System: The Raspbian operating system [116] was used. It was chosen because it has a rich library, and it is easy to work with.

4. Server: The Apache server [232] was adopted to obtain data and send the data to the edge environment for the local users and a cloud-based environment for remote users. For the cloud environment, Amazon Web Services (AWS) [118] was chosen as it provides a good environment for the deployment of IoT systems.

5. Programming Languages: The Python programming language [233] was used for writing programmes to capture the images using the Pi camera. It has various libraries [119] and is open source [205]. Some of the libraries adopted included the following: TensorFlow [213], Keras [207], Seaborn [209], Matplotlib [211], pandas [207], and NumPy [215]. Additionally, the Laravel PHP framework was adopted in writing application programming interfaces (APIs). HTML5 (the hypertext markup language) and CSS (cascading style sheets) were used in designing the web interfaces of the model.

9.3.3 Image Database

As discussed in Section 9.1, TeaNet was trained, validated, and evaluated using data set from [97]. The data set contained 6000 images of three classes of tea, i.e. underfermented, fermented, and overfermented. Fermentation experts provided the ground truths of all the images, which enabled the classification of all the images into the three classes. From the experts' judgement, fermentation degrees of tea with time depends on the following factors: time [234], temperature and humidity level at which fermentation takes place, the clones of the tea, nutrition levels of the tea, age of tea, stage of growth of tea, plucking standards, and post-harvesting handling. Presently, more than 20 clones of tea are grown in Kenya [40]. We presented an example of each image of the classes in Figure 7.13.

9.3.4 Majority Voting for TeaNet

We developed a deep learning model that is dubbed TeaNet [107] and is based on CNNs. AlexNet [199], the widely used network architecture in CNNs, inspired the development process. We designed the model for simplicity and to reduce the computational needs. TeaNet was chosen over the traditional methods since it outperformed the traditional methods in the simulation experiments reported in Chapter 7 and published in [107]. Additionally, TeaNet as a deep-learning-based model is trained rather than programmed; thus, it does not require much fine-tuning. TeaNet is flexible as it be can be retrained using other data sets for other domain applications, unlike OpenCV algorithms that are domain-specific [235]. With transfer learning [204], TeaNet can be applied to solve challenges in other fields. Figure 7.13 shows the architecture of the developed TeaNet model. The architecture of the TeaNet that we propose for optimum detection of tea fermentation was discussed in Chapter,7 and illustrated in Figure 7.13.

Gaussian distribution was adopted in initializing the weights of the network. A stochastic gradient descent [204] technique, with a batch size of 16 and a momentum value of 0.9, was chosen. The rate of learning was 0.1, with a threshold minimum of 0.0001. The network learning iterations were set at 50, with a weight decay of 0.0005. The model registered a steady increase in accuracy with increasing epoch numbers, registering a perfect precision of 1.0 at epoch 10. The accuracy of the model increased with each iteration as the weights of the neurons were turned after every iteration. The validation accuracy of the model was 1.0 at epoch 10 (7, Figure 7.14 (a)). The loss of the model during training and validation was illustrated in 7, Figure 7.14 (b). The loss steadily reduces, with epoch increases up to epoch 10 where there is no loss. These values are promising, as they depict a stable loss rate.

During real-deployment, TeaNet did not show promising results as it did during the simulation and reported in Chapter 7. To solve this challenge, we proposed a region-based majority. (RBM) voting for TeaNet (RBM-TeaNet) composed of three steps.The steps were image segmentation, the training of the TeaNet model, and majority voting (Figure 9.2). After the image collection, as discussed in Section 3.3, the images were prepared for input into the CNN network by resizing them to 150 × 150 pixels. The semantic segmentation annotation method was followed to annotate the images according to [136]. Some of the common types of noise in the images include photon noise, readout noise, and dark noise [236]. To perform denoising, the linear filtering method was adopted. For the region-based majority voting for TeaNet (RBMTeaNet), each region was labelled by voter data generated by the region majority voting system. Each of the patches had three voters. One of the voters was in the centre of the patch, with the others being generated randomly within the patch. Finally, the classification results were arrived at from the candidate label that had the highest vote numbers (Figure 9.2).

9.3.5 Deployment of the model

We deployed the developed tea fermentation monitoring system in a tea fermentation bed in the Sisibo Tea Factory, Kenya (Figure 9.3). The Pi camera was attached to a Raspberry Pi model and used to take images of the tea in the fermentation bed at an interval of 1 min. The learned model developed in [107]was trained and deployed in the Sisibo Tea Factory for validation

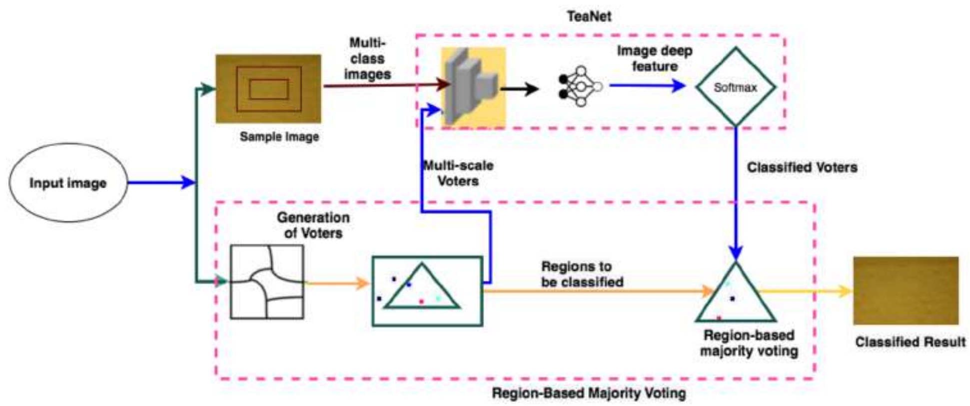


Figure 9.2: The proposed workflow of the region-based majority voting for TeaNet.

between 1 and 30 July 2020 and thereafter evaluated between 10 and 16 August 2020. In the tea fermentation bed, a Pi camera was deployed to take tea fermentation images in real time. The server side contained the Raspberry Pi, a Wi-Fi router, and internet wall.

Figure 9.3: Application of Raspberry Pi 3 Model B+ with the Pi camera for taking pictures during the fermentation of black tea in the Sisibo Tea Factory, Kenya.

Every collected image was sent to the Raspberry Pi through the jumper wires. In Raspberry, the TeaNet model predicted the classes of images based on the knowledge gained during training. The image was sent to the cloud servers by the Raspberry Pi with the use of Wi-Fi technology for use by remote users. The internet wall was used to secure the connection to cloud servers. A copy of each image was then sent to the edge servers locally for use by local users. Each realtime image was displayed on the web page, which is accessible through both mobile phones and computers. Additionally, the web page displayed images alongside their predicted classes.

9.4 Evaluation Results

We evaluated the model based on precision, accuracy, and confusion matrix. These metrics have been discussed in Chapter 7 Section 7.5 . Figure 9.4 shows the evaluation results of the model based on average precision. The evaluation results showed that TeaNet produced a perfect simulation precision of 1.0, with an average of between 0.8485 and 0.9589 in real deployment. The model produced an average precision of between 0.9260 and 0.9737 when TeaNet and majority voting were combined. From the results, TeaNet performed better in terms of precision when evaluated offline compared to when it was evaluated in real deployment.

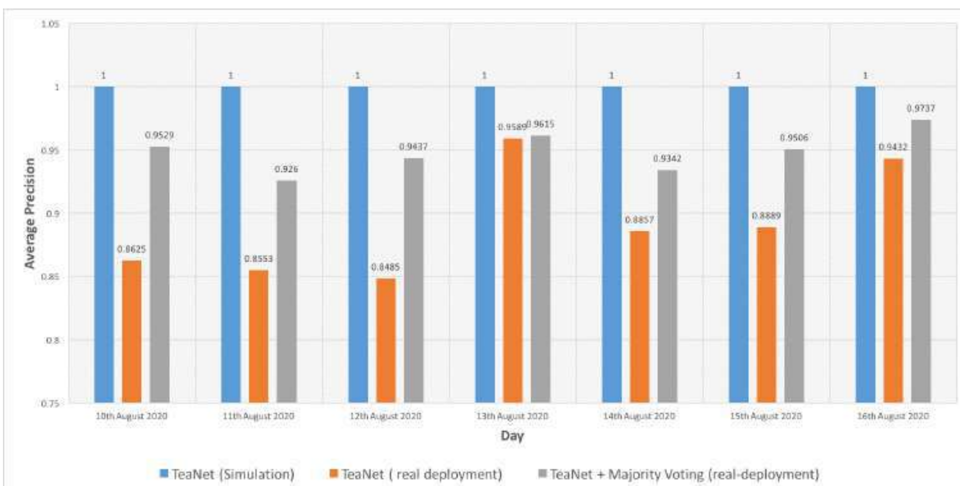


Figure 9.4: The average accuracy of the TeaNet system in monitoring optimum tea fermentation.

Figure 9.5 shows the evaluation results of the TeaNet model based on accuracy. Also, TeaNet showed high effectiveness when evaluated offline, based on the achieved average accuracy of 1.0 across the scanning days. When TeaNet was evaluated in a real deployment environment, it achieved an average accuracy of between 0.7179 and 0.8646 across the scanning days. Additionally, when majority voting was employed to aid in the decision-making process of TeaNet, performance improved to an average ranging between 0.8372 and 0.8916.

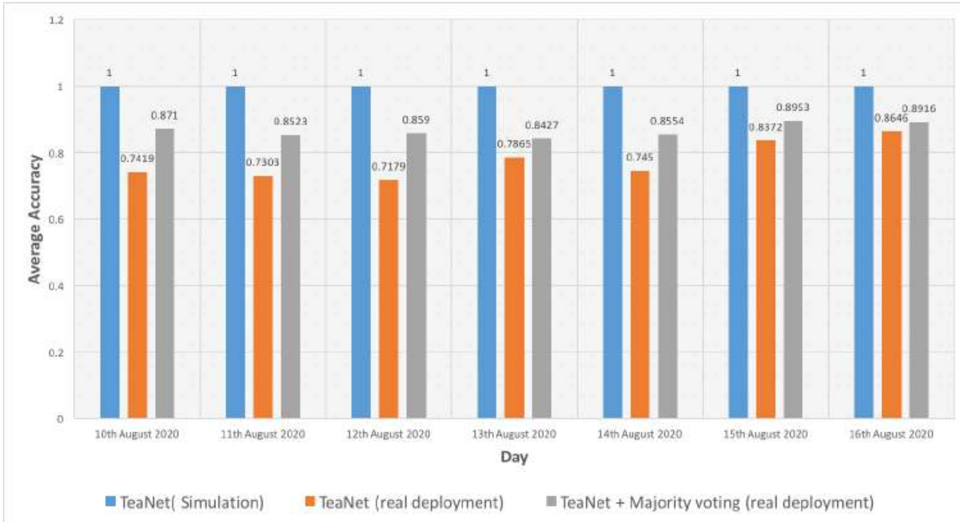


Figure 9.5: The average accuracy of the TeaNet system in monitoring optimum tea fermentation

Table 9.1 presents the performance of the RBM-TeaNet model in terms of sensitivity. In the tea factory, overfermented tea was not found, since such tea is low in quality and no tea factory allows the fermentation of tea to reach that level. Generally, the model had good sensitivity across days, with a minimum of 70.3% being achieved on 13 August 2020 where the model classified six fermented images as underfermented and five of the fermented images were classified as overfermented. More promisingly, the model could not confuse unfermented and overfermented tea. This is because of the clear distinction in the two classes in terms of colour.

Table 9.1: Confusion matrix of RBM-TeaNet model during the fermentation of tea

Class	**fermented**	**overfermented**	**underfermented**	**Sensitivity**
10thAugust 2020 (unfermented)	41	4	0	91.1%
10thAugust 2020(fermented)	3	39	5	83.0%
11thAugust 2020 (unfermented)	45	6	0	88.2%
11thAugust 2020(fermented)	6	29	2	78.4%
12thAugust 2020 (unfermented)	42	5	0	89.4%
12thAugust 2020(fermented)	3	23	4	76.7%
13thAugust 2020 (unfermented)	48	3	0	94.1%
13thAugust 2020(fermented)	6	26	5	70.3%
14th August 2020 (unfermented)	42	5	0	89.4%
14th August 2020(fermented)	3	28	4	80.0%
15th August 2020 (unfermented)	48	4	0	92.3%
15th August 2020(fermented)	0	28	5	84.8%
16th August 2020 (unfermented)	45	2	0	95.7%
16th August 2020(fermented)	4	27	3	79.4%

9.5 Discussion

From the evaluation results TeaNet recorded lower performances across the days compared to when it was evaluated during simulation as reported in Chapter 7. This could be due to noise in images when the camera was taking the images. However, the performance of the model was improved when majority voting technique was introduced to add an extra layer of decisions. The results thus confirmed that TeaNet was a good fit for the task. Further, the proposed model was powered the the electrical energy from the grid, thus, the solution could only be provided in tea factories which are connected to the grid.

9.6 Conclusion and Recommendations

This chapter has proposed a tea fermentation detection system based on IoT, deep learning, and majority voting techniques. The IoT components were Raspberry Pi 3 Model B+ and a Pi camera. The deep learner model was composed of three convolutional layers and three pooling layers. The model developed was deployed to monitor tea fermentation in real time in a tea factory in Kenya. The capabilities of the system were assessed based on the ground truths provided by tea experts. The results from the evaluation are promising and signify a breakthrough in the application of IoT, CNNs, and majority voting techniques in the real-time monitoring of tea fermentation. The same technique can be applied to monitor the processing of other categories of tea that undergo fermentation, including oolong tea. Additionally, the model can be used for monitoring the fermentation of coffee and cocoa, since all of them have a distinction in colour based on the fermentation degrees. It is recommended that future studies monitor the physical parameters (temperature and humidity) of tea during fermentation to find their effect on the quality of the made tea.

Chapter 10

Prologue to Fifth Article

10.1 Article Details

G. Kimutai, A. Ngenzi, S. R. Ngoga and A. Forster, ”Offloading an Energy-Efficient IoT Solution to the Edge: A Practical Solution for Developing Countries,” 2021 IEEE Global Humanitarian Technology Conference (GHTC), pp. 265-272, doi: 10.1109/GHTC53159.2021

Personal Contribution. The idea of writing this paper came out of a meeting with Anna Förster,A. Ngenzi and R. Ngoga Said. Formal analysis was done by A. Förster; I did methodology together with A. Förster; i implemented the software for data collection with supervision, from A. Ngenzi, R. Ngoga Said, and A. Förster. I did the validation and visualization of the results. I also write the original draft while A. Ngenzi, R. Ngoga Said, and A. Förster performed writing review and editing. I produced all of the figures and tables. All authors read and agreed to the published version of the manuscript.

10.2 Context

At the time that we wrote this article, machine learning was showing a lot of promise in various domains. However, from the first article, we realised that the same success had not be realised in the detection of optimum tea fermentation. This motivated us to explore machine learning for the detection of optimal tea fermentation.

The study in this chapter was presented in the 2021 Global Humanitarian Technology Conference(GHTC 2021) and won the best paper award. The work is published in IEEE Xplorer as a conference proceedings and is available in [237]

10.3 Contributions

The contribution of this paper are threefold: first a proposed approach for deploying an IoT-based model on an Edge environment which is closer to the end-users. Second, implementation of the proposed approach including the necessary hardware and software changes and adaptations and its evaluation on deployment environment. Third, powering the IoT-based solution using Photo-voltaic (PV)-based power. This makes the solution usable in developing countries where the power supply is unstable.

10.4 Recent Developments

This paper is very recent, therefore no more recent developments to report.

Chapter 11

Offloading an Energy-Efficient IoT Solution to the Edge: A Practical Solution for Developing Countries

Agriculture contributes to the economies of many developing countries. Tea is the most popular crop in Kenya as it contributes majorly to her economy. Among the various stages of processing tea, fermentation is the most important as it determines the final quality of the processed tea. Presently, the process of monitoring is done manually by tea tasters by tasting, smelling, and touching tea which compromises the quality of tea. In this Chapter, a deep learner dubbed ”TeaNet” is deployed in Edge and Fog environments for real-time monitoring of tea fermentation. We power the system using a Photovoltaic (PV) energy source to overcome the challenge of unreliable power supply from the grid. Furthermore, the energy consumption of the solution is reduced by applying duty cycling, where idle components are designed to sleep. We used the Analysis of variance (ANOVA) and Post-hoc for data analysis. From the results, Edge registered the lowest latency compared to the Cloud and Fog environments. During deployment of the energy optimized model, 50.6559Wh amount of energy was saved. This study recommends that the task offloading model proposed in this study be explored in offloading tasks in other fields.

11.1 Introduction

Although the IoT Solution presented in Chapter 9 was successful, it could not be deployed in many tea factories as most of them were not connected to the electrical grid [238], [239]. Furthermore, unreliable internet connection and costs associated with internet connectivity limited its functionality as most of the areas in developing countries are connected with only 2G or with some areas with 3G and 4G, which are intermittent and expensive [240]. Thus it was essential to achieve independence of energy, optimize the available power and eradicate the need for internet connections in operations by offloading the tasks to the Edge environment.

11.2 Related Work

Presently, several studies have been conducted on offloading tasks from Cloud to Edge and Fog environments. Edge environment have been adopted forhosting Machine Learning-based solution for medication application has been proposed in [241]. The feasibility study showed that the model which was running on an ensemble of KNN, NB, and SVM was able to work in the resource-constraint environment. Most importantly the medication solution could provide solution in real-time because of the reduced latency in the Edge as it is near the creation of service. Authors in [242] proposed a model for automatically decide whether to offload tasks to the Edge or not. The model is in the form of a mathematical model and it will be interesting to test its performance when implemented. Research in [243] proposed an approach of offloading tasks to the fog devices when the power consumption of the mobile devices are greater than the consumption of the wifi devices. The approach promises to reduce the energy consumption as most of the computations are offloaded to the Fog environment. Multiple Edge servers complementing each other have been experimented in a study reported in [244]. These approach of having multiple edge devices working together ensured that computations can be done by another devices when one of them is down and assures availability. A Fog computing model encompassing Cloud and local nodes has been proposed in [245]. Thus computationally intensive tasks were done by the cloud while lighter ones by the fog devices. This reduced the cost of computation.

Minimizing the power consumption of Raspberry Pi (RP*i*) is an active research area [246]. A mathematical model on RP*i* power consumption is presented in [247].Reducing power consumption of RP*i* by sleeping when not active is reported in [248]. Generally, the RP*i* is accepted for application in various domains. However, the major limitation is that it is power-hungry compared to other IoT Platforms. From these studies, it is evident that Fog and Edge environments are being considered as an alternative to the Cloud as they promise lower latency and cheaper in terms of cost.

11.3 Materials and Methods

This section discusses proposed system architecture, energy harvesting components, IoT-based data collection components, proposed energy optimization approach, and evaluation metrics.

11.3.1 System Architecture

The proposed system was composed of the following components: energy harvesting model, IoT-based data collection module, and the hosting environment (Figure 11.1).

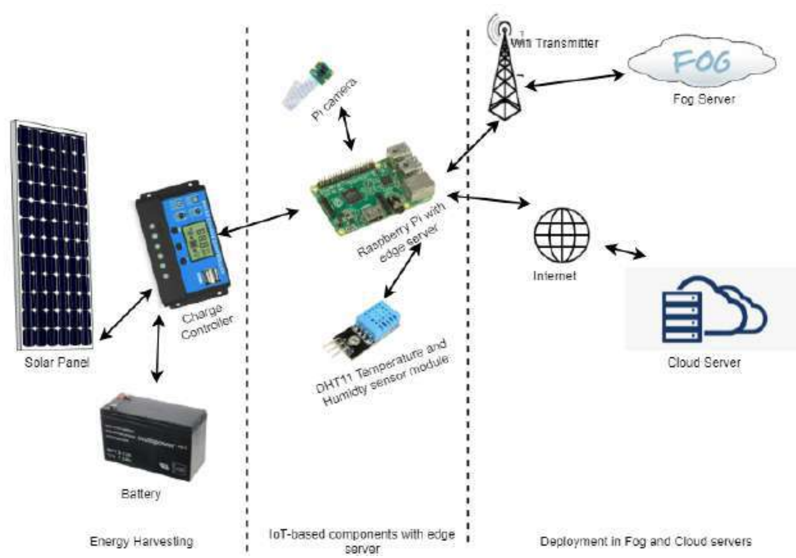


Figure 11.1: The system architecture of the task offloading model for IoT

The hosting environment involved Edge, Fog and Cloud. Edge and Fog servers are based on offloading computations closer to the source of data [249]. Edge server takes computation and storage to the device while Fog takes the computation to the gateways for processing by the resources at the Local Area Network (LAN) [250]. Edge and Fog are designed to complement the cloud by providing cheap and secure deployment. Additionally, due to their low latency, they are ideal for deploying real-time applications.

11.3.2 Energy Harvesting

Figure 11.2 shows the deployment of the PV-based energy harvesting model for IoT.

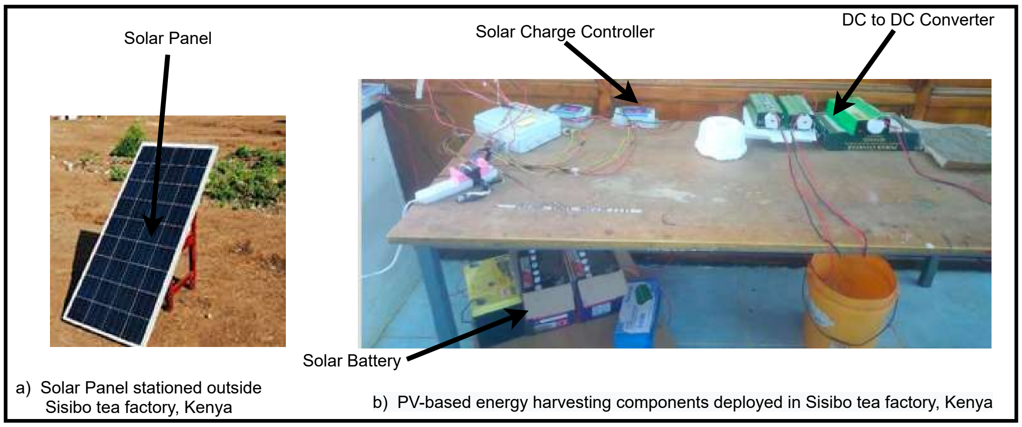


Figure 11.2: Pv-based power harvesting for IoT

The energy harvested must power the IoT-based solution and a bulb to illuminate the fermentation bed for uniform lighting. The longest time a fermentation could take place in the Sisibo tea factory was 10 hours. Equation 11.1 gives the total amount of energy used by the IoT-based solution during the period.

$$Energy\_IoT = 3.44w \times 10h = 34.4Wh \tag{11.1}$$

The bulb had a rating of 15w thus the total energy consumed is given by equation 11.2:

$$Energy\_bulb = 15w \times 10h = 150Wh \tag{11.2}$$

The total amount of energy consumed by the bulb and the IoT solution daily is given by Equation 11.3:

$$Energy\_total = 150wh + 34.4 = 184.4Wh \tag{11.3}$$

The efficiency of the battery, DC-DC convertor, and the Charge Controller was 80%, 90%, and 85% respectively. Therefore the solar array load is given by Equation 11.4:

$$PV\_array = \frac{184.4wh}{0.8 \times 0.9 \times 0.85} = 301.3071Wh \tag{11.4}$$

The worst solar radiation in the Eldoret region where the Sisibo factory is located is witnessed in the months of August, with daily radiation of 5 hours. Thus we take this as the peak solar radiation. Therefore the array load is given by Equation 11.5. Since Kenya lies within the equator, the solar mismatch factor was minimal and thus ignored. Thus a solar panel of more than 60 w could serve the model. We, therefore, chose a solar panel of 100w more than the minimum of 60.26 and was also most widely available in the market at a fair price.

$$Array\_Load = \frac{301.3071wh}{5h} = 60.26W \tag{11.5}$$

A deep discharge solar battery was chosen as it is durable and works well with solar panels. The battery had a maximum allowable discharge at 60% and had a rating of 12V. Thus the good daily shot and the energy store is given by equation 11.6:

$$Energy\_Required = \frac{184.4wh}{0.6} = 307.3333Wh \tag{11.6}$$

Therefore the rating of the battery is given by equation 11.7. A battery of 30A/h was chosen as it was the nearest rating available in the market.

$$rating\_battery = \frac{307.3333wh}{12v} = 25.61A/h \tag{11.7}$$

The specifications of the Energy harvesting system were: 100w solar panel, 12v 30 A/h battery with its charge controller, and DC-DC converter.

11.3.3 IoT-Based Components

The IoT-based hardware used in this experiment was the same as those reported in our prior study [251]. They had: Raspberry pi 4 model B, Micro SD card, Pi camera, and DHT11 temperature and humidity sensors. The micro SD card of memory size 16 GB and a Pi camera of 5 megapixels were used.

11.3.4 Optimization of Energy Consumption of IoT-Based Devices

To reduce the energy consumption of the model, we applied duty cycling or sleeping, where idle components sleep when not performing tasks. In our case, we took an image every 60 seconds and the processing time takes approximately 100 micro-seconds. Thus, our system was idle for more than 60% of the time in an hour. By putting the GPIO pins, Pi-camera, Temperature and humidity sensor, to sleep when idle, we potentially saved a lot of energy, even if also sleeping modes consume some, but insignificant amount of energy.

11.3.5 Deployment Environments

This Chapter proposes offloading of the TeaNet model from the Cloud to the Edge. The Edge node is closest to the source of data and does not require internet connectivity. When hosted in the Edge, the performance of the model is compared to its performance when hosted in the Fog and Cloud environments. With Fog, the computations take place within the resources in the LAN while the Cloud is physically located far from the users, thus making transmission delay large. Therefore, local users within the network of the factory could access the solution in Edge and Fog environments. In contrast, remote users access the solution in the Cloud through the Internet (Fig 11.3).

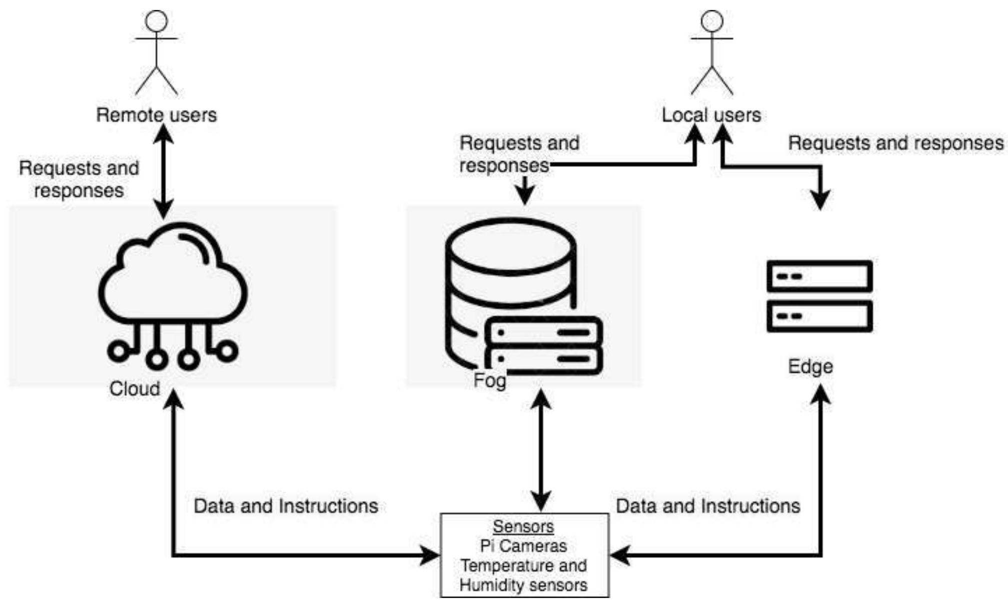


Figure 11.3: Proposed Task offloading model for TeaNet

11.3.6 Evaluation Metrics

After deploying the IoT-based solution, it was necessary to evaluate its performance. The following metrics were used: to measure latency, ten ping measurements were taken from an end-user to Edge, Fog, and Cloud servers, and the minimum, average and maximum round trip time was selected. The Cloud servers considered were from 5 popular Cloud service providers: Alibaba Cloud, Google Cloud, Amazon Web Services (AWS), IBM Cloud, and Microsoft Azure.

A one-way Analysis of variance (ANOVA) [252] and a Post-hoc analysis [253] using Tukey multiple comparisons of means pairwise at 95% confidence interval was done to compare the latency of the environments. These statistical tools were implemented in the R software environment for statistical computing and graphics as it is open-source, platform-independent, and easy to use [254].

Furthermore we used the following metrics during evaluation: accuracy and precision. We discussed these metrics in Chapter 7 Section 7.5.

11.4 Evaluation Results

The proposed solution was evaluated based on latency, energy consumption, accuracy and precision.

11.4.1 Latency

To test the latency of the Edge server, we ran ten pings on the Edge server for 23 iterations. We test the performance of the Edge using Shapiro–Wilk on the minimum, average and maximum values of the Edge server (Figure 11.4). The null hypothesis(H0) is a variable in Shapiro–Wilk test, distributed normally in a population, and its P-value should be greater than 0.05 for a confidence level of 95%. From the results in 11.4(a), the P-value for the minimum latency was 0.02338 ms; thus, we rejected the null hypothesis since the P-value was less than 0.05. We conclude that the data is not normally distributed. The P-value for the average latency was 0.478, which is greater than 0.05; thus, we accept the H0 and conclude that the latency of the mean values is distributed normally. The average of the mean value was 28.22024 ms(Fig 11.4, (b)). The latency of the maximum latency values is illustrated in Fig 11.4 (c). The P-value was 0.983, which was greater than 0.05; thus, the distribution was normal. At the 95% confidence level, the upper limit of the minimum latency was 22.27891 ms, the mean was 21.9344, and its lower limit was 21.58997ms.

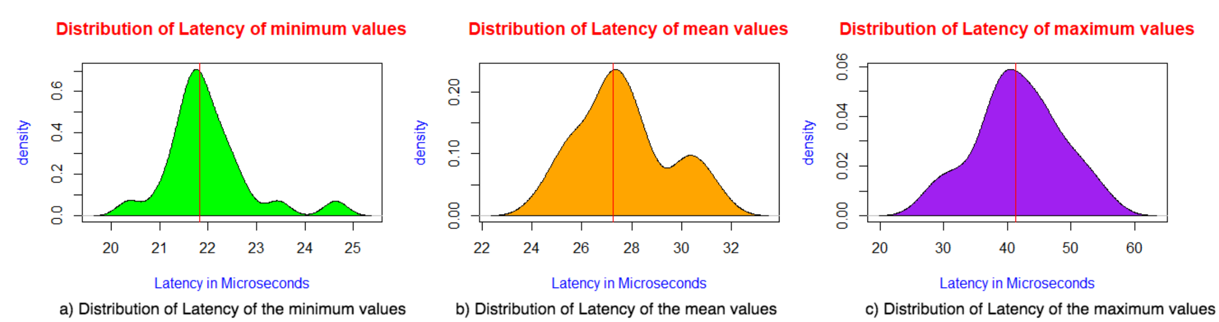


Figure 11.4: The system architecture of the task offloading model for IoT

Fig 11.5 shows the mean latency in microseconds of the proposed Edge model in comparison with that of the five popular cloud service providers and the Fog. The Edge had the least mean latency, followed by the Fog environment. Microsoft Azure, AWS, Google Cloud, and IBM cloud had almost similar mean latency. Alibaba Cloud had the highest latency mean.

Further, results in Fig 11.5 show little within-group variance but high among-group variance. The results show that there was no much difference in terms of the latency recorded by the cloud service providers except Alibaba Cloud, which recorded high latency values. This is because it did not have a data center in Africa like the rest of the cloud service providers, which had data centers in Cape Town by the time of performing these experiments.

Further, a one-way ANOVA between AWS, Alibaba Cloud, Microsoft Azure, Google Cloud, IBM Cloud, Edge, and Fog at 95% confidence level was conducted to compare the means of latency time (Table 11.1). Results indicate evidence of a statistically significant difference in the mean latency in the different environments as determined by one-way ANOVA ($F(6,63)=709.4$ and p-value $< 2.2e-16$).

Table 11.1: ANOVA of latencies of the environments

type	Estimate	Std Error	t-value	Pr (>t)
(Intercept)	150.838	4.495	33.557	< 2.2e-16
Alibaba Cloud	227.793	6.357	35.834	< 2.2e-16
Microsoft Azure	17.361	6.357	2.731	0.00818
Google Cloud	17.555	6.357	2.762	0.00753
IBM Cloud	137.315	6.357	21.601	< 2.2e-16
Edge	-125.747	6.357	-19.781	< 2.2e-16
Fog	-69.492	6.357	-10.932	3.45e-16

Results in Table 11.1 show that there were significant differences in the mean latency of the environments but do not tell where the difference lie. Therefore, a Post-hoc analysis using Tukey multiple comparisons of means pairwise at 95% confidence interval was done (Table 11.2). The analysis indicated that the mean latency time for Alibaba Cloud, IBM Cloud, Edge, and Fog was significantly different from AWS, Microsoft Azure, and Google Cloud. However, AWS did not significantly differ in mean latency time from both Microsoft

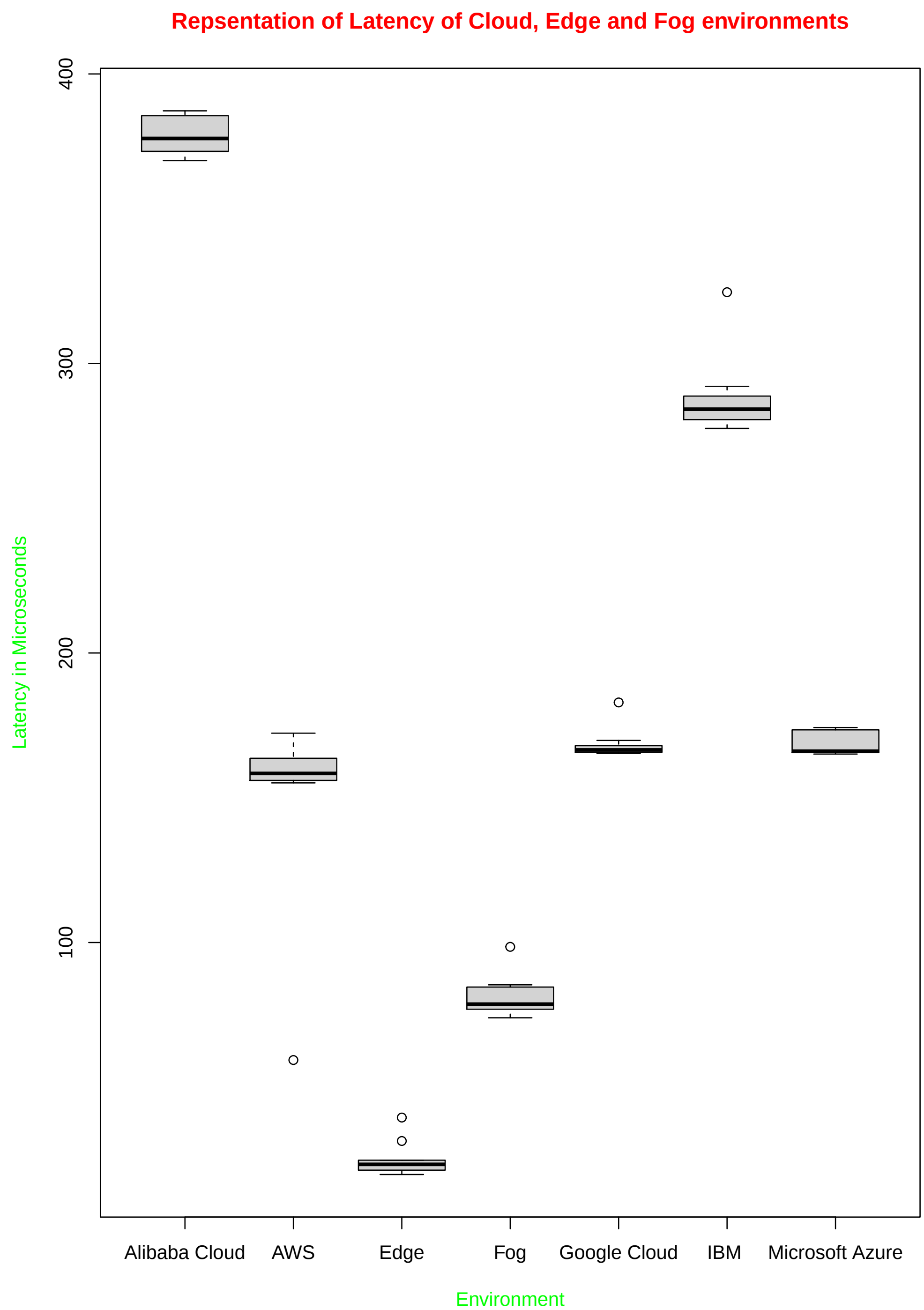


Figure 11.5: A box plot comparison of latency recorded by the environments

Azure and Google Cloud. The confidence intervals (CIs) for Microsoft Azure-AWS, Google cloud-AWS, and Google cloud-Microsoft Azure (-1.999303,36.7215; -1.805603,369152 and

-19.166703,19.5541, respectively) overlap zero implying a non-significant difference. These results suggest that location affects the latency time. However, the location of the Edge gave the least latency time. These results are in agreement with the findings from [255].

Table 11.2: Post-hoc analysis comparison of means the environments

type	diff	lower	Upper	P adj
Alibaba-AWS	227.7927	208.4323	247.1531	0.0000
Microsoft-AWS	17.3611	-1.9993	36.7215	0.1073
Google-AWS	17.5548	-1.8056	36.9152	0.1001
IBM-AWS	137.3152	117.9548	156.6756	0.0000
Edge-AWS	-125.7469	-145.1073	-106.3865	0.0000
Fog-AWS	-69.4920	-88.8524	-50.1316	0.0000
Microsoft-Alibaba	-210.4316	-229.7920	-191.0712	0.0000
Google-Alibaba	-210.2379	-229.5983	-190.8775	0.0000
IBM-Alibaba	-90.4775	-109.8379	-71.1171	0.0000
Edge-Alibaba	-353.5396	-372.9000	-334.1792	0.0000
Fog-Alibaba	-297.2487	-316.6451	-277.9243	0.0000
Google-Microsoft	0.1937	-19.1667	19.5541	1.0000
IBM-Microsoft	119.9541	100.5937	139.3145	0.0000
Edge-Microsoft	-143.1080	-162.4684	-123.7476	0.0000
Fog-Microsoft	-86.8531	-106.2135	-67.4927	0.0000
IBM-Google	119.7604	100.3999	139.1208	0.0000
Edge-Google	-143.3017	-162.6621	-123.9413	0.0000
Fog-Google	-87.0468	-106.4072	-67.6864	0.0000
Edge-IBM	-263.0621	-282.4225	-243.7017	0.0000
Fog-IBM	-206.8072	-226.1676	-187.4468	0.00000
Fog-Edge	56.2549	36.8945	75.6153	0.0000

11.4.2 Accuracy and Precision

The TeaNet model was deployed in Sisibo tea factory between 10th May 2021 and 16 th May 2021. Experiments were done on the model to test its performance in terms of accuracy and precision when powered by the Grid and when powered by the PV-based energy source(Fig 11.6).

A t-test paired comparison of the mean latency at a confidence level of 95% was conducted on precision and accuracy results. For the precision, the lower limit was -0.153393, and the upper limit was 0.0084250. From these results, the confidence interval included zero; thus, there was no significant difference in the precision recorded by the model when powered by PV and

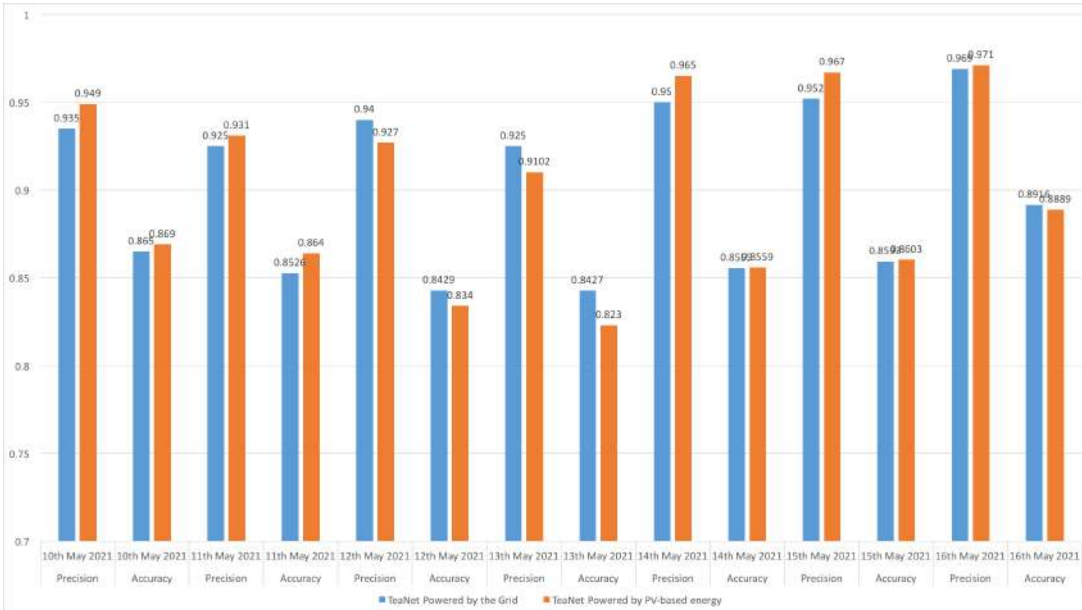


Figure 11.6: Average accuracy and precision of TeaNet when powered with Grid and PV-based energy

Grid. Additionally, for the accuracy, the confidence interval was (-0.019084, 0.0110512), which also included zero, implying that there was no significant difference in the model's accuracy when powered by the Grid and PV. Therefore the improved precision and accuracy of the model when powered by the PV could be because the bulb stationed at the top of the fermentation bed gave good lighting that enabled the model to distinguish the classes of tea images effectively. This experiment was performed to ensure that no hidden effects occur, like power outages or unstable energy supply, which might potentially influence the system

11.4.3 Energy Consumption

TAs discussed in Section 11.3.4, RP*i* power adapter supplies 5.0V; therefore, the total energy (in W-h) consumed in these modes is: Active mode (When RP*i* is capturing, analyzing, and sending data):

$$Energy = V \times I \times t \tag{11.8}$$

Where V is the Voltage, t is the time in hours and I is the current. Thus energy consumed by the RP*i* is given by equation 11.9:

$$Energy\_rpi = 5v \times 0.6A \times 1hour = 3Wh \tag{11.9}$$

The energy consumed by the Pi camera is given by equation 11.10:

$$Energy\_camera = 5v \times 0.25A \times \frac{20}{60} = 0.4167Wh \tag{11.10}$$

The energy consumed by the DHT11 is given by equation 11.11:

$$Energy\_DHT11 = 5v \times 0.016A \times \frac{20}{60} = 0.0267Wh \tag{11.11}$$

Therefore the total energy consumed when the model is optimised is given by Equation 11.12:

$$Total\_Energy = 3wh + 0.4167wh + 0.0267wh = 3.44Wh \tag{11.12}$$

The total amount of energy consumed when the RP*i*, Pi camera, and DHT11 were active was as follows:

Thus energy consumed by the RP*i* is given by Equation 11.13:

$$Energy = 5 \times 0.6A \times 1hour = 3Wh \tag{11.13}$$

The energy consumed by the Pi camera is given by Equation 11.14:

$$Energy = 5 \times 0.25A \times 1hour = 1.25Wh \tag{11.14}$$

The energy consumed by the DHT11 is given by Equation 11.15:

$$Energy = 5 \times 0.016A \times 1hour = 0.08Wh \tag{11.15}$$

Therefore the total energy consumed without optimisation is given by equation 11.16:

$$Total\_Energy = 3wh + 1.25wh + 0.08wh = 4.33Wh \tag{11.16}$$

Thus the total power saved during optimization is given by equation 11.17.

$$Total\_Energy\_saved = 4.33wh - 3.44Wh = 0.89Wh \tag{11.17}$$

After energy consumption optimization, TeaNet was deployed in the Sisibo tea factory between 10th May 2021 and 16th May 2021. Figure 11.7 shows the amount of energy consumed during the period. The duration of the fermentation cycles of tea is dependent on the amount of tea to undergo fermentation. The energy saved during the days depended on the number of hours that the model was operational. With energy optimization, by having the components sleep when they were idle, the total amount of energy saved was 50.6559 Wh which was sufficient to power the optimized model for 14 hours, 43 minutes, and 32 seconds. These results confirm that the energy harvesting and energy optimization were done efficiently.

11.4.4 Discussion of the Results

This study offloaded a deep learning from the cloud to Edge. Furthermore we harvesting energy for use by the model and also optimise code so as to reduce on the energy consumption. From the results, energy harvesting is a non-trivial process. In our case we first had to consider the longest time that fermentation could take place in the factory as this dictates the amount of energy that will be required by the model. Furthermore, the solar radiation of the area was an important consideration and thus we took the worst solar radiation of 5 hours which is ex-

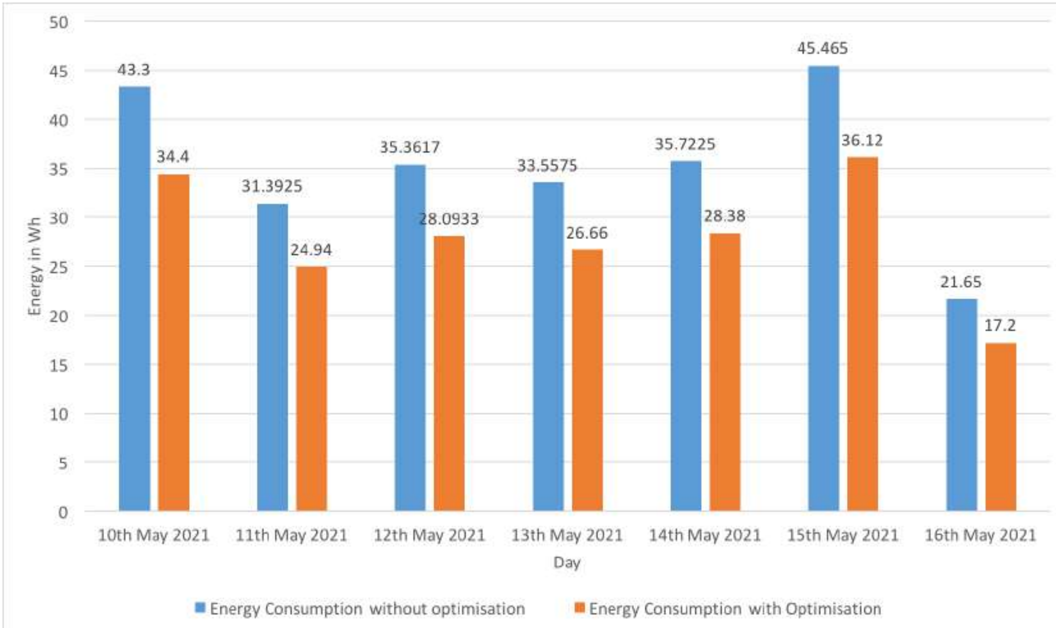


Figure 11.7: Energy consumption of the model between 10$^{\text{th}}$ May 2021 and 16$^{\text{th}}$ May 2021

perienced in August. Since Eldoret is near the equator, we ignored the solar mismatch factor although its an important consideration in some areas. Additionally, optimising code to ensure that components sleep during their idle time saved a lot of energy. Additionally, the results confirmed that the solar modelling was done efficiently and that the model was not compromised in the end as it still record good performance. Thus, PV-based energy harvesting is efficient for powering devices. However, it is essential to measure the required power, optimize code, and consider the battery's efficiency, mismatch factor, peak solar radiation, and discharge levels of the battery.

From this chapter, we found out that Cloud service providers prefer similar locations to deploy their data centers. Four of the five Cloud service providers had a data center in Cape Town, South Africa, serving the continent of Africa. Similarly, in Asia are Tokyo, Singapore, Seoul, and Hong Kong, while in European countries London, Paris, Amsterdam, and Frankfurt while America served by data centers in Texas and Virginia. This could be because the customers they target are the same and that some of them share hosting infrastructure so as to reduce on the cost of operation. Therefore, end-users in Africa would experience minor enhancement in latency when they use multiple service providers since their data centers are all in similar locations; Cape Town, South Africa.

End-users in Africa need to consider deploying their real-time applications in the Edge environment as the Cloud platforms have high latency, which is undesirable for hosting real-time applications.

11.5 Conclusion and Recommendations

This study offloaded an IoT-based tea fermentation monitoring model to the Edge. These results confirm that edge servers can provide lower latency compared to cloud and fog environments. Additionally, the results demonstrate that Edge can be used in developing countries with intermittent internet connections and scarce resources. Further, TeaNet recorded good accuracy and precision results when powered by the PV-based energy source. Thus the PV-based energy can be applied in powering systems in developing countries where the majority of the users are not connected to the grid or have intermittent power supply.

Chapter 12

General conclusion

This chapter gives the conclusions and the recommendations of the study.

12.1 Conclusions

The main aim of this study was to develop and deploy an IoT based tea fermentation monitoring model based on machine learning and image processing techniques for improved quality of black tea. This objective was progressively achieved as follows. In our first study, we analyzed existing proposals for monitoring tea fermentation and found out that most of them were in the form of simulation models. Furthermore, it was evident that the most of the studies adopted color change to detect optimum fermentation levels of tea due to their ease of distinction. In our second study, we collected tea fermentation images for use in training, validating and evaluation of Machine learning models. After data collection, we performed feasibility study on the major machine learning models where a deep learner that we proposed outperformed the rest of the models. The deep learner was then implemented in real-deployment where it showed good results by correlating well with the expert decision when majority voting techniques were used. However, during deployment, we encountered power outages and many of the tea factories are not connected to the grid, thus we modeled a Photo-voltaic energy source to power the solution. The developed model's predictions correlated well with the experts' decisions and confirmed that machine learning and the IoT are viable in monitoring optimum fermentation levels of tea and can be applied in other areas including in monitoring fermentation of other crops including cocoa and coffee since fermentation process is similar for more crops as they experience a change in color. Furthermore, it confirms that PV-based energy source are appropriate when modeled correctly. Thus this study is significant in the field as it explores the deployment of a deep learner with IoT and registered a lot of success. The practicability of applying PV-based energy source in areas which are not connected to the grid or areas experiencing intermittent connection is a game changer and outlines how these solutions can be delivered in developing countries which are not connected to the grid. Furthermore, offloading of tasks from the cloud to the edge resulted in the reduction in latency and thus it is practical to run time-critical applications in Africa when they are offloaded to the edge.

12.2 Limitations of the Study

The following are the limitations of this study:

- The proposed model may not be transferred for use in the detection of fermentation of tea in other regions without transfer learning because the clones of tea grown differs in how they ferment.
- The energy harvesting reported in this study was done in Eldoret region,Kenya, near the equator thus solar mismatch factor was ignore. The same cannot be said of other regions which are not near the equator

12.3 Recommendations

This study makes the following recommendations:

- The developed solution be applied in the monitoring of the optimum fermentation levels of tea while it is offloaded to the edge for reduced latency and costs. This will result in the improvement of the quality of the made tea and consequently their prices and thus economies.
- The developed deep learner (TeaNet) be applied to other comparable tasks by adopting instance-based transfer learning on it to tune its parameters. The model is general enough

as it was trained using two distinct datasets and it did not suffer from overfitting. This will ensure that there is accelerated deployments of solutions.

- The use of Photo-voltaic energy sources in developing countries be explored. This will ensure that the internet of things components are autonomous and also that the developing countries benefits from technological advancements even as they experience electricity connection challenges.
- This study recommends that for reduced latency, tasks be offloaded from the cloud to the edge in African environments. This is because cloud data centers were located in our locality serving all african countries by the time of performing the experiments in this study.
- Last but not least, we recommend that future research work be on the design of an energy harvesting circuit for use in powering the ultra-low power IoT devices.

References

[1] J. D. B. Gil, P. Reidsma, K. Giller, L. Todman, A. Whitmore, and M. van Ittersum, "Sustainable development goal 2: Improved targets and indicators for agriculture and food security," *Ambio*, vol. 48, no. 7, pp. 685–698, Jul. 2019, ISSN: 16547209. DOI: 10.1007/S13280-018-1101-4/TABLES/3. [Online]. Available: https://link.springer.com/article/10.1007/s13280-018-1101-4.

[2] P. Caron, G. Ferrero y de Loma-Osorio, D. Nabarro, *et al.*, "Food systems for sustainable development: proposals for a profound four-part transformation," *Agronomy for Sustainable Development*, vol. 38, no. 4, pp. 1–12, Aug. 2018, ISSN: 17730155. DOI: 10.1007/S13593-018-0519-1/FIGURES/2. [Online]. Available: https://link.springer.com/article/10.1007/s13593-018-0519-1.

[3] W. Lutz and S. Kc, "Dimensions of global population projections: what do we know about future population trends and structures?" *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 365, no. 1554, p. 2779, Sep. 2010, ISSN: 14712970. DOI: 10.1098/RSTB.2010.0133. [Online]. Available: /pmc/articles/PMC2935115/ /pmc/articles/PMC2935115/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC2935115/.

[4] L. Christiaensen, L. Demery, and J. Kuhl, "The Role of Agriculture in Poverty Reduction : An Empirical Perspective," Policy Research Working Papers, Sep. 2006. DOI: 10.1596/1813-9450-4013. [Online]. Available: https://openknowledge.worldbank.org/handle/10986/9306.

[5] Y. R. Waarts, V. Janssen, R. Aryeetey, *et al.*, "Multiple pathways towards achieving a living income for different types of smallholder tree-crop commodity farmers," *Food Security*, vol. 13, no. 6, pp. 1467–1496, Dec. 2021, ISSN: 18764525. DOI: 10.1007/S12571-021-01220-5/FIGURES/11. [Online]. Available: https://link.springer.com/article/10.1007/s12571-021-01220-5.

[6] J. Adriano Cruz, "Enhancement of growth and yield of upland rice (Oryza sativa L.) by Actinomycetes," *Agrotechnology*, vol. s1, no. 01, p. 4, 2014, ISSN: 1686-9141. DOI: 10.4172/2168-9881.s1.008. [Online]. Available: http://dx.doi.org/10.4172/2168-9881.S1.008.

[7] Z. Mitei, "Growing sustainable tea on Kenyan smallholder farms," *International Journal of Agricultural Sustainability*, vol. 9, no. 1, pp. 59–66, 2011, ISSN: 1473-5903. DOI: 10.3763/ijas.2010.0550. [Online]. Available: https://www.tandfonline.com/action/journalInformation?journalCode=tags20.

[8] M. Eichsteller, T. Njagi, and E. Nyukuri, "The role of agriculture in poverty escapes in Kenya – Developing a capabilities approach in the context of climate change," *World*

Development, vol. 149, p. 105 705, Jan. 2022, ISSN: 0305-750X. DOI: 10.1016/J.WORLDDEV.2021.105705.

[9] G. Sigley, "Tea and China's rise: tea, nationalism and culture in the 21st century," *International Communication of Chinese Culture*, vol. 2, no. 3, pp. 319–341, Dec. 2015, ISSN: 2197-4233. DOI: 10.1007/s40636-015-0037-7.

[10] E. Yiannakopoulou, "Green Tea Catechins: Proposed Mechanisms of Action in Breast Cancer Focusing on the Interplay Between Survival and Apoptosis," *Anti-Cancer Agents in Medicinal Chemistry*, vol. 14, no. 2, pp. 290–295, Feb. 2014, ISSN: 18715206. DOI: 10.2174/18715206113136660339. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/24069935.

[11] K. Miura, M. C. B. Hughes, N. I. Arovah, J. C. Van Der Pols, and A. C. Green, "Black Tea Consumption and Risk of Skin Cancer: An 11-Year Prospective Study," *Nutrition and Cancer*, vol. 67, no. 7, pp. 1049–1055, Oct. 2015, ISSN: 15327914. DOI: 10.1080/01635581.2015.1073759.

[12] F. Hajiaghaalipour, M. S. Kanthimathi, J. Sanusi, and J. Rajarajeswaran, "White tea (Camellia sinensis) inhibits proliferation of the colon cancer cell line, HT-29, activates caspases and protects DNA of normal cells against oxidative damage," *Food Chemistry*, vol. 169, pp. 401–410, Feb. 2015, ISSN: 18737072. DOI: 10.1016/j.foodchem.2014.07.005.

[13] E. Sironi, L. Colombo, A. Lompo, *et al.*, "Natural compounds against neurodegenerative diseases: Molecular characterization of the interaction of catechins from green tea with Aβ1-42, PrP106-126, and ataxin-3 oligomers," *Chemistry - A European Journal*, vol. 20, no. 42, pp. 13 793–13 800, Oct. 2014, ISSN: 15213765. DOI: 10.1002/chem.201403188.

[14] J. B. Lazaro, A. Ballado, F. P. F. Bautista, J. K. B. So, and J. M. J. Villegas, "Chemometric data analysis for black tea fermentation using principal component analysis," in *AIP Conference Proceedings*, vol. 2045, AIP Publishing LLC, Dec. 2018, p. 020 050. DOI: 10.1063/1.5080863. [Online]. Available: http://aip.scitation.org/doi/abs/10.1063/1.5080863.

[15] D. Saikia, P. K. Boruah, and U. Sarma, "A Sensor Network to Monitor Process Parameters of Fermentation and Drying in Black Tea Production," *MAPAN*, vol. 30, no. 3, pp. 211–219, Sep. 2015, ISSN: 0970-3950. DOI: 10.1007/s12647-015-0142-4. [Online]. Available: http://link.springer.com/10.1007/s12647-015-0142-4.

[16] D. D. Onduru, A De Jager, S Hiller, and R Van Den Bosch, "Sustainability of smallholder tea production in developing countries: Learning experiences from farmer field schools in Kenya," Tech. Rep. 3, 2012, pp. 714–742. [Online]. Available: www.isdsnet.com/ijds.

[17] R. Zhi, L. Zhao, and D. Zhang, "A framework for the multi-level fusion of electronic nose and electronic tongue for tea quality assessment," *Sensors (Switzerland)*, vol. 17, no. 5, May 2017, ISSN: 14248220. DOI: 10.3390/s17051007.

[18] A. Akuli, A. Pal, G. Bej, *et al.*, "A Machine Vision System for Estimation of Theaflavins and Thearubigins in orthodox black tea," *International Journal on Smart Sensing and Intelligent Systems*, vol. 9, no. 2, pp. 709–731, Jun. 2016. DOI: 10.21307/ijssis-2017-891. [Online]. Available: https://www.exeley.com/in_jour_smart_sensing_and_intelligent_systems/doi/10.21307/ijssis-2017-891.

[19] S. Okamura, Y. Zhang, and N. Tsukamoto, *A new microstripline-type moisture sensor for heavily wet tea leaves*, Apr. 2007. DOI: 10.1088/0957-0233/18/4/009. [Online]. Available: http://stacks.iop.org/0957-0233/18/i=4/a=009?key=crossref.344b0debb89b83838183329495659502.

[20] B. Maiti, R. Mahobia, D. K. Kushwaha, and R. Soni, "A novel technique for withering of tea leaf," in *AIP Conference Proceedings*, vol. 1851, American Institute of Physics Inc., Jun. 2017, ISBN: 9780735415256. DOI: 10.1063/1.4984654.

[21] S. Ghosh, B. Tudu, N. Bhattacharyya, and R. Bandyopadhyay, "A recurrent Elman network in conjunction with an electronic nose for fast prediction of optimum fermentation time of black tea," *Neural Computing and Applications*, vol. 31, pp. 1165–1171, Feb. 2019, ISSN: 09410643. DOI: 10.1007/s00521-017-3072-y.

[22] M. Xu, J. Wang, and S. Gu, "Rapid identification of tea quality by E-nose and computer vision combining with a synergetic data fusion strategy," *Journal of Food Engineering*, vol. 241, pp. 10–17, Jan. 2019, ISSN: 02608774. DOI: 10.1016/j.jfoodeng.2018.07.020.

[23] R. Jayabalan, R. V. Malbaša, E. S. Lončar, J. S. Vitas, and M. Sathishkumar, "A Review on Kombucha Tea-Microbiology, Composition, Fermentation, Beneficial Effects, Toxicity, and Tea Fungus," *Comprehensive Reviews in Food Science and Food Safety*, vol. 13, no. 4, pp. 538–550, Jul. 2014, ISSN: 15414337. DOI: 10.1111/1541-4337.12073. [Online]. Available: http://doi.wiley.com/10.1111/1541-4337.12073.

[24] K. Jolvis Pou, "Fermentation: The Key Step in the Processing of Black Tea," *Journal of Biosystems Engineering*, vol. 41, no. 2, pp. 85–92, Jun. 2016, ISSN: 1738-1266. DOI: 10.5307/JBE.2016.41.2.085. [Online]. Available: http://koreascience.or.kr/journal/view.jsp?kj=NOGGB5&py=2016&vnc=v41n2&sp=85.

[25] M. H. Asil, B. Rabiei, and R. Ansari, *Optimal fermentation time and temperature to improve biochemical composition and sensory characteristics of black tea.* 2012. [Online]. Available: https://www.semanticscholar.org/paper/Optimal-fermentation-time-and-temperature-to-and-of-Asil-Rabiei/.

[26] S. Borah and M. Bhuyan, "Non-destructive testing of tea fermentation using image processing," *Insight: Non-Destructive Testing and Condition Monitoring*, vol. 45, no. 1, pp. 55–58, Jan. 2003, ISSN: 13542575. DOI: 10.1784/insi.45.1.55.52593.

[27] P. P. Shinde and S. Shah, "A Review of Machine Learning and Deep Learning Applications," in *Proceedings - 2018 4th International Conference on Computing, Communication Control and Automation, ICCUBEA 2018*, Institute of Electrical and Electronics Engineers Inc., Jul. 2018, ISBN: 9781538652572. DOI: 10.1109/ICCUBEA.2018.8697857.

[28] M. N. S. Miazi, Z. Erasmus, M. A. Razzaque, M. Zennaro, and A. Bagula, "Enabling the Internet of Things in developing countries: Opportunities and challenges," in *2016 5th International Conference on Informatics, Electronics and Vision (ICIEV)*, IEEE, May 2016, pp. 564–569, ISBN: 978-1-5090-1269-5. DOI: 10.1109/ICIEV.2016.7760066. [Online]. Available: http://ieeexplore.ieee.org/document/7760066/.

[29] A. Khanna and S. Kaur, "Evolution of Internet of Things (IoT) and its significant impact in the field of Precision Agriculture," *Computers and Electronics in Agriculture*, vol. 157, pp. 218–231, Feb. 2019, ISSN: 0168-1699. DOI: 10.1016/J.COMPAG.2018.12.039. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168169918316417.

[30] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of Things: A Survey on Enabling Technologies, Protocols, and Applications," *IEEE Communications Surveys and Tutorials*, vol. 17, no. 4, pp. 2347–2376, Oct. 2015. DOI: 10.1109/COMST.2015.2444095.

[31] V. Hassija, V. Chamola, V. Saxena, D. Jain, P. Goyal, and B. Sikdar, *A Survey on IoT Security: Application Areas, Security Threats, and Solution Architectures*, 2019. DOI: 10.1109/ACCESS.2019.2924045.

[32] D. F. Josué, D. F. Josué, F. M. Guy-De-Patience, K. Bienvenu, and D. Bonaventure, "Design and development of cocoa pod breaking and beans extraction machine," *International Journal of Engineering & Technology*, vol. 8, no. 3, pp. 357–366, Sep. 2019. DOI: 10.14419/ijet.v8i3.29548. [Online]. Available: https://www.sciencepubco.com/index.php/ijet/article/view/29548.

[33] M. Elsisi, M. Q. Tran, K. Mahmoud, M. Lehtonen, and M. M. Darwish, "Deep Learning-Based Industry 4.0 and Internet of Things towards Effective Energy Management for Smart Buildings," *Sensors 2021, Vol. 21, Page 1038*, vol. 21, no. 4, p. 1038, Feb. 2021, ISSN: 1424-8220. DOI: 10.3390/S21041038. [Online]. Available: https://www.mdpi.com/1424-8220/21/4/1038/htmhttps://www.mdpi.com/1424-8220/21/4/1038.

[34] C. C. Aggarwal, "An Introduction to Neural Networks," *Neural Networks and Deep Learning*, pp. 1–52, 2018. DOI: 10.1007/978-3-319-94463-0{_}1. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-94463-0_1.

[35] F. M. Awaysheh, M. Alazab, S. Garg, D. Niyato, and C. Verikoukis, "Big Data Resource Management Networks: Taxonomy, Survey, and Future Directions," *IEEE Communications Surveys and Tutorials*, vol. 23, no. 4, pp. 2098–2130, 2021, ISSN: 1553877X. DOI: 10.1109/COMST.2021.3094993.

[36] S. Pouyanfar, S. Sadiq, Y. Yan, *et al.*, *A survey on deep learning: Algorithms, techniques, and applications*, Aug. 2018. DOI: 10.1145/3234150. [Online]. Available: http://dl.acm.org/citation.cfm?doid=3271482.3234150.

[37] C. Chen, P. Zhang, H. Zhang, *et al.*, "Deep Learning on Computational-Resource-Limited Platforms: A Survey," *Mobile Information Systems*, vol. 2020, 2020, ISSN: 1875905X. DOI: 10.1155/2020/8454327.

[38] I. H. Sarker, "Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions," *SN Computer Science 2021 2:6*, vol. 2, no. 6, pp. 1–20, Aug. 2021, ISSN: 2661-8907. DOI: 10.1007/S42979-021-00815-1. [Online]. Available: https://link.springer.com/article/10.1007/s42979-021-00815-1.

[39] S. Borah and M. Bhuyan, "A computer based system for matching colours during the monitoring of tea fermentation," *International Journal of Food Science and Technology*, vol. 40, no. 6, pp. 675–682, Jun. 2005, ISSN: 0950-5423. DOI: 10.1111/j.1365-2621.2005.00981.x. [Online]. Available: http://doi.wiley.com/10.1111/j.1365-2621.2005.00981.x.

[40] S. M. Kamunya, F. N. Wachira, R. S. Pathak, R. C. Muoki, and R. K. Sharma, "Tea Improvement in Kenya," in *Advanced Topics in Science and Technology in China*, Springer, 2012, pp. 177–226. DOI: 10.1007/978-3-642-31878-8{_}5.

[41] L. Lvova, "Electronic Tongue Principles and Applications in the Food Industry," in *Electronic Noses and Tongues in Food Science*, Elsevier Inc., Jan. 2016, pp. 151–160, ISBN: 9780128004029. DOI: 10.1016/B978-0-12-800243-8.00015-9.

[42] Y. h. Zhong, S. Zhang, R. He, *et al.*, "A Convolutional Neural Network Based Auto Features Extraction Method for Tea Classification with Electronic Tongue," *Applied Sciences*, vol. 9, no. 12, p. 2518, Jun. 2019, ISSN: 2076-3417. DOI: 10.3390/app9122518. [Online]. Available: https://www.mdpi.com/2076-3417/9/12/2518.

[43] H. Li, B. Zhang, W. Hu, Y. Liu, C. Dong, and Q. Chen, "Monitoring black tea fermentation using a colorimetric sensor array-based artificial olfaction system," *Journal of Food*

Processing and Preservation, vol. 42, no. 1, e13348, Jan. 2018, ISSN: 01458892. DOI: 10.1111/jfpp.13348. [Online]. Available: http://doi.wiley.com/10.1111/jfpp.13348.

[44] J. Zobel, *Writing for Computer Science*. London: Springer London, 2014, ISBN: 978-1-4471-6638-2. DOI: 10.1007/978-1-4471-6639-9. [Online]. Available: http://link.springer.com/10.1007/978-1-4471-6639-9.

[45] K. Peffers, T. Tuunanen, and B. Niehaves, "Design science research genres: introduction to the special issue on exemplars and criteria for applicable design science research," *European Journal of Information Systems*, vol. 27, no. 2, pp. 129–139, Mar. 2018, ISSN: 0960-085X. DOI: 10.1080/0960085X.2018.1458066. [Online]. Available: https://doi.org/10.1080/0960085X.2018.1458066.

[46] A. R. Hevner, S. T. March, J. Park, and S. Ram, "Design science in information systems research," *MIS Quarterly: Management Information Systems*, vol. 28, no. 1, pp. 75–105, 2004, ISSN: 02767783. DOI: 10.2307/25148625.

[47] P. Johannesson and E. Perjons, *An Introduction to Design Science*, en. Springer International Publishing, Dec. 2014, ISBN: 978-3-319-10631-1. [Online]. Available: file://www.springer.com/gp/book/9783319106311.

[48] T. Sharmilan, I. Premarathne, I. Wanniarachchi, S. Kumari, and D. Wanniarachchi, "Electronic Nose Technologies in Monitoring Black Tea Manufacturing Process," *Journal of Sensors*, vol. 2020, Y. Li, Ed., p. 3 073 104, 2020, ISSN: 1687-725X. DOI: 10.1155/2020/3073104. [Online]. Available: https://doi.org/10.1155/2020/3073104.

[49] S. Deb and K. R. Jolvis Pou, "A Review of Withering in the Processing of Black Tea," *J. of Biosystems Eng*, vol. 41, no. 4, p. 12, 2016, ISSN: 1738-1266. DOI: 10.5307/JBE.2016.41.4.365. [Online]. Available: https://doi.org/10.5307/JBE.2016.41.4.365.

[50] G. S. Gill, A. Kumar, and R. Agarwal, "Monitoring and grading of tea by computer vision – A review," *Journal of Food Engineering*, vol. 106, no. 1, pp. 13–19, Sep. 2011, ISSN: 0260-8774. DOI: 10.1016/J.JFOODENG.2011.04.013. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0260877411002093.

[51] M. K. Meegahakumbura, M. C. Wambulwa, M. M. Li, *et al.*, "Domestication origin and breeding history of the tea plant (Camellia sinensis) in China and India based on nuclear microsatellites and cpDNA sequence data," *Frontiers in Plant Science*, vol. 8, p. 2270, Jan. 2018, ISSN: 1664462X. DOI: 10.3389/FPLS.2017.02270/FULL. [Online]. Available: /pmc/articles/PMC5788969//pmc/articles/PMC5788969/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC5788969/.

[52] F. Ozdemir, I. Tontul, F. Balci-Torun, and A. Topuz, "Effect of rolling methods and storage on volatile constituents of Turkish black tea," *Flavour and Fragrance Journal*, vol. 32, no. 5, pp. 362–375, Sep. 2017, ISSN: 1099-1026. DOI: 10.1002/FFJ.3385. [Online]. Available: https://onlinelibrary.wiley.com/doi/full/10.1002/ffj.3385https://onlinelibrary.wiley.com/doi/abs/10.1002/ffj.3385https://onlinelibrary.wiley.com/doi/10.1002/ffj.3385.

[53] D. Gonçalves Bortolini, C. Windson Isidoro Haminiuk, A. Cristina Pedro, I. de Andrade Arruda Fernandes, and G. Maria Maciel, "Processing, chemical signature and food industry applications of Camellia sinensis teas: An overview," *Food Chemistry: X*, vol. 12, p. 100 160, Dec. 2021, ISSN: 2590-1575. DOI: 10.1016/J.FOCHX.2021.100160.

[54] L. Capelli, S. Sironi, P. Céntola, R. Del Rosso, and M. Il Grande, "Electronic noses for the continuous monitoring of odours from a wastewater treatment plant at specific receptors: Focus on training methods," *Sensors and Actuators, B: Chemical*, vol. 131, no. 1, pp. 53–62, Apr. 2008, ISSN: 09254005. DOI: 10.1016/j.snb.2007.12.004.

[55] A. Catarina Bastos and N. Magan, "Potential of an electronic nose for the early detection and differentiation between Streptomyces in potable water," *Sensors and Actuators, B: Chemical*, vol. 116, no. 1-2, pp. 151–155, Jul. 2006, ISSN: 09254005. DOI: 10.1016/j.snb.2005.11.073.

[56] N. S. Ramgir, "Electronic Nose Based on Nanomaterials: Issues, Challenges, and Prospects," *ISRN Nanomaterials*, vol. 2013, pp. 1–21, 2013, ISSN: 2090-8741. DOI: 10.1155/2013/941581.

[57] K. R. Kashwan and M. Bhuyan, "Robust electronic-nose system with temperature and humidity drift compensation for tea and spice flavour discrimination," *2005 Asian Conference on Sensors and the International Conference on New Techniques in Pharmaceutical and Biomedical Research - Proceedings*, vol. 2005, pp. 154–158, 2005. DOI: 10.1109/ASENSE.2005.1564528.

[58] N. Bhattacharyya, S. Seth, B. Tudu, *et al.*, "Monitoring of black tea fermentation process using electronic nose," *Journal of Food Engineering*, vol. 80, no. 4, pp. 1146–1156, Jun. 2007, ISSN: 02608774. DOI: 10.1016/j.jfoodeng.2006.09.006. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0260877406005917.

[59] B. Tudu, B. Das, N. Bhattacharyya, A. Jana, D. Ghosh, and R. Bandyopadhyay, "Smell peak prediction during black tea fermentation process using time-delay neural network on electronic nose data," in *Proceedings - International Conference on Computing: Theory and Applications, ICCTA 2007*, 2007, pp. 257–260, ISBN: 0769527701. DOI: 10.1109/ICCTA.2007.119.

[60] S. S. Chowdhury, B. Tudu, R. Bandyopadhyay, and N. Bhattacharyya, "Portable electronic nose system for aroma classification of black tea," in *IEEE Region 10 Colloquium*

and 3rd International Conference on Industrial and Information Systems, ICIIS 2008, 2008, ISBN: 9781424428069. DOI: 10.1109/ICIINFS.2008.4798403.

[61] D. G. Mohit Sharma and Nabarun Bhattacharya, "Electronic Nose – A new way for predicting the optimum point of fermentation of Black Tea," *International Journal of Engineering Science Invention*, vol. 12, no. 1, 2012.

[62] S. N. Hidayat, K. Triyana, I. Fauzan, *et al.*, "The Electronic Nose Coupled with Chemometric Tools for Discriminating the Quality of Black Tea Samples In Situ," *Chemosensors 2019, Vol. 7, Page 29*, vol. 7, no. 3, p. 29, Jul. 2019, ISSN: 22279040. DOI: 10.3390/CHEMOSENSORS7030029. [Online]. Available: https://www.mdpi.com/2227-9040/7/3/29/htmhttps://www.mdpi.com/2227-9040/7/3/29.

[63] T. S. Tseng, M. H. Hsiao, P. A. Chen, S. Y. Lin, S. W. Chiu, and D. J. Yao, "Utilization of a Gas-Sensing System to Discriminate Smell and to Monitor Fermentation during the Manufacture of Oolong Tea Leaves," *Micromachines*, vol. 12, no. 1, pp. 1–11, Jan. 2021, ISSN: 2072666X. DOI: 10.3390/MI12010093. [Online]. Available: /pmc/articles/PMC7829750//pmc/articles/PMC7829750/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC7829750/.

[64] R. Glatz and K. Bailey-Hill, *Mimicking nature's noses: From receptor deorphaning to olfactory biosensing*, Feb. 2011. DOI: 10.1016/j.pneurobio.2010.11.004.

[65] M. Peris and L. Escuder-Gilabert, *A 21st century technique for food control: Electronic noses*, Apr. 2009. DOI: 10.1016/j.aca.2009.02.009.

[66] F. Röck, N. Barsan, and U. Weimar, *Electronic nose: Current status and future trends*, Feb. 2008. DOI: 10.1021/cr068121q.

[67] M. Podrazka, E. Báczyńska, M. Kundys, P. S. Jeleń, and E. W. Nery, "Electronic Tongue—A Tool for All Tastes?" *Biosensors*, vol. 8, no. 1, Dec. 2018, ISSN: 20796374. DOI: 10.3390/BIOS8010003. [Online]. Available: /pmc/articles/PMC5872051//pmc/articles/PMC5872051/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC5872051/.

[68] P. Ciosek and W. Wróblewski, *Sensor arrays for liquid sensing - Electronic tongue systems*, 2007. DOI: 10.1039/b705107g.

[69] K. H. Chang, R. L. Chen, B. C. Hsieh, *et al.*, "A hand-held electronic tongue based on fluorometry for taste assessment of tea," *Biosensors and Bioelectronics*, vol. 26, no. 4, pp. 1507–1513, Dec. 2010, ISSN: 09565663. DOI: 10.1016/j.bios.2010.07.100.

[70] S. Ai, R. Wu, L. Yan, and Y. Wu, "Evaluation of taste quality in green tea infusion using electronic tongue combined with LS-SVM," in *Advanced Materials Research*, vol. 301-

303, 2011, pp. 1643–1647, ISBN: 9783037851975. DOI: 10.4028/www.scientific.net/AMR.301-303.1643.

[71] A. Ghosh, A. K. Bag, P. Sharma, *et al.*, "Monitoring the fermentation process and detection of optimum fermentation time of black tea using an electronic tongue," *IEEE Sensors Journal*, vol. 15, no. 11, pp. 6255–6262, Nov. 2015, ISSN: 1530437X. DOI: 10.1109/JSEN.2015.2455535.

[72] P. Saha, S. Ghorai, B. Tudu, R. Bandyopadhyay, and N. Bhattacharyya, "WHT-based tea quality prediction using electronic tongue signals," *Computer, Communication and Electrical Technology - Proceedings of the International Conference on Advancement of Computer Communication and Electrical Technology, ACCET 2016*, pp. 267–272, 2017. DOI: 10.1201/9781315400624-52.

[73] S. Buratti, A. Casiraghi, P. Minghetti, and G. Giovanelli, "The Joint Use of Electronic Nose and Electronic Tongue for the Evaluation of the Sensorial Properties of Green and Black Tea Infusions as Related to Their Chemical Composition," *Food and Nutrition Sciences*, vol. 04, no. 06, pp. 605–615, 2013, ISSN: 2157-944X. DOI: 10.4236/fns.2013.46078.

[74] P. Saha, S. Ghorai, B. Tudu, R. Bandyopadhyay, and N. Bhattacharyya, "SVD based tea quality prediction using electronic tongue signal," *2016 International Conference on Intelligent Control, Power and Instrumentation, ICICPI 2016*, pp. 79–83, Feb. 2017. DOI: 10.1109/ICICPI.2016.7859678.

[75] S Saranka, K. Thangathurai, C Wanniarachchi, and W. K. Wanniarachchi, "Monitoring Fermentation of Black Tea with Image Processing Techniques," *IPSL*, vol. 32, Dec. 2016.

[76] A. Akuli, A. Pal, T. Dey, *et al.*, "Assessment of Black Tea Using Low-Level Image Feature Extraction Technique," in Springer, Singapore, 2020, pp. 453–467. DOI: 10.1007/978-981-15-2188-1{_}36.

[77] C. Dong, G. Liang, B. Hu, *et al.*, "Prediction of Congou Black Tea Fermentation Quality Indices from Color Features Using Non-Linear Regression Methods," *Scientific Reports*, vol. 8, no. 1, p. 10 535, Dec. 2018, ISSN: 2045-2322. DOI: 10.1038/s41598-018-28767-2. [Online]. Available: http://www.nature.com/articles/s41598-018-28767-2.

[78] P. T. Binh, D. H. Du, and T. C. Nhung, "Control and Optimize Black Tea Fermentation Using Computer Vision and Optimal Control Algorithm," in *Lecture Notes in Networks and Systems*, vol. 104, Springer, Dec. 2020, pp. 310–319. DOI: 10.1007/978-3-030-37497-6{_}36.

[79] G. Jin, Y. j. Wang, M. Li, *et al.*, "Rapid and real-time detection of black tea fermentation quality by using an inexpensive data fusion system," *Food Chemistry*, vol. 358, p. 129 815, Oct. 2021, ISSN: 0308-8146. DOI: 10.1016/J.FOODCHEM.2021.129815.

[80] Z. Yang, J. Tang, L. Xue, and Y. Peng, "Study on High Efficiency Black Tea Fermentation Control System Based on Multiple Regression," *ACM International Conference Proceeding Series*, May 2021. DOI: 10.1145/3469213.3471321.

[81] Y. Zou, G. Qi, S. Chen, L. Tan, W. Li, and T. Liu, "A simple method based on image processing to estimate primary pigment levels of Sichuan Dark Tea during post-fermentation," *European Food Research and Technology*, vol. 239, no. 2, pp. 357–363, Jun. 2014, ISSN: 14382385. DOI: 10.1007/S00217-014-2219-7/FIGURES/5. [Online]. Available: https://link.springer.com/article/10.1007/s00217-014-2219-7.

[82] T Baranwal, Nitika, and P. K. Pateriya, "Development of IoT based smart security and monitoring devices for agriculture," in *2016 6th International Conference - Cloud System and Big Data Engineering (Confluence)*, 2016, pp. 597–602. DOI: 10.1109/CONFLUENCE.2016.7508189. [Online]. Available: https://ieeexplore.ieee.org/document/7508189.

[83] P. Sethi and S. R. Sarangi, "Internet of Things: Architectures, Protocols, and Applications," *Journal of Electrical and Computer Engineering*, vol. 2017, R. Khanna, Ed., p. 9 324 035, 2017, ISSN: 2090-0147. DOI: 10.1155/2017/9324035. [Online]. Available: https://doi.org/10.1155/2017/9324035.

[84] D. Saikia, P. Boruah, and U. Sarma, "Development and implementation of a sensor network to monitor fermentation process parameter in tea processing," *International Journal on Smart Sensing and Intelligent Systems*, vol. 7, no. 3, pp. 1254–1270, Sep. 2014, ISSN: 1178-5608. DOI: 10.21307/ijssis-2017-704. [Online]. Available: https://www.exeley.com/in_jour_smart_sensing_and_intelligent_systems/doi/10.21307/ijssis-2017-704.

[85] Y. Uehara and S. Ohtake, "Factory Environment Monitoring: A Japanese Tea Manufacturer's Case," in *2019 IEEE International Conference on Consumer Electronics (ICCE)*, IEEE, Jan. 2019, pp. 1–3, ISBN: 978-1-5386-7910-4. DOI: 10.1109/ICCE.2019.8661967. [Online]. Available: https://ieeexplore.ieee.org/document/8661967/.

[86] Y. Kumar, A. Aggarwal, S. Tiwari, and K. Singh, "An efficient and robust approach for biomedical image retrieval using Zernike moments," *Biomedical Signal Processing and Control*, vol. 39, pp. 459–473, Jan. 2018, ISSN: 1746-8094. DOI: 10.1016/j.bspc.2017.08.018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S174680941730191Xhttps://www.sciencedirect.com/science/article/abs/pii/S174680941730191X.

[87] H. CO and G. W. Sanderson, “Biochemistry of tea Fermentation: conversion of amino acids to Black tea aroma constituents,” *Journal of Food Science*, vol. 35, no. 2, pp. 160–164, Mar. 1970, ISSN: 0022-1147. DOI: 10.1111/j.1365-2621.1970.tb12128.x. [Online]. Available: http://doi.wiley.com/10.1111/j.1365-2621.1970.tb12128.x.

[88] J. B. Lazaro, A. Ballado, F. P. Bautista, J. K. So, and J. M. Villegas, “Chemometric data analysis for black tea fermentation using principal component analysis,” *AIP Conference Proceedings*, vol. 2045, no. 1, p. 020 050, Dec. 2018, ISSN: 0094-243X. DOI: 10.1063/1.5080863. [Online]. Available: https://aip.scitation.org/doi/abs/10.1063/1.5080863.

[89] S. V. Patel and V. N. Jokhakar, “A random forest based machine learning approach for mild steel defect diagnosis,” in *2016 IEEE International Conference on Computational Intelligence and Computing Research, ICCIC 2016*, Institute of Electrical and Electronics Engineers Inc., May 2017, ISBN: 9781509006113. DOI: 10.1109/ICCIC.2016.7919549.

[90] M. Al Hamad and A. M. Zeki, “Accuracy vs. cost in decision trees: A survey,” in *2018 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies, 3ICT 2018*, Institute of Electrical and Electronics Engineers Inc., Nov. 2018, ISBN: 9781538692073. DOI: 10.1109/3ICT.2018.8855780.

[91] M. Mahmud, M. S. Kaiser, A. Hussain, and S. Vassanelli, “Applications of Deep Learning and Reinforcement Learning to Biological Data,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 6, pp. 2063–2079, Jun. 2018, ISSN: 21622388. DOI: 10.1109/TNNLS.2018.2790388.

[92] G. Florimbi, H. Fabelo, E. Torti, *et al.*, “Accelerating the K-Nearest neighbors filtering algorithm to optimize the real-time classification of human brain tumor in hyperspectral images,” *Sensors (Switzerland)*, vol. 18, no. 7, Jul. 2018, ISSN: 14248220. DOI: 10.3390/s18072314.

[93] H. Zhang, G. Zou, X. Liu, Y. Xiao, and W. Wang, “Identification of Xinyang Maojian tea taste using electronic tongue,” *Sensors and Materials*, vol. 31, no. 7, pp. 2347–2356, 2019, ISSN: 09144935. DOI: 10.18494/SAM.2019.2244.

[94] S. Karthika and N. Sairam, “A Naïve Bayesian classifier for educational qualification,” *Indian Journal of Science and Technology*, vol. 8, no. 16, 2015, ISSN: 09745645. DOI: 10.17485/ijst/2015/v8i16/62055.

[95] Ji Qin Ni and Z. Hang, “Energy consumption and conservation in tea processing in China,” *Renewable Energy*, vol. 2, no. 1, pp. 1–5, Feb. 1992, ISSN: 09601481. DOI: 10.1016/0960-1481(92)90053-6.

[96] P. Jayaraman, A. Yavari, D. Georgakopoulos, *et al.*, "Internet of Things Platform for Smart Farming: Experiences and Lessons Learnt," *Sensors*, vol. 16, no. 11, p. 1884, Nov. 2016, ISSN: 1424-8220. DOI: 10.3390/s16111884. [Online]. Available: http://www.mdpi.com/1424-8220/16/11/1884.

[97] G. Kimutai, A. Ngenzi, R. Ngoga Said, R. C. Ramkat, and A. Förster, "A Data Descriptor for Black Tea Fermentation Dataset," *Data*, vol. 6, no. 3, p. 34, Mar. 2021, ISSN: 2306-5729. DOI: 10.3390/data6030034. [Online]. Available: https://www.mdpi.com/2306-5729/6/3/34.

[98] Gibson Kimutai and Förster Anna, "tea sickness dataset," Tech. Rep., 2022. DOI: 10.17632/J32XDT2FF5.1. [Online]. Available: https://data.mendeley.com/datasets/j32xdt2ff5/1.

[99] g. hu, h. wu, y. zhang, and m. wan, "Data for: A Low Shot Learning Method for Tea Leaf's Disease Identification," vol. 1, 2019. DOI: 10.17632/DBJYFKN6JR.1.

[100] A. N. Karuri, "Adaptation of Small-Scale Tea and Coffee Farmers in Kenya to Climate Change," *African Handbook of Climate Change Adaptation*, pp. 29–47, 2021. DOI: 10.1007/978-3-030-45106-6{_}70. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-030-45106-6_70.

[101] E. Asil, M. V. Yılmaz, and H. Yardimci, "Effects of black tea consumption and caffeine intake on depression risk in black tea consumers," *African Health Sciences*, vol. 21, no. 2, p. 858, Jun. 2021, ISSN: 16806905. DOI: 10.4314/AHS.V21I2.47. [Online]. Available: /pmc/articles/PMC8568251//pmc/articles/PMC8568251/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC8568251/.

[102] M. A. Hossain, T. Ahmed, M. S. Hossain, P. Dey, S. Ahmed, and M. M. Hossain, "Optimization of the factors affecting BT-2 black tea fermentation by observing their combined effects on the quality parameters of made tea using Response Surface Methodology (RSM)," *Heliyon*, vol. 8, no. 2, e08948, Feb. 2022, ISSN: 2405-8440. DOI: 10.1016/J.HELIYON.2022.E08948.

[103] H. Deka, P. P. Sarmah, A. Devi, P. Tamuly, and T. Karak, "Changes in major catechins, caffeine, and antioxidant activity during CTC processing of black tea from North East India," *RSC Advances*, vol. 11, no. 19, pp. 11 457–11 467, Mar. 2021, ISSN: 20462069. DOI: 10.1039/D0RA09529J. [Online]. Available: https://pubs.rsc.org/en/content/articlehtml/2021/ra/d0ra09529jhttps://pubs.rsc.org/en/content/articlelanding/2021/ra/d0ra09529j.

[104] M. K. Lee, H. W. Kim, S. H. Lee, *et al.*, "Characterization of catechins, theaflavins, and flavonols by leaf processing step in green and black teas (Camellia sinensis) using UPLC-DAD-QToF/MS," *European Food Research and Technology*, vol. 245, no. 5, pp. 997–1010, May 2019, ISSN: 14382385. DOI: 10.1007/S00217-018-3201-6/

FIGURES/5. [Online]. Available: https://link.springer.com/article/10.1007/s00217-018-3201-6.

[105] S. Zhang, Y. Yang, X. Cheng, *et al.*, "Prediction of suitable brewing cuppages of Dahongpao tea based on chemical composition, liquor colour and sensory quality in different brewing," *Scientific Reports 2020 10:1*, vol. 10, no. 1, pp. 1–11, Jan. 2020, ISSN: 2045-2322. DOI: 10.1038/s41598-020-57623-5. [Online]. Available: https://www.nature.com/articles/s41598-020-57623-5.

[106] N. Bhattacharyya, S. Seth, B. Tudu, *et al.*, "Detection of optimum fermentation time for black tea manufacturing using electronic nose," *Sensors and Actuators B: Chemical*, vol. 122, no. 2, pp. 627–634, Mar. 2007, ISSN: 0925-4005. DOI: 10.1016/J.SNB.2006.07.013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925400506004965.

[107] G. Kimutai, A. Ngenzi, R. N. Said, A. Kiprop, and A. Förster, "An Optimum Tea Fermentation Detection Model Based on Deep Convolutional Neural Networks," *Data*, vol. 5, no. 2, p. 44, Apr. 2020, ISSN: 2306-5729. DOI: 10.3390/data5020044. [Online]. Available: https://www.mdpi.com/2306-5729/5/2/44.

[108] Tea Board of Kenya, "Kenya Tea Yearly Report," Tea Board of Kenya, Nairobi, Tech. Rep., 2018, pp. 51–59.

[109] E. K. Kagira, S. W. Kimani, and K. S. Githii, "Sustainable Methods of Addressing Challenges Facing Small Holder Tea Sector in Kenya: A Supply Chain Management Approach," *Journal of Management and Sustainability*, vol. 2, no. 2, May 2012, ISSN: 1925-4733. DOI: 10.5539/jms.v2n2p75. [Online]. Available: http://www.ccsenet.org/journal/index.php/jms/article/view/17455.

[110] L. N. Hlahla, F. N. Mudau, and I. K. Mariga, "Effect of fermentation temperature and time on the chemical composition of bush tea (Athrixia phylicoides DC.)," *Journal of Medicinal Plants Research*, vol. 4, no. 9, pp. 824–829, May 2010, ISSN: 1996-0875. DOI: 10.5897/JMPR10.118. [Online]. Available: https://academicjournals.org/journal/JMPR/article-abstract/7DF679B20952.

[111] A. Ghosh, P. Sharma, B. Tudu, *et al.*, "Detection of Optimum Fermentation Time of Black CTC Tea Using a Voltammetric Electronic Tongue," *IEEE Transactions on Instrumentation and Measurement*, vol. 64, no. 10, pp. 2720–2729, Oct. 2015. DOI: 10.1109/TIM.2015.2415113. [Online]. Available: http://ieeexplore.ieee.org/document/7119610/.

[112] N Manigandan, "Handheld Electronic Nose (HEN) for detection of optimum fermentation time during tea manufacture and assessment of tea quality.," *International Journal of Advanced Research*, vol. 7, no. 7, pp. 697–702, Jul. 2019. DOI: 10.21474/ijar01/9413.

[113] L. Shen, L. R. Margolies, J. H. Rothstein, E. Fluder, R. McBride, and W. Sieh, "Deep Learning to Improve Breast Cancer Detection on Screening Mammography," *Scientific Reports*, vol. 9, no. 1, Dec. 2019, ISSN: 2045-2322. DOI: 10.1038/s41598-019-48995-4.

[114] G. Kimutai, P. W. Cheruiyot, and D. C. Otieno, "A Content Based Image Retrieval Model for E-Commerce," en, *International Journal of Engineering and Computer Science*, vol. 7, no. "11, pp. 24 392–24 396, Nov. 2018, ISSN: 2319-7242. DOI: 10.18535/ijecs/v7i11.01. [Online]. Available: http://www.ijecs.in/index.php/ijecs/article/view/4217http://www.ijecs.in/index.php/ijecs/article/download/4217/3884.

[115] Kimutai Gibson and Anna Förster, "Black Tea Fermentation Dataset," Tech. Rep., Mar. 2020. DOI: 10.17632/CBPF92VR5J.1.

[116] J. Marot and S. Bourennane, "Raspberry Pi for image processing education," in *25th European Signal Processing Conference, EUSIPCO 2017*, vol. 2017-January, Institute of Electrical and Electronics Engineers Inc., Oct. 2017, pp. 2364–2368, ISBN: 9780992862671. DOI: 10.23919/EUSIPCO.2017.8081633.

[117] S. K. Thangavel and M. Murthi, "A semi automated system for smart harvesting of tea leaves," in *2017 4th International Conference on Advanced Computing and Communication Systems, ICACCS 2017*, Institute of Electrical and Electronics Engineers Inc., Aug. 2017, ISBN: 9781509045594. DOI: 10.1109/ICACCS.2017.8014724.

[118] S. Narula, A. Jain, and Prachi, "Cloud computing security: Amazon web service," in *International Conference on Advanced Computing and Communication Technologies, ACCT*, vol. 2015-April, Institute of Electrical and Electronics Engineers Inc., Apr. 2015, pp. 501–505, ISBN: 9781479984886. DOI: 10.1109/ACCT.2015.20. [Online]. Available: http://ieeexplore.ieee.org/document/7079135/.

[119] F. Dubosson, S. Bromuri, and M. Schumacher, "A python framework for exhaustive machine learning algorithms and features evaluations," in *Proceedings - International Conference on Advanced Information Networking and Applications, AINA*, vol. 2016-May, Institute of Electrical and Electronics Engineers Inc., May 2016, pp. 987–993, ISBN: 9781509018574. DOI: 10.1109/AINA.2016.160.

[120] G. Kimutai, A. Ngenzi, R. N. Said, and A. Förster, "Dataset for Black Tea Fermentation Detection based on Image Processing and Machine Learning Techniques," Tech. Rep., Jan. 2021. DOI: 10.5281/ZENODO.4469326. [Online]. Available: https://zenodo.org/record/4469326.

[121] K. Bhargava, S. Ivanov, and W. Donnelly, "Internet of Nano Things for Dairy Farming," ser. NANOCOM' 15, ACM, Jan. 2015, pp. 1–24, ISBN: 978-1-4503-3674-1. DOI: 10.

1145/2800795.2800830. [Online]. Available: http://doi.acm.org/10.1145/2800795.2800830.

[122] M. Gao, M. Shi, and C. Li, "Research and Implementation of Image Recognition of Tea Based on Deep Learning," *Proceedings - 2021 21st ACIS International Semi-Virtual Winter Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, SNPD-Winter 2021*, pp. 63–68, Jan. 2021. DOI: 10.1109/SNPDWINTER52325.2021.00021.

[123] J. Wang, C. Zhang, M. Chang, *et al.*, "Optimization of electronic nose sensor array for tea aroma detecting based on correlation coefficient and cluster analysis," *Chemosensors*, vol. 9, no. 9, p. 266, Sep. 2021, ISSN: 22279040. DOI: 10.3390/CHEMOSENSORS9090266/S1. [Online]. Available: https://www.mdpi.com/2227-9040/9/9/266/htmhttps://www.mdpi.com/2227-9040/9/9/266.

[124] W. Xu, L. Zhao, J. Li, S. Shang, X. Ding, and T. Wang, "Detection and classification of tea buds based on deep learning," *Computers and Electronics in Agriculture*, vol. 192, p. 106 547, Jan. 2022, ISSN: 0168-1699. DOI: 10.1016/J.COMPAG.2021.106547.

[125] D. Saikia, P. K. Boruah, and U. Sarma, "A Sensor Network to Monitor Process Parameters of Fermentation and Drying in Black Tea Production," *MAPAN*, vol. 30, no. 3, pp. 211–219, Sep. 2015, ISSN: 0970-3950. DOI: 10.1007/s12647-015-0142-4. [Online]. Available: http://link.springer.com/10.1007/s12647-015-0142-4.

[126] R. Memisevic, "Deep learning: Architectures, algorithms, applications," in *2015 IEEE Hot Chips 27 Symposium, HCS 2015*, Institute of Electrical and Electronics Engineers Inc., May 2016, ISBN: 9781467388856. DOI: 10.1109/HOTCHIPS.2015.7477319.

[127] A. Luckow, K. Kennedy, M. Ziolkowski, *et al.*, "Artificial Intelligence and Deep Learning Applications for Automotive Manufacturing," in *Proceedings - 2018 IEEE International Conference on Big Data, Big Data 2018*, Institute of Electrical and Electronics Engineers Inc., Jan. 2019, pp. 3144–3152, ISBN: 9781538650356. DOI: 10.1109/BigData.2018.8622357.

[128] W. Zhu, L. Xie, J. Han, and X. Guo, "The Application of Deep Learning in Cancer Prognosis Prediction," *Cancers*, vol. 12, no. 3, p. 603, Mar. 2020, ISSN: 2072-6694. DOI: 10.3390/cancers12030603. [Online]. Available: https://www.mdpi.com/2072-6694/12/3/603.

[129] N. Khan and H. Mukhtar, "Tea and Health: Studies in Humans," *Current Pharmaceutical Design*, vol. 19, no. 34, pp. 6141–6147, Sep. 2013, ISSN: 13816128. DOI: 10.2174/1381612811319340008.

[130] S. Dargan, M. Kumar, M. R. Ayyagari, and G. Kumar, "A Survey of Deep Learning and Its Applications: A New Paradigm to Machine Learning," *Archives of Computational*

Methods in Engineering, pp. 1–22, Jun. 2019, ISSN: 18861784. DOI: 10.1007/s11831-019-09344-w.

[131] J. M. Barcelo-Ordinas, J. P. Chanet, K. M. Hou, and J García-Vidal, "A survey of wireless sensor technologies applied to precision agriculture," en, J. V. Stafford, Ed., Wageningen Academic Publishers, 2013, pp. 801–808, ISBN: 978-90-8686-778-3.

[132] P. N. Samarakoon, E. Promayon, and C. Fouard, "Light Random Regression Forests for automatic multi-organ localization in CT images," in *Proceedings - International Symposium on Biomedical Imaging*, IEEE Computer Society, Jun. 2017, pp. 371–374, ISBN: 9781509011711. DOI: 10.1109/ISBI.2017.7950540.

[133] A. Kumar, H. Singh, S. P. Sharma, and A. Kumar, *Color Analysis of Black Tea Liquor using Image Processing Techniques*, 2011. [Online]. Available: https://www.semanticscholar.org/paper/Color-Analysis-of-Black-Tea-Liquor-using-Image-Kumar-Singh/db8be924724706a2ac2c686561ea24455ef3f8e5.

[134] H. M. Ghodake, T. K. Goswami, and A. Chakraverty, "Mathematical Modeling of Withering Characteristics of Tea Leaves," *Drying Technology*, vol. 24, no. 2, pp. 159–164, Mar. 2006, ISSN: 0737-3937. DOI: 10.1080/07373930600558979. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/07373930600558979.

[135] J. Li and J. Z. Wang, "Automatic linguistic indexing of pictures by a statistical modeling approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1075–1088, Sep. 2003, ISSN: 01628828. DOI: 10.1109/TPAMI.2003.1227984.

[136] R. Tylecek and R. Fisher, "Consistent Semantic Annotation of Outdoor Datasets via 2D/3D Label Transfer," *Sensors*, vol. 18, no. 7, p. 2249, Jul. 2018, ISSN: 1424-8220. DOI: 10.3390/s18072249. [Online]. Available: http://www.mdpi.com/1424-8220/18/7/2249.

[137] A. Fabijańska and D. Sankowski, "Image noise removal - The new approach," in *The Experience of Designing and Application of CAD Systems in Microelectronics - Proceedings of the 9th International Conference, CADSM 2007*, 2007, pp. 457–459, ISBN: 9789665535874. DOI: 10.1109/CADSM.2007.4297618.

[138] K. Ahmad, J. Khan, and M. S. U. D. Iqbal, "A comparative study of different denoising techniques in digital image processing," in *2019 8th International Conference on Modeling Simulation and Applied Optimization, ICMSAO 2019*, Institute of Electrical and Electronics Engineers Inc., Apr. 2019, ISBN: 9781538676844. DOI: 10.1109/ICMSAO.2019.8880389.

[139] Y. Mistry, D. T. Ingole, and M. D. Ingole, "Content based image retrieval using hybrid features and various distance metric," *Journal of Electrical Systems and Information*

Technology, Jun. 2017, ISSN: 2314-7172. DOI: 10.1016/j.jesit.2016.12.009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2314717216301155.

[140] Y.-G. Yang, L. Zou, Y.-H. Zhou, and W.-M. Shi, "Visually meaningful encryption for color images by using Qi hyper-chaotic system and singular value decomposition in YCbCr color space," *Optik*, p. 164 422, Feb. 2020, ISSN: 0030-4026. DOI: 10.1016/j.ijleo.2020.164422.

[141] N. Alamgir, K. Nguyen, V. Chandran, and W. Boles, "Combining multi-channel color space with local binary co-occurrence feature descriptors for accurate smoke detection from surveillance videos," *Fire Safety Journal*, vol. 102, pp. 1–10, Dec. 2018, ISSN: 03797112. DOI: 10.1016/j.firesaf.2018.09.003.

[142] Z. H. Li, D. Han, C. J. Yang, T. Y. Zhang, and H. Q. Yu, "Probing operational conditions of mixing and oxygen deficiency using HSV color space," *Journal of Environmental Management*, vol. 232, pp. 985–992, Feb. 2019, ISSN: 10958630. DOI: 10.1016/j.jenvman.2018.12.025.

[143] C. L. Novak and S. A. Shafer, "Anatomy of a color histogram," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1992-June, IEEE Computer Society, 1992, pp. 599–605, ISBN: 0818628553. DOI: 10.1109/CVPR.1992.223129.

[144] H. Jung, K. Koo, and H. Yang, "Measurement-Based Power Optimization Technique for OpenCV on Heterogeneous Multicore Processor," *Symmetry*, vol. 11, no. 12, p. 1488, Dec. 2019, ISSN: 2073-8994. DOI: 10.3390/sym11121488. [Online]. Available: https://www.mdpi.com/2073-8994/11/12/1488.

[145] G. A. S. Saroja and C. H. Sulochana, "Texture analysis of non-uniform images using GLCM," in *2013 IEEE Conference on Information and Communication Technologies, ICT 2013*, 2013, pp. 1319–1322, ISBN: 9781467357586. DOI: 10.1109/CICT.2013.6558306.

[146] G. Kumar and P. K. Bhatia, "A detailed review of feature extraction in image processing systems," in *International Conference on Advanced Computing and Communication Technologies, ACCT*, Institute of Electrical and Electronics Engineers Inc., 2014, pp. 5–12, ISBN: 9781479949106. DOI: 10.1109/ACCT.2014.74.

[147] G. Preethi and V. Sornagopal, "MRI image classification using GLCM texture features," in *Proceeding of the IEEE International Conference on Green Computing, Communication and Electrical Engineering, ICGCCEE 2014*, Institute of Electrical and Electronics Engineers Inc., Oct. 2014, ISBN: 9781479949823. DOI: 10.1109/ICGCCEE.2014.6922461.

[148] O. A. Akanbi, I. S. Amiri, E. Fazeldehkordi, O. A. Akanbi, I. S. Amiri, and E. Fazeldehkordi, "Chapter 4 – Feature Extraction," in *A Machine-Learning Approach to Phishing Detection and Defense*, Syngress, Jan. 2015, pp. 45–54, ISBN: 9780128029275. DOI: 10.1016/B978-0-12-802927-5.00004-6.

[149] F. Xiaofeng and W. Wei, "Centralized binary patterns embedded with image euclidean distance for facial expression recognition," in *Proceedings - 4th International Conference on Natural Computation, ICNC 2008*, vol. 4, IEEE, 2008, pp. 115–119, ISBN: 9780769533049. DOI: 10.1109/ICNC.2008.94. [Online]. Available: http://ieeexplore.ieee.org/document/4667260/.

[150] C. Saravanan, "Color image to grayscale image conversion," in *2010 2nd International Conference on Computer Engineering and Applications, ICCEA 2010*, vol. 2, 2010, pp. 196–199, ISBN: 9780769539829. DOI: 10.1109/ICCEA.2010.192.

[151] EMC Education Services, *Data Science & Big Data Analytics*. Indianapolis, IN, USA: John Wiley & Sons, Inc, Jan. 2015, ISBN: 9781119183686. DOI: 10.1002/9781119183686. [Online]. Available: http://doi.wiley.com/10.1002/9781119183686.

[152] H. Dahan, S. Cohen, L. Rokach, and O. Maimon, *Proactive Data Mining with Decision Trees*, ser. SpringerBriefs in Electrical and Computer Engineering. New York, NY: Springer New York, 2014, ISBN: 978-1-4939-0538-6. DOI: 10.1007/978-1-4939-0539-3. [Online]. Available: http://link.springer.com/10.1007/978-1-4939-0539-3.

[153] A. Navada, A. N. Ansari, S. Patil, and B. A. Sonkamble, "Overview of use of decision tree algorithms in machine learning," in *Proceedings - 2011 IEEE Control and System Graduate Research Colloquium, ICSGRC 2011*, 2011, pp. 37–42, ISBN: 9781457703379. DOI: 10.1109/ICSGRC.2011.5991826.

[154] A. Brunello, E. Marzano, A. Montanari, and G. Sciavicco, "J48SS: A novel decision tree approach for the handling of sequential and time series data," *Computers*, vol. 8, no. 1, Mar. 2019, ISSN: 2073431X. DOI: 10.3390/computers8010021.

[155] S. Park, S.-Y. Hamm, and J. Kim, "Performance Evaluation of the GIS-Based Data-Mining Techniques Decision Tree, Random Forest, and Rotation Forest for Landslide Susceptibility Modeling," *Sustainability*, vol. 11, no. 20, pp. 1–20, 2019.

[156] W. Alajali, W. Zhou, S. Wen, and Y. Wang, "Intersection Traffic Prediction Using Decision Tree Models," *Symmetry*, vol. 10, no. 9, p. 386, Sep. 2018, ISSN: 2073-8994. DOI: 10.3390/sym10090386. [Online]. Available: http://www.mdpi.com/2073-8994/10/9/386.

[157] J. Han, M. Fang, S. Ye, C. Chen, Q. Wan, and X. Qian, “Using Decision Tree to Predict Response Rates of Consumer Satisfaction, Attitude, and Loyalty Surveys,” *Sustainability*, vol. 11, no. 8, pp. 1–13, 2019.

[158] D. Feng, Z. Deng, T. Wang, Y. Liu, and L. Xu, “Identification of disturbance sources based on random forest model,” in *2018 International Conference on Power System Technology, POWERCON 2018 - Proceedings*, Institute of Electrical and Electronics Engineers Inc., Jan. 2019, pp. 3370–3375, ISBN: 9781538664612. DOI: 10.1109/POWERCON.2018.8602245.

[159] S. Georganos, T. Grippa, A. Gadiaga, *et al.*, “An application of geographical random forests for population estimation in Dakar, Senegal using very-high-resolution satellite imagery,” in *2019 Joint Urban Remote Sensing Event, JURSE 2019*, Institute of Electrical and Electronics Engineers Inc., May 2019, ISBN: 9781728100098. DOI: 10.1109/JURSE.2019.8809049.

[160] Y. Liu, L. Liu, Y. Gao, and L. Yang, “An improved random forest algorithm based on attribute compatibility,” in *Proceedings of 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference, ITNEC 2019*, Institute of Electrical and Electronics Engineers Inc., Mar. 2019, pp. 2558–2561, ISBN: 9781538662434. DOI: 10.1109/ITNEC.2019.8729146.

[161] A. Saffari, C. Leistner, J. Santner, M. Godec, and H. Bischof, “On-line random forests,” in *IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops*, 2009, pp. 1393–1400, ISBN: 9781424444427. DOI: 10.1109/ICCVW.2009.5457447.

[162] A. Z. Kouzani, S. Nahavandi, and K. Khoshmanesh, “Face classification by a random forest,” in *IEEE Region 10 Annual International Conference, Proceedings/TENCON*, 2007, ISBN: 1424412722. DOI: 10.1109/TENCON.2007.4428937.

[163] A. S. More and D. P. Rana, “Review of random forest classification techniques to resolve data imbalance,” in *Proceedings - 1st International Conference on Intelligent Systems and Information Management, ICISIM 2017*, vol. 2017-Janua, Institute of Electrical and Electronics Engineers Inc., Nov. 2017, pp. 72–78, ISBN: 9781509042647. DOI: 10.1109/ICISIM.2017.8122151.

[164] T. Qunzhu, Z. Rui, Y. Yufei, Z. Chengyao, and L. Zhijun, “Improvement of random forest cascade regression algorithm and its application in fatigue detection,” in *2019 2nd International Conference on Electronics Technology, ICET 2019*, Institute of Electrical and Electronics Engineers Inc., May 2019, pp. 499–503, ISBN: 9781728116181. DOI: 10.1109/ELTECH.2019.8839317.

[165] B. Xu, Y. Ye, and L. Nie, "An improved random forest classifier for image classification," in *EEE International Conference on Information and Automation, ICIA*, 2012, pp. 795–800, ISBN: 9781467322386. DOI: 10.1109/ICInfA.2012.6246927.

[166] M. S. Alam and S. T. Vuong, "Random forest classification for detecting android malware," in *Proceedings - 2013 IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing, GreenCom-iThings-CPSCom*, 2013, pp. 663–669, ISBN: 9780769550466. DOI: 10.1109/GreenCom-iThings-CPSCom.2013.122.

[167] P. Thanh Noi and M. Kappas, "Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery," *Sensors*, vol. 18, no. 2, p. 18, Dec. 2017, ISSN: 1424-8220. DOI: 10.3390/s18010018. [Online]. Available: http://www.mdpi.com/1424-8220/18/1/18.

[168] H. Saadatfar, S. Khosravi, J. H. Joloudari, A. Mosavi, and S. Shamshirband, "A New K-Nearest Neighbors Classifier for Big Data Based on Efficient Data Pruning," *Mathematics*, vol. 8, no. 2, p. 286, Feb. 2020, ISSN: 2227-7390. DOI: 10.3390/math8020286. [Online]. Available: https://www.mdpi.com/2227-7390/8/2/286.

[169] G.-F. Fan, Y.-H. Guo, J.-M. Zheng, and W.-C. Hong, "Application of the Weighted K-Nearest Neighbor Algorithm for Short-Term Load Forecasting," *Energies*, vol. 12, no. 5, pp. 1–19, 2019.

[170] M. S. Aldayel, "K-Nearest Neighbor classification for glass identification problem," in *International Conference on Computer Systems and Industrial Informatics, ICCSII*, 2012, ISBN: 9781467351577. DOI: 10.1109/ICCSII.2012.6454522.

[171] P. Viswanath and T. Hitendra Sarma, "An improvement to k-nearest neighbor classifier," in *IEEE Recent Advances in Intelligent Computational Systems, RAICS*, 2011, pp. 227–231, ISBN: 9781424494774. DOI: 10.1109/RAICS.2011.6069307.

[172] S. Sun and R. Huang, "An adaptive k-nearest neighbor algorithm," in *Proceedings - 2010 7th International Conference on Fuzzy Systems and Knowledge Discovery, FSKD*, vol. 1, 2010, pp. 91–94, ISBN: 9781424459346. DOI: 10.1109/FSKD.2010.5569740.

[173] S. Zhang, X. Li, M. Zong, X. Zhu, and R. Wang, "Efficient kNN classification with different numbers of nearest neighbors," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 5, pp. 1774–1785, May 2018, ISSN: 21622388. DOI: 10.1109/TNNLS.2017.2673241.

[174] P. P. Jayaraman, A. Zaslavsky, and J. Delsing, "Intelligent processing of K-nearest neighbors queries using mobile data collectors in a location aware 3D wireless sensor network," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 6098 LNAI, Springer,

Berlin, Heidelberg, 2010, pp. 260–270, ISBN: 3642130321. DOI: 10.1007/978-3-642-13033-5{_}27.

[175] A. Cord and S. Chambon, “Automatic Road Defect Detection by Textural Pattern Recognition Based on AdaBoost,” *Computer-Aided Civil and Infrastructure Engineering*, vol. 27, no. 4, pp. 244–259, Apr. 2012, ISSN: 10939687. DOI: 10.1111/j.1467-8667.2011.00736.x. [Online]. Available: http://doi.wiley.com/10.1111/j.1467-8667.2011.00736.x.

[176] H. R. Roth, A. Farag, L. Lu, E. B. Turkbey, and R. M. Summers, “Deep convolutional networks for pancreas segmentation in CT imaging,” *Medical Imaging 2015: Image Processing*, vol. 9413, 94131G, Apr. 2015. DOI: 10.1117/12.2081420. [Online]. Available: http://arxiv.org/abs/1504.03967http://dx.doi.org/10.1117/12.2081420.

[177] W. Wang, B. Wu, S. Yang, and Z. Wang, “Road Damage Detection and Classification with Faster R-CNN,” in *Proceedings - 2018 IEEE International Conference on Big Data, Big Data 2018*, Institute of Electrical and Electronics Engineers Inc., Jan. 2019, pp. 5220–5223, ISBN: 9781538650356. DOI: 10.1109/BigData.2018.8622354.

[178] S. C. Radopoulou and I. Brilakis, “Automated Detection of Multiple Pavement Defects,” *Journal of Computing in Civil Engineering*, vol. 31, no. 2, Mar. 2017, ISSN: 08873801. DOI: 10.1061/(ASCE)CP.1943-5487.0000623.

[179] C. Chen, Y. Ma, and G. Ren, “A Convolutional Neural Network with Fletcher–Reeves Algorithm for Hyperspectral Image Classification,” *Remote Sensing*, vol. 11, no. 11, p. 1325, Jun. 2019, ISSN: 2072-4292. DOI: 10.3390/rs11111325. [Online]. Available: https://www.mdpi.com/2072-4292/11/11/1325.

[180] M. P. Véstias, “A Survey of Convolutional Neural Networks on Edge with Reconfigurable Computing,” *Algorithms*, vol. 12, no. 8, p. 154, Jul. 2019, ISSN: 1999-4893. DOI: 10.3390/a12080154. [Online]. Available: https://www.mdpi.com/1999-4893/12/8/154.

[181] J. Song, S. Gao, Y. Zhu, and C. Ma, “A survey of remote sensing image classification based on CNNs,” *Big Earth Data*, vol. 3, no. 3, pp. 232–254, Jul. 2019, ISSN: 2096-4471. DOI: 10.1080/20964471.2019.1657720.

[182] D. C. Cire\csan, U. Meier, J. Masci, L. M. Gambardella, and J. Schmidhuber, “Flexible, High Performance Convolutional Neural Networks for Image Classification,” in *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence - Volume Volume Two*, ser. IJCAI’11, AAAI Press, 2011, 1237–1242, ISBN: 9781577355144.

[183] M. Paoletti, J. Haut, J. Plaza, and A. Plaza, “Deep&Dense Convolutional Neural Network for Hyperspectral Image Classification,” *Remote Sensing*, vol. 10, no. 9, p. 1454, Sep. 2018, ISSN: 2072-4292. DOI: 10.3390/rs10091454. [Online]. Available: http://www.mdpi.com/2072-4292/10/9/1454.

[184] G. Hinton, L. Deng, D. Yu, *et al.*, *Deep Neural Networks for Acoustic Modeling in Speech Recognition*, 2012.

[185] H. Elaidi, Y. Elhaddar, Z. Benabbou, and H. Abbar, “An idea of a clustering algorithm using support vector machines based on binary decision tree,” in *2018 International Conference on Intelligent Systems and Computer Vision, ISCV 2018*, vol. 2018-May, Institute of Electrical and Electronics Engineers Inc., May 2018, pp. 1–5, ISBN: 9781538643969. DOI: 10.1109/ISACV.2018.8354024.

[186] C. F. Chao and M. H. Horng, “The construction of support vector machine classifier using the firefly algorithm,” *Computational Intelligence and Neuroscience*, vol. 2015, Feb. 2015, ISSN: 16875273. DOI: 10.1155/2015/212719.

[187] G. Cheng and X. Tong, “Fuzzy clustering multiple kernel support vector machine,” in *International Conference on Wavelet Analysis and Pattern Recognition*, vol. 2018-July, IEEE Computer Society, Nov. 2018, pp. 7–12, ISBN: 9781538652176. DOI: 10.1109/ICWAPR.2018.8521307.

[188] S. Shahbudin, M. Zamri, M. Kassim, S. A. C. Abdullah, and S. I. Suliman, “Weed classification using one class support vector machine,” in *2017 International Conference on Electrical, Electronics and System Engineering, ICEESE 2017*, vol. 2018-Janua, Institute of Electrical and Electronics Engineers Inc., Feb. 2018, pp. 7–10, ISBN: 9781538609088. DOI: 10.1109/ICEESE.2017.8298404.

[189] N. Kranjčić, D. Medak, R. Župan, and M. Rezo, “Support Vector Machine Accuracy Assessment for Extracting Green Urban Areas in Towns,” *Remote Sensing*, vol. 11, no. 6, p. 655, Mar. 2019, ISSN: 2072-4292. DOI: 10.3390/rs11060655. [Online]. Available: https://www.mdpi.com/2072-4292/11/6/655.

[190] H. Walia, A. Rana, and V. Kansal, “A Naïve Bayes Approach for working on Gurmukhi Word Sense Disambiguation,” in *2017 6th International Conference on Reliability, Infocom Technologies and Optimization: Trends and Future Directions, ICRITO 2017*, vol. 2018-Janua, Institute of Electrical and Electronics Engineers Inc., Apr. 2018, pp. 432–435, ISBN: 9781509030125. DOI: 10.1109/ICRITO.2017.8342465.

[191] A. H. Jahromi and M. Taheri, “A non-parametric mixture of Gaussian naive Bayes classifiers based on local independent features,” in *19th CSI International Symposium on Artificial Intelligence and Signal Processing, AISP 2017*, vol. 2018-Janua, Institute of Electrical and Electronics Engineers Inc., Mar. 2018, pp. 209–212, ISBN: 9781538625859. DOI: 10.1109/AISP.2017.8324083.

[192] F. Harahap, A. Y. N. Harahap, E. Ekadiansyah, R. N. Sari, R. Adawiyah, and C. B. Harahap, "Implementation of Naïve Bayes Classification Method for Predicting Purchase," in *2018 6th International Conference on Cyber and IT Service Management, CITSM 2018*, Institute of Electrical and Electronics Engineers Inc., Mar. 2019, ISBN: 9781538654330. DOI: 10.1109/CITSM.2018.8674324.

[193] Y. N. Dewi, D. Riana, and T. Mantoro, "Improving Naïve Bayes performance in single image pap smear using weighted principal component analysis (WPCA)," in *3rd International Conference on Computing, Engineering, and Design, ICCED 2017*, vol. 2018-March, Institute of Electrical and Electronics Engineers Inc., Mar. 2018, pp. 1–5, ISBN: 9781538634950. DOI: 10.1109/CED.2017.8308130.

[194] H. Zhao, Z. Wang, and F. Nie, "A New Formulation of Linear Discriminant Analysis for Robust Dimensionality Reduction," *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, no. 4, pp. 629–640, Apr. 2019, ISSN: 15582191. DOI: 10.1109/TKDE.2018.2842023.

[195] M. Singh and K. L. Maharjan, "Crop Production and Net Return from Organic and Conventional Farming Systems," in *Sustainability of Organic Farming in Nepal*, Singapore: Springer Singapore, 2017, pp. 133–149. DOI: 10.1007/978-981-10-5619-2{_}9. [Online]. Available: http://link.springer.com/10.1007/978-981-10-5619-2_9.

[196] V. M. Jerkovic, V. Kojic, and M. B. Popovic, "Linear discriminant analysis: Classification of on-surface and in-air handwriting," in *2015 23rd Telecommunications Forum, TELFOR 2015*, Institute of Electrical and Electronics Engineers Inc., Jan. 2016, pp. 460–463, ISBN: 9781509000548. DOI: 10.1109/TELFOR.2015.7377506.

[197] N. A. A. Shashoa, N. A. Salem, I. N. Jleta, and O. Abusaeeda, "Classification depend on linear discriminant analysis using desired outputs," in *2016 17th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering, STA 2016 - Proceedings*, Institute of Electrical and Electronics Engineers Inc., Jun. 2017, pp. 328–332, ISBN: 9781509034079. DOI: 10.1109/STA.2016.7952041.

[198] P. P. Markopoulos, "Linear Discriminant Analysis with few training data," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, Institute of Electrical and Electronics Engineers Inc., Jun. 2017, pp. 4626–4630, ISBN: 9781509041176. DOI: 10.1109/ICASSP.2017.7953033.

[199] J. Sun, X. Cai, F. Sun, and J. Zhang, "Scene image classification method based on Alex-Net model," in *2016 3rd International Conference on Informative and Cybernetics for Computational Social Systems, ICCSS 2016*, Institute of Electrical and Electronics Engineers Inc., Oct. 2016, pp. 363–367, ISBN: 9781509033676. DOI: 10.1109/ICCSS.2016.7586482.

[200] S. Thalagala and C. Walgampaya, "Application of AlexNet convolutional neural network architecture-based transfer learning for automated recognition of casting surface defects," *Proceedings - International Research Conference on Smart Computing and Systems Engineering, SCSE 2021*, pp. 129–136, Sep. 2021. DOI: 10.1109/SCSE53661.2021.9568315.

[201] M. Z. Alom, T. M. Taha, C. Yakopcic, *et al.*, "The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches," Mar. 2018. DOI: 10.48550/arxiv.1803.01164. [Online]. Available: https://arxiv.org/abs/1803.01164v2.

[202] Y. Beeharry and V. Bassoo, "Performance of ANN and AlexNet for weed detection using UAV-based images," *2020 3rd International Conference on Emerging Trends in Electrical, Electronic and Communications Engineering, ELECOM 2020 - Proceedings*, pp. 163–167, Nov. 2020. DOI: 10.1109/ELECOM49001.2020.9296994.

[203] P. Kim, "Machine Learning," in *MATLAB Deep Learning*, Berkeley, CA: Apress, 2017, pp. 1–18. DOI: 10.1007/978-1-4842-2845-6{_}1. [Online]. Available: http://link.springer.com/10.1007/978-1-4842-2845-6_1.

[204] I. Sutskever, J. Martens, G. Dahl, and G. Hinton, "On the Importance of Initialization and Momentum in Deep Learning," in *Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28*, ser. ICML'13, JMLR.org, 2013, III–1139–III–1147.

[205] B. R. Samal, A. K. Behera, and M. Panda, "Performance analysis of supervised machine learning techniques for sentiment analysis," in *Proceedings of 2017 3rd IEEE International Conference on Sensing, Signal Processing and Security, ICSSS 2017*, Institute of Electrical and Electronics Engineers Inc., Oct. 2017, pp. 128–133, ISBN: 9781509049295. DOI: 10.1109/SSPS.2017.8071579.

[206] R. Tohid, B. Wagle, S. Shirzad, *et al.*, "Asynchronous execution of python code on task-based runtime systems," in *Proceedings of ESPM2 2018: 4th International Workshop on Extreme Scale Programming Models and Middleware, Held in conjunction with SC 2018: The International Conference for High Performance Computing, Networking, Storage and Analysis*, Institute of Electrical and Electronics Engineers Inc., Feb. 2019, pp. 37–45, ISBN: 9781728101781. DOI: 10.1109/ESPM2.2018.00009.

[207] I. Stancin and A. Jovic, "An overview and comparison of free Python libraries for data mining and big data analysis," in *2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO 2019 - Proceedings*, Institute of Electrical and Electronics Engineers Inc., May 2019, pp. 977–982, ISBN: 9789532330984. DOI: 10.23919/MIPRO.2019.8757088.

[208] M. Abadi, P. Barham, J. Chen, *et al.*, *TensorFlow: A system for large-scale machine learning*, 2016.

[209] S. K. Fahad and A. E. Yahya, “Big Data Visualization: Allotting by R and Python with GUI Tools,” in *2018 International Conference on Smart Computing and Electronic Enterprise, ICSCEE 2018*, Institute of Electrical and Electronics Engineers Inc., Nov. 2018, ISBN: 9781538648360. DOI: `10.1109/ICSCEE.2018.8538413`.

[210] B. Baker and L. Pan, “Overview of the Model and Observation Evaluation Toolkit (MONET) Version 1.0 for Evaluating Atmospheric Transport Models,” *Atmosphere*, vol. 8, no. 12, p. 210, Oct. 2017, ISSN: 2073-4433. DOI: `10.3390/atmos8110210`. [Online]. Available: `http://www.mdpi.com/2073-4433/8/11/210`.

[211] H.-C. Hung, I.-F. Liu, C.-T. Liang, and Y.-S. Su, “Applying Educational Data Mining to Explore Students’ Learning Patterns in the Flipped Learning Approach for Coding Education,” *Symmetry*, vol. 12, no. 2, p. 213, Feb. 2020, ISSN: 2073-8994. DOI: `10.3390/sym12020213`. [Online]. Available: `https://www.mdpi.com/2073-8994/12/2/213`.

[212] F. Pedregosa, G. Varoquaux, A. Gramfort, *et al.*, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, Oct. 2011, ISSN: 15324435.

[213] Z. Zeng, Q. Gong, and J. Zhang, “CNN model design of gesture recognition based on tensorflow framework,” in *Proceedings of 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference, ITNEC 2019*, Institute of Electrical and Electronics Engineers Inc., Mar. 2019, pp. 1062–1067, ISBN: 9781538662434. DOI: `10.1109/ITNEC.2019.8729185`.

[214] G. Li, Z. Liu, L. Cai, and J. Yan, “Standing-Posture Recognition in Human–Robot Collaboration Based on Deep Learning and the Dempster–Shafer Evidence Theory,” *Sensors*, vol. 20, no. 4, p. 1158, Feb. 2020, ISSN: 1424-8220. DOI: `10.3390/s20041158`. [Online]. Available: `https://www.mdpi.com/1424-8220/20/4/1158`.

[215] H. G. Momm, R. ElKadiri, and W. Porter, “Crop-Type Classification for Long-Term Modeling: An Integrated Remote Sensing and Machine Learning Approach,” *Remote Sensing*, vol. 12, no. 3, p. 449, Feb. 2020, ISSN: 2072-4292. DOI: `10.3390/rs12030449`. [Online]. Available: `https://www.mdpi.com/2072-4292/12/3/449`.

[216] A. Kosmopoulos, I. Partalas, E. Gaussier, G. Paliouras, and I. Androutsopoulos, “Evaluation Measures for Hierarchical Classification: a unified view and novel approaches,” Jun. 2013. DOI: `10.1007/s10618-014-0382-x`. [Online]. Available: `http://arxiv.org/abs/1306.6802http://dx.doi.org/10.1007/s10618-014-0382-x`.

[217] P. Flach, “Performance Evaluation in Machine Learning: The Good, the Bad, the Ugly, and the Way Forward,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, pp. 9808–9814, Jul. 2019, ISSN: 2159-5399. DOI: `10.1609/aaai.v33i01.33019808`.

[218] M. Hai, Y. Zhang, and Y. Zhang, "A Performance Evaluation of Classification Algorithms for Big Data," in *Procedia Computer Science*, vol. 122, Elsevier B.V., 2017, pp. 1100–1107. DOI: 10.1016/j.procs.2017.11.479.

[219] H. Yue, J. Guo, and X. Yin, "A Comparison of Traditional Machine Learning and Deep Learning in Image Recognition," *Journal of Physics: Conference Series*, vol. 1314, no. 1, p. 012 148, Oct. 2019, ISSN: 1742-6596. DOI: 10.1088/1742-6596/1314/1/012148. [Online]. Available: https://iopscience.iop.org/article/10.1088/1742-6596/1314/1/012148https://iopscience.iop.org/article/10.1088/1742-6596/1314/1/012148/meta.

[220] S. Smith, "Manifesto for the Humanities," *Manifesto for the Humanities*, May 2015. DOI: 10.3998/DCBOOKS.13607059.0001.001.

[221] S. Khan and T. Yairi, "A review on the application of deep learning in system health management," *Mechanical Systems and Signal Processing*, vol. 107, pp. 241–265, Jul. 2018, ISSN: 10961216. DOI: 10.1016/j.ymssp.2017.11.024.

[222] J. Hernandez-Gonzalez, I. Inza, and J. A. Lozano, "A Note on the Behavior of Majority Voting in Multi-Class Domains with Biased Annotators," *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, no. 1, pp. 195–200, Jan. 2019, ISSN: 15582191. DOI: 10.1109/TKDE.2018.2845400.

[223] E. C. Too, L. Yujian, S. Njuki, and L. Yingchun, "A comparative study of fine-tuning deep learning models for plant disease identification," *Computers and Electronics in Agriculture*, vol. 161, pp. 272–279, Jun. 2019, ISSN: 01681699. DOI: 10.1016/j.compag.2018.03.032.

[224] S. Sarkar, N. Bhattacharyya, and V. K. Palakurthi, "Taste recognizer by multi sensor electronic tongue: A case study with tea quality classification," *Proceedings - 2nd International Conference on Emerging Applications of Information Technology, EAIT 2011*, pp. 138–141, 2011. DOI: 10.1109/EAIT.2011.42.

[225] Q. Chen, J. Zhao, Z. Chen, H. Lin, and D. A. Zhao, "Discrimination of green tea quality using the electronic nose technique and the human panel test, comparison of linear and nonlinear classification tools," *Sensors and Actuators, B: Chemical*, vol. 159, no. 1, pp. 294–300, Nov. 2011, ISSN: 09254005. DOI: 10.1016/j.snb.2011.07.009.

[226] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, Jun. 2017, ISSN: 15577317. DOI: 10.1145/3065386. [Online]. Available: http://dl.acm.org/citation.cfm?doid=3098997.3065386.

[227] A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, "CNN Features off-the-shelf: an Astounding Baseline for Recognition," *IEEE Computer Society Conference on Com-*

puter Vision and Pattern Recognition Workshops, pp. 512–519, Mar. 2014. [Online]. Available: http://arxiv.org/abs/1403.6382.

[228] M. H. Kamrul, M. Rahman, M. Risul Islam Robin, M. Safayet Hossain, M. H. Hasan, and P. Paul, "A deep learning based approach on categorization of tea leaf," in *ACM International Conference Proceeding Series*, New York, NY, USA: Association for Computing Machinery, Jan. 2020, pp. 1–8, ISBN: 9781450377782. DOI: 10.1145/3377049.3377122. [Online]. Available: https://dl.acm.org/doi/10.1145/3377049.3377122.

[229] H. Zhou, F. Ni, Z. Wang, F. Zheng, and N. Yao, "Classification of Tea Pests Based on Automatic Machine Learning," in *Lecture Notes in Electrical Engineering*, vol. 653, Springer Science and Business Media Deutschland GmbH, 2021, pp. 296–306, ISBN: 9789811585982. DOI: 10.1007/978-981-15-8599-9{_}35.

[230] G. Hu, H. Wu, Y. Zhang, and M. Wan, "A low shot learning method for tea leaf's disease identification," *Computers and Electronics in Agriculture*, vol. 163, p. 104 852, Aug. 2019, ISSN: 0168-1699. DOI: 10.1016/J.COMPAG.2019.104852.

[231] B Karunamoorthy and D Somasundereswari, "A Defect Tea Leaf Identification Using Image Processing," *Przegląd Elektrotechniczny*, pp. 318–320, 2013.

[232] J. Zhao and K. S. Trivedi, "Performance modeling of apache web server affected by aging," in *Proceedings - 2011 3rd International Workshop on Software Aging and Rejuvenation, WoSAR 2011*, 2011, pp. 56–61, ISBN: 9780769546162. DOI: 10.1109/WoSAR.2011.13.

[233] H. M. P. C. Kumarihami and K. J. Song, "Review on Challenges and Opportunities in Global Tea Industry," *The Korean Tea Society*, vol. 24, no. 3, pp. 79–87, Sep. 2018, ISSN: 12259640. DOI: 10.29225/jkts.2018.24.3.79.

[234] M. Obanda, P. Okinda Owuor, and R. Mang'oka, "Changes in the chemical and sensory quality parameters of black tea due to variations of fermentation time and temperature," *Food Chemistry*, vol. 75, no. 4, pp. 395–404, Dec. 2001, ISSN: 03088146. DOI: 10.1016/S0308-8146(01)00223-0.

[235] N. O'Mahony, S. Campbell, A. Carvalho, *et al.*, "Deep Learning vs. Traditional Computer Vision," in *Advances in Intelligent Systems and Computing*, vol. 943, Springer Verlag, 2020, pp. 128–144, ISBN: 9783030177942. DOI: 10.1007/978-3-030-17795-9{_}10.

[236] S. Khan and T. Yairi, *A review on the application of deep learning in system health management*, Jul. 2018. DOI: 10.1016/j.ymssp.2017.11.024.

[237] G. Kimutai, A. Ngenzi, N. Rutabayiro, and A. Förster, "Offloading an Energy Efficient IoT Solution to the Edge: A practical Solution for Developing Countries," in *IEEE Global Humanitarian Technology Conference (GHTC)*, Virtual: IEEE, Oct. 2021. [Online]. Available: https://www.researchgate.net/publication/355667998_Offloading_an_Energy_Efficient_IoT_Solution_to_the_Edge_A_practical_Solution_for_Developing_Countries.

[238] W. R. Kahoreria, S. Bett, and K. Wanjala, "Power Connectivity and Performance of Kenya Power and Lighting Company in Western Kenya," *The International Journal of Science & Technoledge*, vol. 7, no. 10, Oct. 2019. DOI: 10.24940/theijst/2019/v7/i10/st1910-026.

[239] K. Lee, E. Brewer, C. Christiano, *et al.*, "Electrification for "under Grid" households in Rural Kenya," *Development Engineering*, vol. 1, pp. 26–35, Jun. 2016, ISSN: 23527285. DOI: 10.1016/j.deveng.2015.12.001.

[240] P. F. Ogola, B. Davidsdottir, and I. B. Fridleifsson, *Potential contribution of geothermal energy to climate change adaptation: A case study of the arid and semi-arid eastern Baringo lowlands, Kenya*, Aug. 2012. DOI: 10.1016/j.rser.2012.01.081.

[241] M. Aazam, S. Zeadally, and E. F. Flushing, "Task offloading in edge computing for machine learning-based smart healthcare," *Computer Networks*, vol. 191, p. 108 019, 2021, ISSN: 1389-1286. DOI: https://doi.org/10.1016/j.comnet.2021.108019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1389128621001298.

[242] M. S. Hossain, C. I. Nwakanma, J. M. Lee, and D. S. Kim, "Edge computational task offloading scheme using reinforcement learning for IIoT scenario," *ICT Express*, vol. 6, no. 4, pp. 291–299, Dec. 2020, ISSN: 24059595. DOI: 10.1016/j.icte.2020.06.002.

[243] D. Rahbari and M. Nickray, "Task offloading in mobile fog computing by classification and regression tree," *Peer-to-Peer Networking and Applications*, vol. 13, no. 1, pp. 104–122, Jan. 2020, ISSN: 19366450. DOI: 10.1007/S12083-019-00721-7/TABLES/6. [Online]. Available: https://link.springer.com/article/10.1007/s12083-019-00721-7.

[244] S. Wang, J. Li, G. Wu, H. Chen, and S. Sun, "Joint Optimization of Task Offloading and Resource Allocation Based on Differential Privacy in Vehicular Edge Computing," *IEEE Transactions on Computational Social Systems*, 2021, ISSN: 2329924X. DOI: 10.1109/TCSS.2021.3074949.

[245] Q. Zhu, B. Si, F. Yang, and Y. Ma, "Task offloading decision in fog computing system," *China Communications*, vol. 14, no. 11, pp. 59–68, Nov. 2017, ISSN: 16735447. DOI: 10.1109/CC.2017.8233651.

[246] F. Astudillo-Salinas, D. Barrera-Salamea, A. Vazquez-Rodas, and L. Solano-Quinde, "Minimizing the power consumption in Raspberry Pi to use as a remote WSN gateway," *2016 8th IEEE Latin-American Conference on Communications, LATINCOM 2016*, 2016. DOI: 10.1109/LATINCOM.2016.7811590.

[247] V. Vujović and M. Maksimović, "Raspberry Pi as a Wireless Sensor node: Performances and constraints," *2014 37th International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO 2014 - Proceedings*, pp. 1013–1018, 2014. DOI: 10.1109/MIPRO.2014.6859717.

[248] S. Ferdoush and X. Li, "Wireless Sensor Network System Design Using Raspberry Pi and Arduino for Environmental Monitoring Applications," *Procedia Computer Science*, vol. 34, pp. 103–110, Jan. 2014, ISSN: 1877-0509. DOI: 10.1016/J.PROCS.2014.07.059.

[249] M. P. Véstias, "A Survey of Convolutional Neural Networks on Edge with Reconfigurable Computing," *Algorithms*, vol. 12, no. 8, p. 154, Jul. 2019, ISSN: 1999-4893. DOI: 10.3390/a12080154. [Online]. Available: https://www.mdpi.com/1999-4893/12/8/154.

[250] R. Khan, S. U. Khan, R. Zaheer, and S. Khan, "Future internet: The internet of things architecture, possible applications and key challenges," in *Proceedings - 10th International Conference on Frontiers of Information Technology, FIT 2012*, 2012, pp. 257–260, ISBN: 9780769549279. DOI: 10.1109/FIT.2012.53.

[251] G. Kimutai, A. Ngenzi, S. R. Ngoga, R. C. Ramkat, and A. Förster, "An internet of things (IoT)-based optimum tea fermentation detection model using convolutional neural networks (CNNs) and majority voting techniques," *Journal of Sensors and Sensor Systems*, vol. 10, no. 2, pp. 153–162, Jul. 2021, ISSN: 2194878X. DOI: 10.5194/JSSS-10-153-2021.

[252] T. K. Kim, "Understanding one-way ANOVA using conceptual figures," *Korean Journal of Anesthesiology*, vol. 70, no. 1, p. 22, Feb. 2017, ISSN: 20057563. DOI: 10.4097/KJAE.2017.70.1.22. [Online]. Available: /pmc/articles/PMC5296382//pmc/articles/PMC5296382/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC5296382/.

[253] A. Mahajan, "Post hoc tests in analysis of variance," *Indian Journal of Occupational and Environmental Medicine*, vol. 20, no. 2, p. 121, May 2016, ISSN: 19983670. DOI: 10.4103/0019-5278.197552. [Online]. Available: /pmc/articles/PMC5299813/https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5299813/.

[254] B. K. Hackenberger, "R software: unfriendly but probably the best," *Croatian Medical Journal*, vol. 61, no. 1, p. 66, Feb. 2020, ISSN: 13328166. DOI: 10.3325/CMJ.2020.

61.66. [Online]. Available: /pmc/articles/PMC7063554/https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7063554/.

[255] B. Charyyev, E. Arslan, and M. Gunes, *Latency Comparison of Cloud Datacenters and Edge Servers*. Dec. 2020. DOI: 10.1109/GLOBECOM42002.2020.9322406.

Appendix A

Awards and Grants

A.1 Award

1. **GHTC 2021 Conference best paper award**. Title of the paper: Offloading an Energy-Efficient IoT Solution to the Edge: A Practical Solution for Developing Countries

A.2 Grants

1. **KENET IoT minigrant**. Title of the project: A black tea fermentation detection model using CNN and image processing processing. Awarded amount: 15,000 USD

2. **University of Bremen BISIP scholarship**. I won a scholarship from the University of Bremen for an internship and exchange at the department of sustainable networks from November to February 2022 Awarded amount: 3650 Euros

Appendix B

Publications

B.1 Published Papers

The following are a list of published papers:

1. **G. Kimutai**, A. Ngenzi, R. N. Said, A. Kiprop, and A. Förster, “An Optimum Tea Fermentation Detection Model Based on Deep Convolutional Neural Networks,” Data, vol. 5, no. 2, p. 44, Apr. 2020.

2. **G. Kimutai**, A. Ngenzi, R. Ngoga Said, R. C. Ramkat, and A. Förster, “A Data Descriptor for Black Tea Fermentation Dataset,” Data, vol. 6, no. 3, p. 34, Mar. 2021.

3. **G. Kimutai**, A.Ngenzi, S. R. Ngoga, R. C. Ramkat, and A. FÖrster, “An internet of things (IoT)-based optimum tea fermentation detection model using convolutional neural networks (CNNs) and majority voting techniques,” Journal of Sensors and Sensor Systems, vol. 10, no. 2, pp. 153–162, Jul. 2021.

4. **G. Kimutai**, A. Ngenzi, N. Rutabayiro, and A. Förster, “Offloading an Energy Ef- ficient IoT Solution to the Edge: A practical Solution for Developing Countries,” in IEEE Global Humanitarian Technology Conference (GHTC), Virtual: IEEE, Oct. 2021. **Best Conference paper award**

5. **Gibson Kimutai** and Anna Förster. 2022. An instance-based deep transfer learning approach for resource-constrained environments. In Proceedings of the ACM SIGCOMM Workshop on Networked Sensing Systems for a Sustainable Society (NET4us ’22). Association for Computing Machinery, New York, NY, USA, 39–45.