EJAS Journal of Science and Technology 1 (2024) 33 – 39

EJAS Journal of Science and Technology
journal homepage: https://ejas.edu.et/

Automatic Tea Leaf Classification Using Deep Learning Models

Dejene Teka

Department of Information Technology, Mattu University, Mattu, Ethiopia
email: dejutek@gmail.com; dejene.teka@meu.edu.et
Submitted: 14 December 2023. Accepted: 15 February 2024. Available online: 5 March 2024.

Abstract

Tea leaf classification is a crucial task in the tea industry for quality control and grading purposes. However, manual classification methods are time-consuming and subjective. This study addresses the challenges of early quality detection for tea leaves by developing an automatic leaf quality detection technique. Three deep learning models, namely CNN, Inception v3, and EfficientNet B0, were utilized to classify tea leaves based on their visual features. The CNN model is widely used for image classification, while Inception v3 employs a complex architecture to improve accuracy. EfficientNet B0, on the other hand, achieves state-of-the-art performance with faster training times. A dataset of tea leaf images was collected and split into training, validation, and testing sets. The models were trained on the training set, fine-tuned using the validation set, and evaluated on the testing set. The results demonstrated that both CNN and EfficientNet B0 surpassed the Inception v3 model, achieving accuracy rates of 96.9% and 97.4%, respectively, compared to Inception v3's accuracy rate of 95.9%. These findings highlight the effectiveness of deep learning algorithms in tea leaf classification and their potential for further improvements and applications in the tea industry.

Key words: Deep learning, tea leaf classification, convolutional neural networks, Inception v3, EfficientNet B0.

1. Introduction

Tea leaves are a popular beverage in the country and one of the most lucrative export products. Tea leaf production, is divided into export, domestic consumption, and utility imports. As one of the major beverages worldwide, tea plays a vital role in our daily lives [1]. Tea harvesting policies include several aspects of plucking, such as the plucking process, style, intensity and frequency. There are various factors that affect the quality of the industry tea, some of reasons being: how to remove green leaves, drying capacity, fermentation time, drying conditions, type of green leaf clones, and cultivation methods. And tea leaves are affected by various conditions such as shear conditions, cultivation conditions, fertilizer application conditions, environmental conditions, plant health protection conditions, and plant age. The tea plant industry that has successfully implemented harvesting policies in a tea estate has a direct impact on tea plantation production, or rather viability. [2] Tea is a drink that is consumed in large amounts worldwide on a daily basis. Its origins can be traced back to China, where it was initially consumed for medicinal purposes. It contains several beneficial compounds, including flavonoids, amino acids, vitamins, caffeine, and polysaccharides. Tea can contribute to good health in various ways, as it does not contain added unhealthy substances such as sodium, fat, or sugar. Tea (of the Assam type) was a new crop introduced to Ethiopia in 1927 and is mostly grown in the Oromia Region, the Ilu Aba Bora Zone, Alle district, and near Gore Town. In 1989, two tea estates called 'Wushwush' and 'Gumero' started to commercially produce the crop in Ethiopia, leading to the establishment of many tea plantations throughout the country. Unlike other countries such as Assam, Ceylon and Indonesia, Ethiopia grows the crop in mostly single cultivation, without tree shade. The optimal climate to grow crops in Ethiopia is in areas with at least 1,500 mm of annual rainfall, a mean air temperature of 18 to 20 degrees Celsius, an average humidity of 70 to 90%, and five hours of sunlight per day. [3]

Tea leaf classification is the process of assessing the quality and condition of tea leaves in the tea industry. This process determines the grade of tea, which ranges from the highest grade, orange pekoe, to the lowest grade, fannings or dust. Depending on the number of adjacent young leaves picked from the leaf buds, different grades of pekoe tea are determined. The highest-quality pekoe grades are made only from leaf buds picked with the fingertips to avoid bruising. Certain grades of leaf are better suited to different types of tea, such as processing buds or shoots of the tea plant for white tea [4]. Fresh tea leaf is unusually rich in the flavanol group of polyphenols known as catechins, which may constitute up to 30% of the dry leaf weight. Other polyphenols include flavanols and their glycosides, and depsides such as chlorogenic acid, coumarylquinic acid, and one unique to tea, theogallin (3- galloylquinic acid) [5].

The classification of tea leaves can be conducted in various ways, such as by size, appearance, verities and weight, orange pekoe, whole-leaf grades, broken-leaf grades, etc. [6]. From these, we are trying to rank the tea leaves on the basis of their size, verities and weight. CNN, InceptionV3 and EffeceintNetB0 are the models we have chosen for tea leaf classification. In recent years, there has been rapid progress in the fields of artificial intelligence, which have played an increasingly important role in agriculture. Scientists have been able to use a type of computer technology called "deep learning" to recognize objects and develop systems for quality control [7].

2. Related work

The Paper is refers to a research paper that proposed a method of plant leaf classification that uses convolutional neural networks (CNNs). The method involves constructing a 10-layer CNN to classify leaf images. Furthermore, the researchers used a technique called sample augment to expand the database, which should help improve the classification accuracy. Finally, the article used a visualization technique to analyze the factors that influence the accuracy rate of the classification method. The results of the research paper showed that the proposed method achieved a high overall accuracy rate of 87.92% for a database of 4,800 leaf images and 32 kinds of leaf. [8]

Convolutional neural networks (CNNs) are models of artificial intelligence, designed to analyze data, identify patterns, and make predictions. They are composed of interconnected neurons and are designed to detect patterns in data using a mathematical process known as convolution. CNNs are often used in image classification and related applications because of their accuracy and speed. In the context of the research paper, these authors used CNN to classify tea leaves based on their fluorescence images. The research investigated the strategies of customers to identify tea varieties and their attributes. Although the technique involves the use of smartphone and computer vision, they came up with the idea that the predictive model may not always be accurate due to variations in lighting. To improve accuracy, this paper proposed a new fluorescence imaging technique that uses an ultraviolet LED lamp with a central wavelength of 370 nm to excite fluorescence in tea samples and capture tea images. Two different computer models, VGG16 and ResNet-34, were tested to classify the shale images. The results showed that the accuracy of classification models is improved when the tea is imaged using fluorescence technique, the accuracy is 97.5%, and the use of this fluorescence imaging method in our daily life proves that it can be useful [9].

The main focus of this research paper is on automated machinery to extract high-quality green tea. To do this, the authors proposed a method that combines semi-supervised learning (which is a type of machine learning that uses both labeled and unlabeled data for training) and image processing (which is manipulating digital images and videos for a desired output) to identify the tender leaves, which are young green tea leaves that have the 32 best quality. The authors used two methods, called the gradient descent method and the Adam algorithm, to optimize the objective function. The results of the experiments showed that the average accuracy of the model was 92.62% with an average misjudgment rate of 18.86%. These results imply that the model, based on semi-supervised learning, can effectively identify tender leaves in early spring and can be used to improve the problem of deep learning requiring a large number of labeled samples. [10]. A computer program called a "convolutional neural network" was created to identify different types of tea leaves by analyzing images of them. This was done because traditional sorting methods were not very accurate. The program has several layers and was trained using pictures of a specific type of tea called "Huoshanhuangya." The paper explains how a machine learning technique called Convolutional neural network (CNN) can be used to sort fresh tea leaves in real time. The researchers designed a CNN model and trained it using 8000 pictures of a specific type of tea. They then tested the model by identifying 2000 fresh tea leaves with an accuracy rate of 95.3% and an average recognition time of 0.3795 seconds per image. The research shows the use of CNN model for tea sorting is a feasible and effective method. [11]

2. Methodology

2.1. Data Collection

Tea leaf image data was collected from Gumaro Tea Leaf Factory. We need a lot of images to develop and evaluate a model. We collected 2,410 images from the Gumero Tea Leaf Factory which is located in the Oromia Region, Ilu Aba Bora Zone, Alle districts and near Gore Town. The images were of good quality and were taken with a smart camera, which was ready to be used to create the dataset.

2.2. Data Set Preparation

Data sets are collections of images that are used for research or application purposes. For this research, images from agricultural fields were used to create five classes: One_plus_bud, two_plus_bud, three_plus_bud, soft-bench, and bad leaf. The dataset has been created in a specific way- 70% of the dataset has been used for training and the remaining 30% has been used for testing the model.

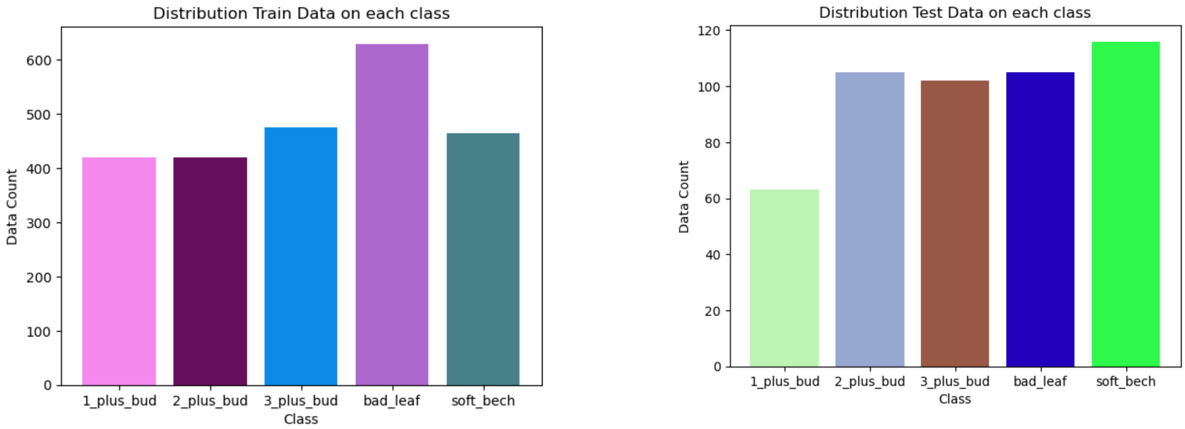

Fig 1: Dataset Distribution

Table 1:Dataset preparation

Class Name	Train	Test	Total
1_plus_bud	420	63	483
2_plus_bud	420	105	525
3_plus_bud	476	102	578
Bad_leaf	630	105	735
Soft_bech	464	116	580
Total	**2,410**	**491**	**2901**

2.3. Proposed System Architecture

The proposed architecture for this research involves two phases: training and testing. In the training phase, the image data is collected, pre-processed, and split into training and test data. Feature extraction is then implemented to reduce the dimensionality of the data. In the testing phase, unseen data is preprocessed and the model detects whether the image is a 1_plus_bud,2_plus_bud,3_plus_bud,bad_leaf,soft_bech, and classifies them.

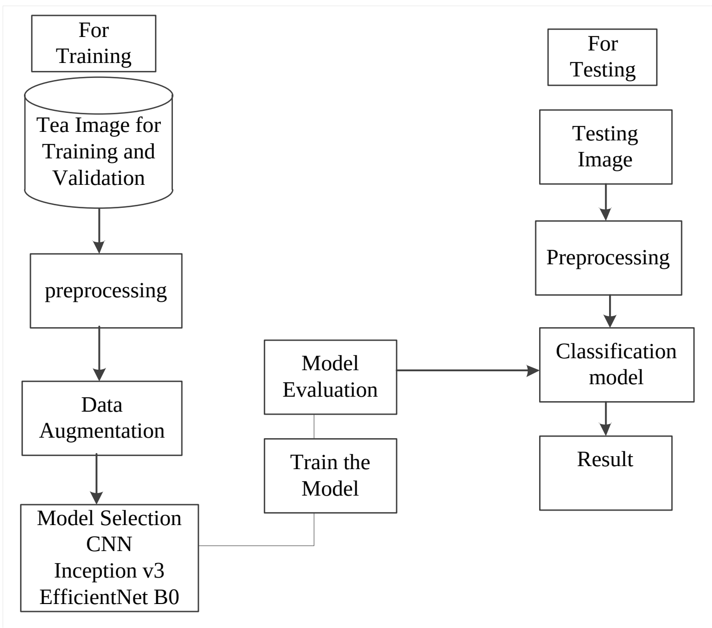

Fig 2: Proposed System Architecture

3. Experimental Result

The content describes the implementation of a classification process using a CNN algorithm to classify tea leaf, such as 1_plus_bud, 2_plus_bud, 3_plus_bud, bad_leaf, and soft_bech, The experimentation details, including the results of each experiment, are presented in graphs and tables. Additionally, the process of feature extraction is explained, which involves selecting and combining variables into features to effectively reduce the amount of data and improve accuracy.

3.1. Result Analysis of CNN Model

In the content, two plots are described that illustrate the CNN model's classification accuracy and loss with respect to epochs. For a dataset of 256 x 256 x 3 images, the plots show metrics such as training and testing accuracy, training loss, and testing loss. The high accuracy of the models with short epochs is demonstrated by the plots, and the testing accuracy and loss match the training accuracy and loss, respectively. By Epoch 25, the model's training accuracy had increased to 96% from a starting point of about 78%. Due to the data set, the high accuracy is achieved with short epochs, and the validation accuracy and loss coincide with the training accuracy and loss, respectively.

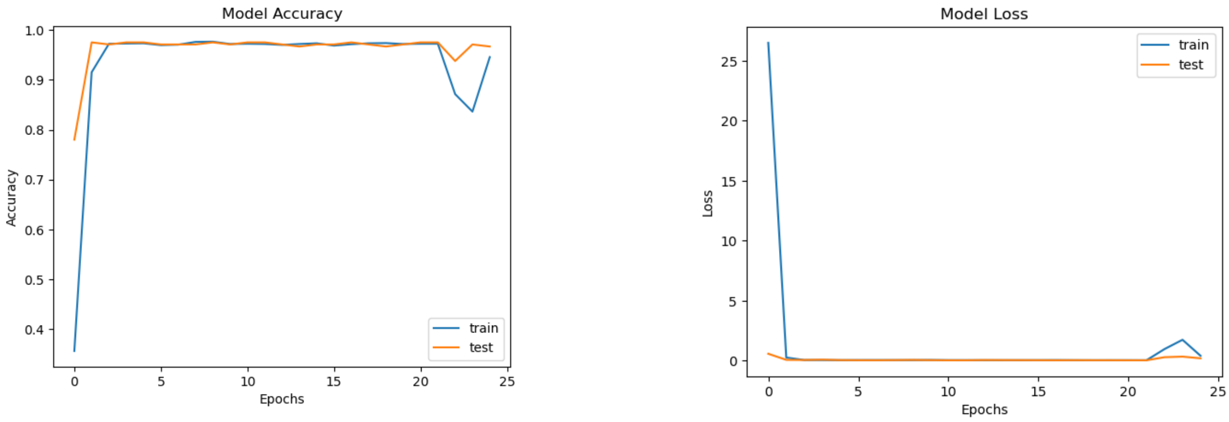

Fig 3: Loss and Accuracy of CNN Model

3.2. Result Analysis of Inception V3 Model

InceptionV3 is a convolutional neural network (CNN) architecture that was developed by Google for image classification tasks. It is a deep learning model that uses multiple layers of convolutional and pooling operations to extract features from images and classify them into different categories. The InceptionV3 model's classification accuracy and loss with respect to epochs are displayed in the next two plots. For an image dataset, the graps show metrics such as testing and training accuracy, testing loss, and training loss.

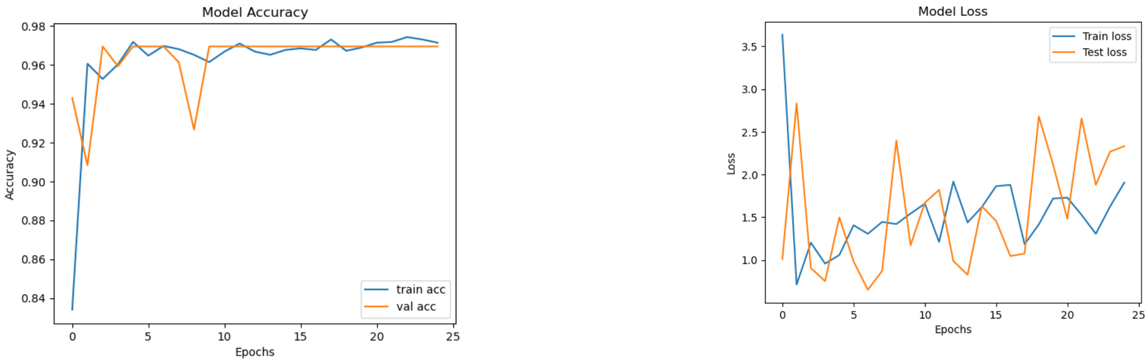

Fig 4: Loss and accuracy of Inception V3 Model

3.3. Result analysis of Efficient Net B0 Model

EfficientNetB0 is a convolutional neural network architecture that was introduced in 2019 by Google AI researchers. It is designed to achieve state-of-the-art accuracy on image classification tasks while being computationally efficient. The model achieves this by using a combination of techniques such as compound scaling, depth-wise separable convolutions, and efficient channel attention.EfficientNetB0 model is a highly efficient and accurate model for image classification tasks. Its performance can be further improved by fine-tuning on specific datasets or using transfer learning techniques.The EfficientNetB0 model's classification accuracy and loss are displayed in these two plots in relation to epochs. For an image dataset, the plots show metrics like testing and training accuracy, testing loss, and training loss.

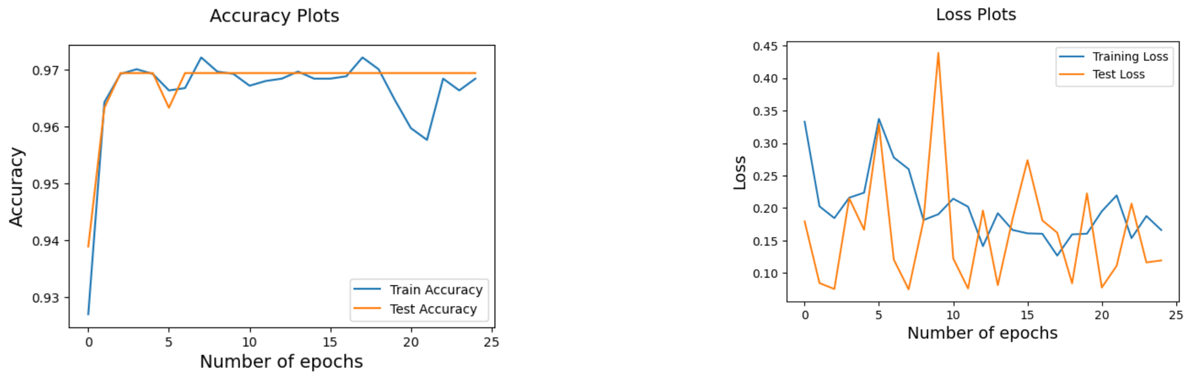

Fig 5: Loss and accuracy of Efficient Net B0 model

3.4. Comparison of the models

Model	Number of Parameter	Number of epochs	Accuracy	Processor Type
Convolutional neural network(CNN) (224X244 pixes)	1,487,488	25	96.9%	GPU
InceptionV3 (299X299 pixes)	22,458,1449	25	95.9%	GPU
EfficientNetB0	4,061,096	25	97.4%	GPU

3.5. Discussion

In this experiment, we evaluate how well a Convolutional Neural Network (CNN) algorithm can classify tea leaf images according to their color, variety, weight, and size. The regions of the tea leaf are defined by the study using RGB color characteristics and GLCM color feature extraction. The findings demonstrate that the suggested model can classify tea leaves with high accuracy while requiring little data sets and little processing time. The study defined the areas of interest using RGB color characteristics in order to detect and classify tea leaf. The CNN algorithm was used to test the detection and classification performances, and the InceptionV3 and EfficientNetB0 architectures were compared in order to determine which model performed best when model evaluation was taken into account.

InceptionV3 and EfficientNetB0 are two types of deep learning models used for image classification tasks. InceptionV3 was developed by Google in 2015 and uses a combination of convolutional layers, pooling layers, and inception modules to extract features from images. EfficientNetB0, developed in 2019, is generally faster and more accurate than InceptionV3, but may require more computational resources during training. The choice between these models depends on the specific requirements of the project and the available resources. The researcher uses a Convolutional Neural Network (CNN) algorithm for image data processing, and use pre- trained model like InceptionV3 and EfficientNetB0 which is suitable for classifying different types of tea leaf. The proposed model achieved high accuracy of 96.9%, 95.9%, 97.4% respectively in training while maintaining time efficiency and effectiveness in classifying 1_plus_bud, 2_plus_bud, 3_plus_bud, bad leaf, and soft_bech.

4. Future Work

Future work in tea leaf classification can focus on several areas to further improve the accuracy and applicability of the automated leaf quality detection technique. Firstly, expanding the dataset by including a wider variety of tea leaf images from different regions and cultivars would enhance the models' ability to generalize and classify diverse tea leaves accurately. Additionally, incorporating other data modalities such as spectral or chemical information can provide complementary features and improve classification performance. Exploring transfer learning techniques by pretraining the models on larger and more diverse image datasets, such as ImageNet, and then fine-tuning them on the tea leaf dataset may also lead to improved results. Furthermore, investigating the interpretability of the deep learning models and understanding the visual features they rely on for classification can provide valuable insights for the tea industry. Lastly, conducting real-world testing and validation of the automated classification system in tea production facilities would be crucial to assess its practicality and integration into existing quality control processes.

5. Conclusion

This study demonstrates the successful application of deep learning algorithms for automating tea leaf classification. By utilizing three different models, namely CNN, Inception v3, and Efficient-Net B0, the researcher achieved high accuracy rates in classifying tea leaves based on their visual features. The CNN and Efficient-Net B0 models outperformed the Inception v3 model, achieving accuracy rates of 96.9% and 97.4%, respectively. The findings highlight the potential of deep learning algorithms to revolutionize the tea industry by improving the efficiency and accuracy of tea grading processes. The automation of tea leaf classification can save time and reduce subjectivity associated with manual classification by human experts. This advancement in technology can significantly contribute to quality control and grading of tea leaves, ultimately enhancing the overall quality of tea products. The study's results provide valuable insights for tea producers and processors to enhance their tea grading systems, ensuring the consistency and quality of tea products. The use of deep learning algorithms opens up possibilities for further research and development in the field, leading to potential improvements in classification accuracy and the application of this technology to other aspects of the tea industry.

References

[1] "Wei K, chen B, li z , and chen D, "Classification of Tea Leaves Based on Fluorescence Imaging and Convolutional Neural Networks.," , vol. 1, China, 20222, p. 7764.".

[2] "Karunasena G and Priyankara D. N. S. Priyankara, "Tea Bud Leaf Identification by Using Machine Learning and Image Processing Techniques," in International Journal of Scientific & Engineering Research 11.8, 2020, pp. 624-628.".

[3] "Dechassa Nagassa and Misgan Marga, "Tea (Camellia sinensis) Production, Opportunities, Challenges," , vol. 2022, Article ID 1942666, 5, p. 2022.".

[4] "Smith and Krisi, "World Atlas of Tea. Great Britain: Mitchell Beazley," p. 22, 2016.".

[5] "Harold N. Graham, "Green tea composition, consumption, and polyphenol chemistry," pp. 334-350.".

[6] "Karl Schapira David Schapira, The book of coffee & tea : a guide to the appreciation of fine coffees, teas, and herbal beverages, 2nd editon.".

[7] "Xu, Wenkai, Longgang Zhao, Juan Li, Shuqi Shang, X, "Detection and classification of tea buds based on deep learning," , 2022.".

[8] "LUIS J, Yuong S, Cheng Y, and Song z, "Plant leaf classification based on deep learning.," in In 2018 Chinese Automation Congress (CAC, 2018, pp. 3165-3169.".

[9] ""Classification of Tea Leaves Based on Fluorescence Imaging and," , vol. 32, 2022, p. 7764.".

[10] "Yung Jie and Yong Chen , "Tender Leaf Identification for Early-Spring Green Tea Based on," in Agronomy, vol. 12, 2022, p. 1958.".

[11] "F. Maosheng and S. Yun, Z. Yanyan, "Study on sorting technology of fresh tea leaves based on convolutional neural network model," pp. 614-617, 2019.".

[12] "Vidura Bandara Wijekoon, "Using AI for Improving Quality of Green Tea Leaves in Sri Lanka - National Level AI project," , 2021, p. 11.".

[13] "Chiu W, "FACTORS AFFECTING THE PRODUCTION AND QUALITY OF PARTIALLY FERMENTED TEA IN TAIWAN," In International Symposium on the Culture of Subtropical and Tropical Fruits and Crops, pp. 57-64.".

[14] "Y., Tadesse, B., Mekonnen, G. and Bekele, T. Tilahun, Sarhad Journal of Agriculture , 2022.".

Authors Profile

Dejene Teka. holds a Bachelor of Technology (B.Tech.) Honours degree in Information Technology from Mattu University. He received his M.Sc. degrees in Information Technology from Mattu University. His present employment is teaching and research as an academic staff in department of Information Technology, College of Engineering, Mattu University.