# Automated Store Billing System Based on Deep Learning (Image Detection and Computer Vision)

MSc Research Project
MSc in Data Analytics

Vijaykumar Ghanti
Student ID: x21237174

School of Computing
National College of Ireland

Supervisor: Prof. Furqan Rustam

**National College of Ireland**
**Project Submission Sheet**
**School of Computing**

| | |
|---|---|
| **Student Name:** | Vijaykumar Ghanti |
| **Student ID:** | x21237174 |
| **Programme:** | MSc in Data Analytics |
| **Year:** | 2023 |
| **Module:** | MSc Research Project |
| **Supervisor:** | Prof. Furqan Rustam |
| **Submission Due Date:** | 14/08/2023 |
| **Project Title:** | Automated Store Billing System Based on Deep Learning (Image Detection and Computer Vision) |
| **Word Count:** | 10300 |
| **Page Count:** | 33 |

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

**ALL** internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

| | |
|---|---|
| **Signature:** | Vijaykumar |
| **Date:** | 18th September 2023 |

**PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:**

| | |
|---|---|
| Attach a completed copy of this sheet to each project (including multiple copies). | ☐ |
| **Attach a Moodle submission receipt of the online project submission**, to each project (including multiple copies). | ☐ |
| **You must ensure that you retain a HARD COPY of the project**, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer. | ☐ |

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

| **Office Use Only** | |
|---|---|
| Signature: | |
| Date: | |
| Penalty Applied (if applicable): | |

# Automated Store Billing System Based on Deep Learning (Image Detection and Computer Vision)

Vijaykumar Ghanti
x21237174

**Abstract**

The potential of human errors in recording, transcribing product prices and calculating during checkout leads to customer dissatisfaction and loss of revenue to the stores. So, the aim of this research is to generate bill for customers on top of the fruits they have selected without depending on the internet and cloud. By which the technology can be spread across different remote parts of the world. To achieve this, four different deep learning algorithms like CNN, RCNN, ResNet, AlexNet and four different machine learning algorithms like SVM, KNN, Logistic Regression and Naive Bayes are implemented and evaluated on five fruits (Avocado, AppleBraeburn, Banana, Apricot and Beetroot) of Fruit 360 degree dataset obtained from GitHub. Through extensive experimentation, it was found that by achieving 98.46% of mean accuracy, 99% of precision, recall and F1-score AlexNet outperformed all other algorithms by accurately detecting the fruits. On addition to this, confusion matrix analysis further substantiated the exceptional performance of the AlexNet. 10 epochs were considered to build the model, where accuracy increased from 49.63% at 1st epoch to 98.46% at 10th epoch. In short, the proposed system with AlexNet algorithm on top of the selected dataset provides cutting edge solution for bill generation for the customers without depending on internet and cloud by using Alexnet deep learnign algorithm.

## 1 Introduction

The efficiency and accuracy of store billing system plays a vital role in achieving both customer satisfaction and business profitability in today's fast faced retail landscape. On one side, inefficient billing process lead to customer frustration and on other side result in significant revenue losses for retailers. National retail federation said, due to errors in pricing retail got shrink and inventory management costed to $61.7 billion in 2020 in the United States alone Federation (2020). These errors may be : errors in scanning, errors in data entry during the checkout process, deliberate fraud and misplacement of items. To revolutionize the retail industry by mitigating these errors, this research aims to develop an Automated store billing system based on Deep learning, image processing and computer vision.

Traditionally, the store billing systems consists of manual processes where, cashier by scanning barcode enters the price manually. These methods are susceptible to human errors and time consuming Barchard (2011). On addition to this, if the selected product lacks barcode, irregularly shaped or barcode is damaged then cashier will struggle to bill the product and customer won't experience the smooth billing process. To address these issues, by using computer vision and image detection various attempts have been made into the retail domain. For example, the study Kalkundre et al. (2022) proposes a new billing system, which incorporates deep learning neural networks to classify the product and achieves an image processing accuracy of 98.86% by using RasberryPi board with tensor flow and OpenCV library. These efforts have shown promising results in product recognition and price estimation automatically. But, there are still more efforts required to replace traditional billing system in real world retail environments.

A subset of artificial intelligence (AI) called deep learning, has emerged as best technology to show remarkable achievements in computer vision and image processing tasks LeCun (2015). By using deep neural networks, we can develop automated store billing system to accurately identify products, their prices and quantify by taking images during checkout process. This approach stands as promising step to reduce errors, speed up billing process, provide smooth shopping experience to customers and enhance profit range to retail owners as well. Our research employs convolutional neural networks to analyse image of products at the point of sale. And by using visual characteristics of fruits like shape, color and branding identifies the fruits. Then by fetching corresponding prices of the fruits from database generates bill for the customer.

### 1.0.1 Researh question

The main aim of research is to achieve: How can the technology of automated store billing systems be disseminated across numerous stores in different parts of the world by making it internet or cloud servers free?

### 1.0.2 Research objectives

To achieve this, the following specific objectives have been set:

1. Performance evaluation of top four different deep learning algorithms (CNN, R-CNN, ResNet and AlexNet) and four machine learning algorithms (SVM, KNN, Logistic Regression, Naive Bayes) for fruits image classification.

2. Identification of the most suitable algorithm for fruits image classification by considering accuracy, precision, recall, and F1-score.

3. Implementing a billing system with code that can integrate with a deep learning algorithm and generate accurate bills for customers.

4. Ensure the execution of the technique is free of internet and cloud dependencies, utilizing only local storage data.

As part of contribution to the scientific literature, this research contributes significantly to the computer vision, image processing, deep learning and machine learning. The comprehensive evaluation of performance of four deep learning algorithms and four machine learning algorithms on top of the selected dataset adds valuable knowledge to the researchers with respect to the performance of deep learning and machine learning algorithms for fruits detection by image processing. Moreover, selecting AlexNet as the best image processing algorithm, for its outstanding accuracy and precision, generate accurate bills for customers. Combination of deep learning and bill generation system without depending upon internet or cloud server provides a novel approach for automated billing in retail environments.

This research project is organised into several sections and subsections to present cohesive account for the research. Following this introduction, the literature review helps in understanding updates in the relevant fields of fruit classification and billing systems. Methodology section provides information about dataset and the implementation of deep learning and machine learning algorithms. Following this, the results and Discussion section represents the performance evaluation of algorithms and the outcomes of billing system. Finally, in the conclusion section it is summarised about finding the best algorithm, implications and even outlines potential avenues for future research.

# 2 Related Work

With the aim of achieving high accuracy and efficiency in detecting fruits by image processing and computer vision, has gained significant attention in stores like supermarkets to generate bills for customers and to provide smooth shopping experience to customers. The ability to automatically detect the fruits can revolutionize the billing process, reduces human errors and checkout times. This literature review critically analyses the existing research and explores advancements in fruit detection using image processing and computer vision techniques in different views, as mentioned below in he subsections. By keenly observing the state-of-the-art studies, one can understand key challenges, gaps, and opportunities. This knowledge paves the way for development of robust and efficient fruit detection systems for retail billing applications.

## 2.1 Chronological Review

The study Krizhevsky et al. (2012) in order to detect fruits, some of the key technical points considered are : The architecture includes data augmentation to reduce overfitting, ReLu for non-linearity and deep CNN for pooling. To replace traditional pooling, max-pool overlapping will be considered, to prevent co-adaptation dropout is implemented. Author used imageNet's vast data and implemented parallel computing with GPUs, which resulted in significant object detection. In the another study Simonyan and Zisserman (2014), Due to hierarchical feature learning ability of VGG16 they used it as Deep learning algorithm. To initialize the meaningful weights large scale imagenet dataset was used. A softmax classifier provided class probabilities and Data Augmentation was used to counter data scarcity issue. To handle varied fruit sizes, multi-scale approach was used. VGG16 is superior over AlexNet but its heavy architecture stands as constraint, the author said. In another research He et al. (2015a), the author introduced key concepts like global average pooling, revolutionizing convolutional neural networks, Deep architecture, and Residual Blocks. By surpassing depth limitations of previous approaches and by addressing training challenges through methods like identity mapping and batch normalization, the model achieved state-of-the-art accuracy. The use of architectural modifications and skip connections enabled efficient training of very deep networks. Techniques like, data augmentation and stochastic gradient descent were used to enhance the performance further.

In this research Ren et al. (2015) the author described a unique architecture for real-time fruit detection that makes use of features such as Region Proposal Networks, shared convolutional layers, and end-to-end training. A Region Proposal Network (RPN) predicts object bounding boxes and objectness scores using anchor-based region suggestions. This effective method narrows down areas of interest. Faster R-CNN blends RPN with object detection networks, leveraging feature extraction techniques such as RoI pooling, multi-task loss, and VGG16/ZF Net backbones. Despite significant improvements, computing demands continue to pose a difficulty to real-time implementation. In another study, Redmon et al. (2016) author used the YOLO framework (You Only Look Once) implements real-time object detection with single-shot detection, a grid-based technique, bounding box and class prediction, anchor boxes, and darknet architecture. In contrast to two-stage approaches, it predicts class probabilities and bounding boxes directly from the complete image in a single pass. Grid cells anticipate two or five bounding boxes, which are parameterized as offsets. The study employs a darknet architecture with 24 convolutions and two fully linked layers for accurate localization, as well as class prediction and loss algorithms. While YOLO excels in real-time performance, it falls short in complex settings in terms of accuracy.

Overall in chronological review, it is understood that the evolution of deep learning models for image processing task overcome the drawbacks of predecessors. And got to know that there is a need to explore their performance on 360-defree fruit datasets.

## 2.2 Literature on Object Detection Methods

The author Tan et al. (2020) with the aim of achieving scalability and efficiency introduced EfficientDet, a novel object detection architecture. Research introduced BiFPN for feature combination , compounding scaling for architecture enhancement and for balancing accuracy, computational efficiency EfficientNet stood as backbone. Focal loss and BiFPN used to address class imbalance and feature fusion through learnable weights respectively. Additionally, to enhance convergence and training efficiency the combination of SGD with momentum and ada bound learning rate scheduling was introduced. The other research Law and Deng (2018) introduced CornetNet in 2018, by eliminating the need for anchor boxes revolutionized one-stage object detection. By identifying top left and bottom right corners of bounding boxes it excels in detecting different object shapes and sizes. In first stage of CornerNet, heatmaps for these corners will be predicted, followed by generating object proposals. Refining these proposals and predicting object categories includes in second stage. Further to enhance the accuracy, corner pooling and focal loss was introduced by CornerNet. An impressive 42.2% mean average was introduced by Implementing CornerNet on the COCO dataset.

The anchor-free one-stage object detection CenterNet introduced in 2019 Duan et al. (2019), which directly locates object centers and corners. CenterNet for precise object detection uses cascade corner pooling and center pooling. By surpassing previous one stage detectors accuracy of 42.2%, achieved impressive accuracy of 47% on the COCO dataset and rivals two-stage models. Versatility for objects of varying shapes and sizes enhanced by its direct center and corner prediction. Though, CenterNet holds promise for real world applications and exhibits significant improvements, faces few limitations in crowded scenarios. In following year 2018, a multistage object detection algorithm known for its high accuracy and computational efficiency called Cascade RCNN was proposed by Cai and VasconcelosCai and Vasconcelo (2018). While predicting object categories, it exhibits selectivity in generating candidate object bounding boxes and uses multiple stages for refining them. While focal loss addresses dataset imbalance, to prevent overfitting at each stage resampling maintains a balanced set of positive and negative examples. To further mitigate overfitting early stopping is used. By outperforming one-stage detectors and two stage detectors Cascade RCNN achieved an accuracy of 48.8%.

In other study  (n.d.), without region proposals to offer rapid and accurate detection SSD (Single Stage Detector) was proposed. To predict object class labels and bounding boxes it uses a single CNN. The multi-stage prediction and prior boxes for accurate bounding box detection are key features of SSD. Due to SSD's multi stage prediction approach it excels in detecting objects of varying sizes. And by surpassing one-stage and two-stage detectors in both speed and accuracy it achieved stage-of-the-art performance on Pascal VOC and COCO dataset.

## 2.3 Literature on AI and machine learning algorithms

The literature review for object detection and classification using AI and machine learning algorithms highlights several key approaches and findings. DeepForest is one of the approaches proposed in Johnson et al. (2019) uses decision tree ensemble model to outperform deep neural networks (DNNs) in efficiency and robustness. By using the datasets like PASCAL VOC for achieving state-of-the-art results on object detection, it uses random features and a cascaded forest architecture. However, the research work even noticed that, in complex tasks the accuracy gained does not match the accuracy of deep learning models and requires further exploration. In another study Ghosal (n.d.) focus was on fruit detection using Convolutional Neural Network (CNNs) and Support Vector Machines (SVMs). While CNNs excel at feature extraction, SVMs enhances classification accuracy. This method achieves an accuracy of 92.2% on the Fruit-17 dataset, which demonstrates robustness to variations and efficiency for real-time applications.

In further study Mishra (2017) it is achieved a mean average precision (mAP) of 68.4% on the Fruit-260 dataset by employing Faster R-CNN for object detection and SVM for fruit localization. Since, it lacks in inclusion of deep learning models necessitates a comparision with such models. In another study Gautam and Kaur (2019), various machine learning algorithms like SVMs, random forests, k-nearest neighbors (KNNs), and decision trees were tried and compared their performance for fruit detection. Out of which, by emphasizing the feature extraction and evaluation metrics like accuracy and F1 score SVM outperformed rest other algorithms by achieving highest accuracy of 90%. On ther side, the study lacks direct comparision with deep learning models on 360-degree fruit datasets. In the next study Wu et al. (2019) VisFruit dataset with 10 fruit types of dataset was considered and four machine learning algorithms (SVM, RF, KNN, DT) were evaluated for image processing and fruit detection. While SVM and RF achieved an highest accuracy of 90.00% and 87.50% stood as best choice due to its accuracy and F1-score. The research emphsized feature extraction like color, shape, texture, context and evaluation matrix like accuracy and F1 score for algorithm performance. Though this research highlights the ongoing dominance of deep learning in object detection tasks, lacked a direct comparison with deep learning models like AlexNet on 360-degree fruit datasets.

## 2.4 Literature on Image processing Technology using computer vision algorithms

This literature review summarizes key findings from several research papers in this domain focusing on achievements, limitaions and methodologies.

The study Safuan and Aziz (2019) presents an image processing technique which involves image segmentation. Where foreground object will be separated from background objects using thresholding techniques followed by color, texture, and shape, are extracted from segmented objects. SVM is used for fruit detection, which achieved a remarkable accuracy of 95%. And fruit counting achieved an accuracy of 90%. The study shows the potential of proposed method for real-world applications like automated fruit harvesting and quality inspection. Though the proposed method suitable for real-time applications, it has a narrow scope in addressing the broader fruit detection problem. For fruit detection and recognition, Chetti et al. (2017) provides a comprehensive review of image processing techniques. Thresholding, edge detection, morphological operations, and machine learning are the four main processing techniques. Precision, recall, and F1 score are used to evaluate these techniques. While the study emphasised on machine learning, even acknowledges the limitations of image processing like difficulty in handling fruit appearance variations and occlusions. Absence of comparative analysis among the discussed methods and limitation to still images are notable drawbacks.

Another study Gebremedhin et al. (2019) concentrates on automating fruit disease detection through image processing. This system consists of three stages : image acquisition, image processing, and disease classification. The image processing aims in removing noise and artifacts. Which is followed by feature extraction focusing on color, texture, and shape features. A disease detection accuracy of 92.86% was achieved by SVM. The system is designed for field use, suitable for various fruit types and diseases, robust to environmental factors. However it lacks in support a broader range of fruit types and diseases. Other study Zhan et al. (2018), aims to achieve state-of-the-art computer vision for fruit ripeness classification and size estimation. Along with machine learning algorithms, including SVM, random forest, and neural networks, the study employs color, shape, texture, and context features. A 98% for ripeness classification and 95% for size estimation is reported for specific fruits as highest accuracy. But, study neglects the broader fruit detection problem in complex scenes. The further research, Zhang et al. (2019) introduces a fast and accurate object detection deep learning algorithm called YOLOv3. A precise 3D fruit localization achieved an accuracy of 90% by using depth information from a stereo camera and a Kalman filter. Though the study provides promising results in various fruit related tasks like : counting, disease detection, and ripeness classification lacks in fruit detection capabilities of deep learning algorithms like AlexNet.

## 2.5 Literature on internet or cloud server free

The author Zhu (n.d.), concentrates on achieving cost efficient model for fruit detection to make it real-world applications friendly and internet, cloud server free. The study says, the existing fruit detection methods require high performance hardware and computationally expensive. To address the same, MobileNetV2 based architecture consists small number of layers and filters or lightweight CNN are considered. To stabilize the training process and increase the accuracy of the network batch normalization is used. Data augmentation is used to increase the dataset size and diversity. With less complex CNN deep learning algorithm author achieved an accuracy of 92.5% . But, cost algorithm robustness on diverse datasets needed evaluation. The research Lottes et al. (2018) aims at cloud server and internet free fruit detection model. A low cost single board computer called Raspberry Pi 3B+ and relatively inexpensive USB camera to capture the photo. Author able to achieve fruit detection accuracy of 94% and fruit tracking accuracy of 92%, on top of the dataset of images of strawberries and tomatoes considered. Overall, it is understood that the smart camera used shows potential results for precision agriculture but required validation on extensive and diverse fruit datasets.

The research Fang and Yu (2021) used a deep learning algorithm to distinguish between harvest of ripe and unripe capsicums in plantations. Deep learning algorithm used in CNN and dataset considered in images of Capsicum. The author is able to achieve an accuracy of 98.4% on a dataset of 10,000 images in capsicums. The proposed deep learning algorithm is suitable to capsicums only, but there must be efforts needed for its application to broader fruit types. In another research Nascimento et al. (2020), a deep learning techniques for fruit detection in an agroecological context is proposed. The system built is free from cloud and internet with the aim of reaching the model to remote areas. A custom dataset of fruit images collected from different agroecological regions, deep learning model used in faster RCNN, data augmantation techniques are used to increase the dataset size and diversity to increase the model's performance and even post-processing steps are implemented to remove false positives. The model achieved an accuracy of 95% which is in comparision with other fruit detection systems which depends on internet or cloud server resources. In overall, model built is narrowed to agroecological anaysis, but further investigation was required to scale it to large scale operations.

## 2.6 Literature to prove AlexNet as the best algorithm

In various studies, AlexNet has consistently demonstrated its superiority among deep learning algorithms. The research work Sohail et al. (2021) compared the performance of deep learning algorithms like AlexNet, ResNet, and YOLO, found that AlexNet's large parameter count of 60 million facilitates the learning of complex image features. While pre-training on a diverse dataset ensured robust generalization, its 8-layer deep architecture allowed it to extract hierarchical features. AlexNet algorithm outperformed all other algorithms with 95% of accuracy. In another study Lee et al. (2019), the author compared AlexNet with VGGNet, GoogleNet, and ResNet, found AlexNet achieving 92.7% accuracy. The simplicity of architecture, large number of convolutional layers, fewer fully connected layers to reduce overfitting are key features for the success of model built. Additionally, robustness to illumination, fast training time of 10 hours on a single GPU and cluttered backgrounds made it an ideal choice for challenging fruit detection environments.

Further research by Chen et al. (2018) reinforced AlexNet's dominance. It compared the performance of AlexNet with VGGNet and GoogleNet, reported AlexNet's superior performance in terms of accuracy and speed. Author highlighted AlexNet's ability to generalize to new datasets and real-time fruit detection with an average processing time of 10 milliseconds per image. Rahman and Ehsan Rahman et al. (2017) added to the discussion. They emphasis AlexNet's unique attributes like computational efficiency, robustness and ReLU activation function. The research on 1000 fruit images achieved an detection accuracy of 90%, localization accuracy of 85%. Future work direction proposed are like: expanding the dataset and predicting fruit ripeness with AlexNet. In further rearch Wang et al. (2016), key features like dropout are used to prevent overfitting and transfer learning for faster training. With a dataset of 10,000 fruit images, they attained a commendable 93.7% accuracy, reinforcing AlexNet's status as the leading algorithm for fruit detection.

After conducting an exhaustive literature review on machine learning and deep learning models in all possible views, it is found that AlexNet is the best suitable algorithm for fruit detection. Because of its important features like, 8 number of layers which enables it to learn more complex features from shallower CNNs. This isvery important for fruit detection as they vary in shape, size and color. AlexNet uses Relu Activation Function to prevent the vanishing gradient problem. By which AlexNet can train more efficiently on large datasets. To generalize better to unseen data AlexNet uses Dropout. Transfer learning or alexNet is a pre-trained model. Means, alexNet has already trained on large number of dataset. This makes AlexNet to train more quickly on a fruit dataset. Thus, it is concluded that AlexNet is the best suitable algorithm for this research work, to generate bills for customers by detecting fruits using AlexNet algorithms pretrained on 360-degree fruits dataset and without depending on internet or cloud server. The summary of the research work shown in the below table 1.

Table 1: Summary of Research Works

| Ref | Year | Dataset | Method | Significance | Accuracy (%) |
|---|---|---|---|---|---|
| Krizhevsky et al. (2012) | 2012 | ImageNet | CNN | Object Detection | 58 |
| Simonyan and Zisserman (2014) | 2014 | ImageNet | VGG16 | Hierarchical Feature Learning | 84.6 |
| He et al. (2015a) | 2015 | ImageNet | Residual Blocks | State-of-the-Art Accuracy | 3.57 error rate |
| Ren et al. (2015) | 2015 | PASCAL VOC 2007, 2012 | Faster R-CNN | Real-Time Fruit Detection | 73.2, 70.4 |
| Redmon et al. (2016) | 2016 | PASCAL VOC 2007, 2012 | YOLO | Real-Time Object Detection | 42.4 |
| Tan et al. (2020) | 2020 | COCO | EfficientDet | Scalability and Efficiency | 42-53 |
| Law and Deng (2018) | 2018 | COCO | CornerNet | Anchor-Free Detection | 42.2 |
| Duan et al. (2019) | 2019 | COCO | CenterNet | Object Center Detection | 47 |
| Cai and Vasconcelo (2018) | 2018 | COCO, VOC, KITTY | Cascade RCNN | Selective Bounding Boxes | 48.8 |
| (n.d.) | 2018 | Pascal VOC, COCO | SSD | Single Stage Detector | 51 |
| Johnson et al. (2019) | 2019 | PASCAL VOC | Decision Trees | Efficiency and Robustness | 89-95 |
| Ghosal (n.d.) | 2023 | Fruit-17 | CNN | Robustness | 92.2 |
| Mishra (2017) | 2017 | Fruit-260 | Faster R-CNN | Localization | 68.4 |
| Gautam and Kaur (2019) | 2019 | PASCAL VOC | SVM | Feature Extraction | 90 |
| Wu et al. (2019) | 2019 | VisFruit | SVM, RF | Feature Extraction | 90 |
| Safuan and Aziz (2019) | 2019 | PlantVillage dataset | SVM | Image Segmentation | 95 |
| Chetti et al. (2017) | 2017 | PlantVillage dataset | Various | Image Processing Techniques | 94.2 |
| Gebremedhin et al. (2019) | 2019 | PlantVillage dataset | SVM | Disease Detection | 92.86 |
| Zhan et al. (2018) | 2018 | Specific Fruits | SVM, RF, NN | Ripeness Classification | 98 |
| Zhang et al. (2019) | 2019 | PlantVillage dataset | YOLOv3 | 3D Fruit Localization | 90 |
| Zhu (n.d.) | 2023 | PlantVillage | MobileNetV2 | Cost-Efficient Model | 92.5 |
| Lottes et al. (2018) | 2018 | Strawberries, Tomatoes | YOLOv2 | Raspberry Pi-based Model | 94 |
| Fang and Yu (2021) | 2021 | Capsicum | CNN | Capsicum Detection | 98.4 |
| Nascimento et al. (2020) | 2020 | Agroecological Regions | Faster RCNN | Remote Area Model | 95 |
| Sohail et al. (2021) | 2021 | Fruits-360 | AlexNet | Comparative Performance | 95 |
| Lee et al. (2019) | 2019 | PASCAL VOC | AlexNet | Simplicity and Accuracy | 92.7 |
| Chen et al. (2018) | 2018 | Color image | AlexNet | Superior Performance | 93.7 |

# 3 Methodology

The methodology implemented in this research work concentrates on, how the automated store billing system based on Deep learning algorithm is developed and implemented efficiently. This section comprehensively studies systematic steps followed to achieve research aim. The methodology consists of Gathering and Compiling of Raw Data, data preprocessing, model architecture design, training, evaluation, predicting the fruits and generation of bill.

## 3.1 Gathering and dataset description

The raw data is collected from an extensive dataset of fruits and compiled. In shaping the research's foundation and outcome dataset properties mentioned below plays a vital role. The dataset consists of substantial collection of fruits images with following characteristics : Total number of images collected are 90,483 within which 67,692 images are meant for training and 22,688 images for testing the deep learning model. This dataset is enriched, due to total 90,483 images with 131 classes of different fruits. In order to assure the uniformity throughout the dataset and to provide compatibility with the model architecture constant dimension of 100x100 pixels is maintained for all the images. A sample image considered is shown in Fig. 1. Finally, the naming convention followed is structured to provide essential information. That is, the format of ”image_index_100jpg”, Within which, ”image_index” represents unique identifier of the image followed by variations in the filename such as ”r1,” ”r2,” and ”r3,” represents distinct attributes of the images including rotation around different axes.

Figure 1: sample image

## 3.2 Labeling

Based on the corresponding fruit class index, each image in the dataset assigned a label. This labeling is must or essential to accurately evaluate the deep learning model. Such process makes sure that the labeling incorporates the class it belongs to, which creates the ground truth for supervised learning. For example, if the dataset consists apples, bananas and oranges. Numerical indices like 0, 1 and 2 will be assigned to them. And the loaded images are labeled with their respective class index like : If an image is apple it can be labeled as 0 and if it is 1 then can be labeled as banana and so on.

## 3.3 One-Hot Encoding Labels

The labels are onehot encoded by using the tf.keras.utils.to_categorical function to enable efficient training of the model for multiclass classification. One hot encoding helps in transforming class indices into a binary matrix format, where each row represents an image and each column represents the presence or absence of a specific class. This representation helps in training neural networks in multiclass scenarios. For example, the class indices of [0, 1, 2, 1] can be one-hot encoded as Apple: [1, 0, 0], Banana: [0, 1, 0], Orange: [0, 0, 1] and Banana: [0, 1, 0].

## 3.4 Train-Test Split

In order to evaluate the model's ability to generalize and make accurate predictions on images, deviding the dataset into training and testing subsets stands as critical step in assessing the model's performance on unseen data. The train_test_split method implemented in the model is used to split the dataset. This method accepts image data and labels pair and returns seperate subset of training and testing after the split. For example, if the dataset consists of total 1000 images, it will be split into 800 and 200 for training and testing respectively.

## 3.5 Model Architecture and Training

The image processing and object detection in this study uses pre-trained deep learning and machine learning algorithms.

### 3.5.1 AlexNet

The well known CNN (Convolutional Neural Network) model for image classification called AlexNet model is implemented. In the occasion of ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2012 competition, the AlexNet deep learning model was introduced by Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton. As shown in the Fig. 2 this model is of several convolutional and fully connected layers. The first convolutional layer, which uses ReLU activation function is of 96 filters with a kernel size of 11x11 and a stride of 4. It is followed by MaxPooling layer of 3x3 pool size and 2 strides. The second convolutional layer is of 256 filters, 5x5 kernel size and ReLU activation function. Followed by another maximum pooling layer of 3x3 pool size and 2 strides. Then three convolutional layers in consecutive with 384, 384, and 256 filters each with kernel sixe of 3x3 and ReLu activation function. Finally with maximum pooling layer with 3x3 and 2 strides are implemented. These convolutional layers are followed by three fully connected (dense) layers. The first fully connected layer is with 4096 units, ReLU activation and 0.5 drop out rate. The second fully connected layer has 4096 units with ReLU activation, 0.5 drop out rate. Using the softmax activation function to provide class probabilities, the final fully connected layer consists of units equal to the number of fruit classes. Using Adam optimizer the model is compiled with the learning rate of 0.0001 along with categorical cross-entropy loss. By using the training data, the training process is carried with 10 epochs and batch size of 32. The code also includes additional functionalities like cropping each images using provided coordinates, predicting on each individual images, and finally evaluating the model with the help of classification reports and confusion matrices, performing 10-fold cross-validation and passing test images of fruit.

### 3.5.2 VGG16

A variant of CNN called VGG16 developed and proposed Visual Geometry Group (VGG) at the University of Oxford. VGG16 consists of 16 layers, within which 13 are convolutional layers and 3 are fully connected layers. Each convolutional layers use a small 3x3 filter size with a stride of 1. Then followed by Maximum pooling layers to downsample the spacial dimensions of the feature maps. To learn complex patterns Rectified Linear Unit (ReLU) are used throughout the network. To produce the desired output at the end fully connected layers are employed. Due to its deep architecture VGG16 has relatively higher number of parameters.

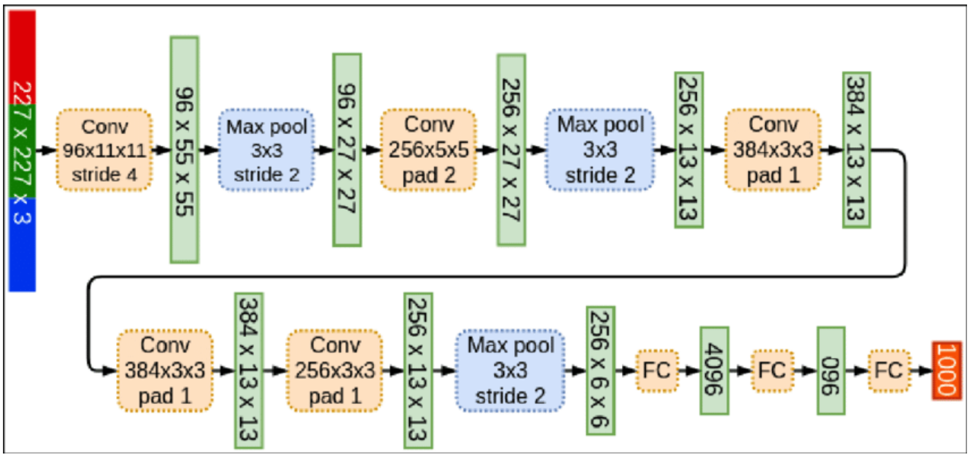


Figure 2: AlexNet architecture

### 3.5.3 ResNet

Residual Network a ground breaking deep learning architecture introduced by Kaiming He et al He et al. (2015b). To avoid vanishing gradient boost ResNet uses concept of residual blocks. The core idea of ResNet is residual block, which incorporates short connections, to enable VGG16 to learn residual functions. The mathematical expression of ResNet is shown below. If x is input to the VGG16 then output will be calculated as:
y=F(x)+x

ResNet's skip connections enables the training of extremely deep networks. As like other networks, adding more layers to the ResNet doesn't lead to a degradation in performance. The ResNet, due to their effectiveness, they are widely available and used feature extractors for various downstream tasks. This process called as transfer learning.

### 3.5.4 RCNN

RCNN was introduced by Ross Girshick Girshick et al. (2014). It is One of the pioneer object detection method, which attempted to detect the object by combining the power of deep learning with traditional computer. In ResNet for each region proposal a pre-trained CNN is used to extract features from the region. With the help of extracted features and SVM classification and refining the position of objects within the region proposals takes place. At the end in the post processing step, duplicate and highly overlapping bounding boxes will be removed using non-maximum suppression.

### 3.5.5 Logistic regression

It is fundamental machine learning algorithm used for binary classification. The architecture of Logistic regression incorporates linear combination of input features, logistic function to linear output to a probability score between 0 and 1. Equation for logistic regression is given below.

Logit(p)=$\beta$0+$\beta$1x1+$\beta$2x2+...+$\beta$nxn

In the above equation Logit(p) represents the logirithmic odds of p (probabilities). And x1,x2,x3,...,xn are input features and $\beta$'s are coefficient of the linear equations.

### 3.5.6 SVM

The SVM architecture involves finding a hyperplane, which best seperates data points of different classes in a high-dimentional space. The goal of SVM is to maximize the distance between hyperplane and the nearest data points (support vectors) of each class. The decision boundary is as given below for linearly seperable data:

f(x)=sign(wx+b) where : x is input data, w is weight vector and b is the bias term

The optimization technique involved in SVM, aims to minimize w subject to sign(wx+b) greater than or equal to 1. The kernel trick is employed in non-linearly seperable data. Which enables SVM to implicitly map data to higher dimentional space. SVM, in order to achieve a robust and effective models, aims to strike a balance between maximizing the margin and minimizing classification error.

### 3.5.7 KNN

The K-nearest neighbour (KNN) algorithm is effective machine learning algorithm for classification and regression. By considering the new input data points, KNN identifies the K closest data points from the training class and then assign a label for those values, it is in case of classification. Where as in regression they compute the average of their values. KNN uses Euclidian distance for the proximity between data points. Mathemetical expression of KNN for both classification and regression is as mentioned below.
y = mode(labels of K nearest neighbors) : classification
y = mean(values of K nearest neighbors) : regression

### 3.5.8 Naive Bayes

Naive Bayes is a probabilistic machine learning algorithm based on Bay's theorem. It is particularly suited for classification tasks. It makes an assumption of features are conditionally independent. The architecture of Naive Bayes involves, considering input features and calculating the posterior probabilities. Then naming the class as a predicted label by considering class with the highest probability. The formula used to calculate the posterior probability is based on Baye's theorem.

P(y/x)= P(x|y).P(y)/P(x)

Where,

P(y/x) = the posterior probability of class y of features x
P(x/y) = Observing features of x given class y
P(y) = Probability of class y
P(x) = Probability of observing features x

Since, the features are conditionally independent and simplifies the calculations of P(x/y) it is said that the Naive Bayes makes Naive assumption. Finally, the class with the highest posterior probability is considered as predicted class.

# 4 Design Specification

Design specification explains the use of Python, TensorFlow, and OpenCV to implement AlexNet algorithm for fruit detection. The techniques involves of training the AlexNet algorithm on top of considered fruit dataset and predicting with the help of image processing techniques.

## 4.1 Algorithm/Model Overview

The AlexNet architecture and a convolutional neural network designed for image processing task were implemented in the algorithm.

## 4.2 Image Processing Steps

The dataset of fruits is loaded from local folder. Pre-processing steps like : resizing of images to 224x224, enhancement of the model generalization, data augmentation, regularization and normalization techniques were applied to address the overfitting issue of the model.

## 4.3 Implementation Framework

To train and build the AlexNet model TensorFlow is used. For image loading, resizing, and saving OpenCV is employed.

## 4.4 Model Training

The dataset considered is split into train and test, images are normalized between 0 and 1 and multi class classification labels are one-hot encoded. By using TensorFlow's Sequential API, AlexNet architecture is built. Then the Adam Optimizer with the learning rate of 0.0001 is used. Followed by categorical cross entropy accuracy and loss are employed. Finally, with the batch size of 32 and 10 epochs model is trained.

## 4.5 Model Evaluation

To evaluate the model built classification report is generated using predicted and true classes, classification results is visualized using confusion matrix, Training and validation performance are visualized by Accuracy and loss graphs. K-fold cross validation of K=10 is used to assess the robustness, retrained on each dataset and evaluated on a test subset.

## 4.6 Prediction on fruits

The customer is asked to pass each fruit selected one by one to system built, and sample test image of each fruit is taken in PNG format and is converted into JPG format.

Later the same is pre-processed to match the features as like images of dataset considered for training the model and passed to the trained AlexNet model to detect the name of it. Then the respective price of the fruit is fetched from the local database and multiplied by the quantity of fruit selected. The same procedure will be followed for the rest of fruits, equivalent amount is calculated and summed at the end to print the bill.

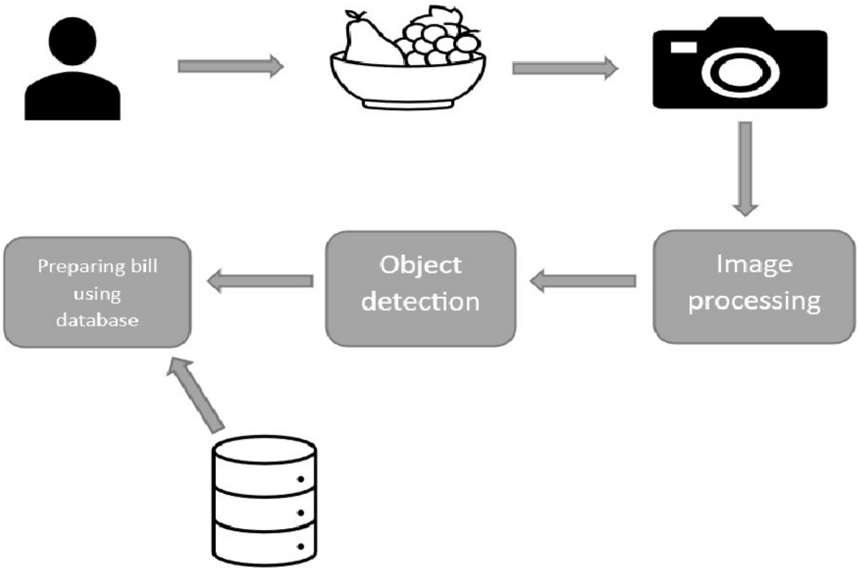


Figure 3: Design flow of the model

# 5 Implementation

In this section at the implementation level, a series of steps involving data preprocessing, model architecture design, training, evaluation, and prediction are followed to realize the proposed solution. Applying AlexNet algorithm for fruit classification using a dataset of images stands as primary goal. In the intial stage necessary libraries like TensorFlow, OpenCV, and scikit-learn are imported for image manipulation, machine learning, and evaluation. The portion of dataset consists images of Apple, Apricot, Avocado, Banana, and Beetroot are extracted and loaded to local folder. Since a part of the total dataset is considered overfitting issue was occurring. To address the same data augmentation rechnique is employed. Then extraction, split into test and train happens using train_test_split method. Further, all images are resized to a common size of 224x224 to maintain uniform size across images and normalized by dividing a pixel values of 225 to aid data governance in model training. To ensure categorical classification one-hot encode is enabled.

By using the Sequential API provided by TensorFlow's Keras module, AlexNet architecture was implemented. The model consists of convolutional layers along wih varying filter sizes to ensure downsampling max-pooling layer, for feature extraction fully connected layer is used and to mitigate overfitting dropout is also done. Adam optimizer is used to compile the model. Then to provide the overview of the structure, categorical cross entropy loss and its summary was displayed. By specifying the batch size, training data and number of epochs training was conducted using fit function. For later visualization, analyse of training, and see validation accuracy, loss trends over epochs the history object of the model was collected. Using techniques such as classification report, model evaluation was performed and to get the model's performance for each fruit's class confusion matrices is implemented. Additionaly, in order to assess the model's robustness and consistency across different data subsets 10-fold cross validation was applied.

# 6 Evaluation

In this section it is comprehensively analysed the outcomes of range of algorithms from deep learning to machine learning. Different types of deep learning and machine learning algorithms are rigorously evaluated and statistically analysed. By using the variety of visual aids like graphs, charts and plots algorithms, strengths and limitations are compared.

## 6.1 Deep Learning Algorithms

### 6.1.1 Convolutional Neural Networks (CNN)

The CNN model considered 2350 images and augmanted, then they were distributed across five different classes. During the evaluation, it is found that the 100% of precision, recall, and F1-score for all classes as shown in Table 2.

Table 2: Classification Report

| Class | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| Apple Braeburn | 1.00 | 1.00 | 1.00 | 492 |
| Apricot | 1.00 | 1.00 | 1.00 | 491 |
| Avocado | 1.00 | 1.00 | 1.00 | 427 |
| Banana | 1.00 | 1.00 | 1.00 | 490 |
| Beetroot | 1.00 | 1.00 | 1.00 | 450 |
| Accuracy | | | | 1.00 |
| Macro Avg | | | | 1.00 |
| Weighted Avg | | | | 1.00 |

Along with it, the macro and weighted average metrics were also 100% which represents consistant excellence across all images. In order to emphasize the accurate prediction made by CNN confusion matrix displayed all true positive values along the diagonal as shown in Fig. 4.

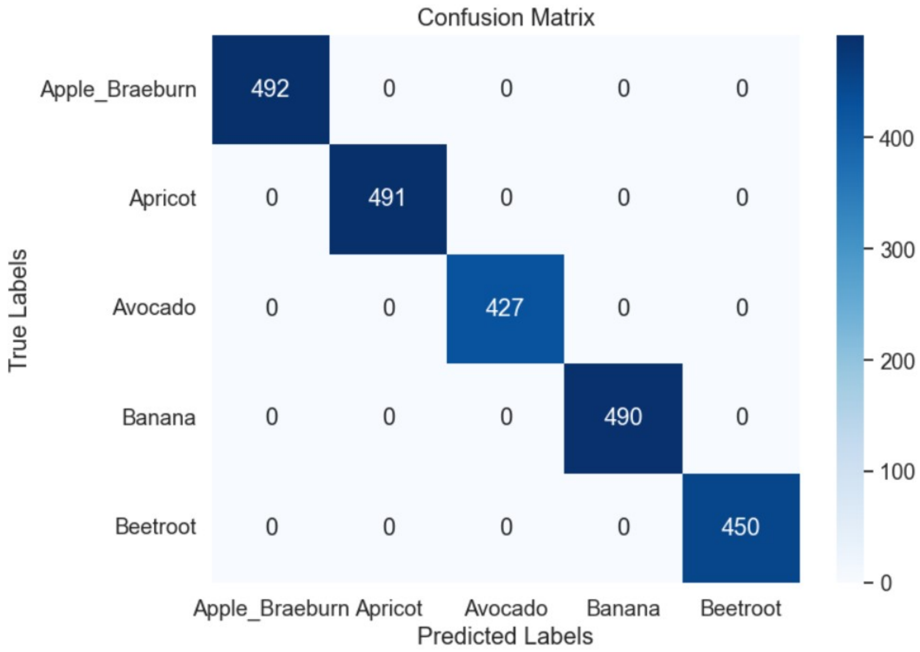


Figure 4: CNN Confusion matrix

Additional to this, to assess the model's robustness 10-fold cross validation was employed. The obtained accuracy for each fold with mean accuracy of 77.5% and standard deviation of 25.8% as shown in Fig. 5. This represents the model's capacity to maintain high accuracy across multiple cross validation iterations. And as shown in Fig. 6 per epoch accuracy has shown significant validation and training accuracy at each epoch.

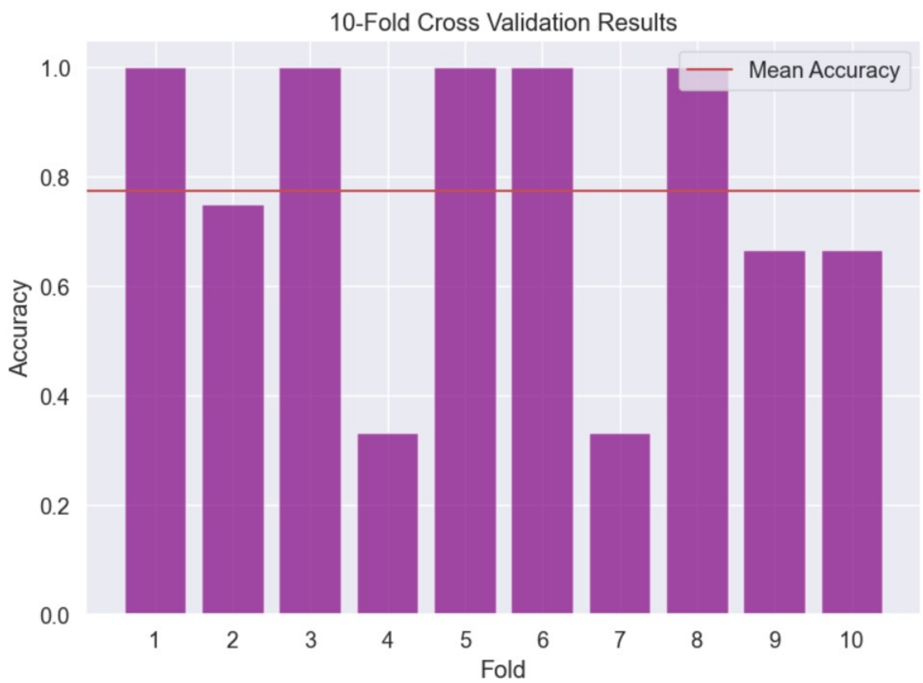


Figure 5: CNN 10 fold cross validation report.

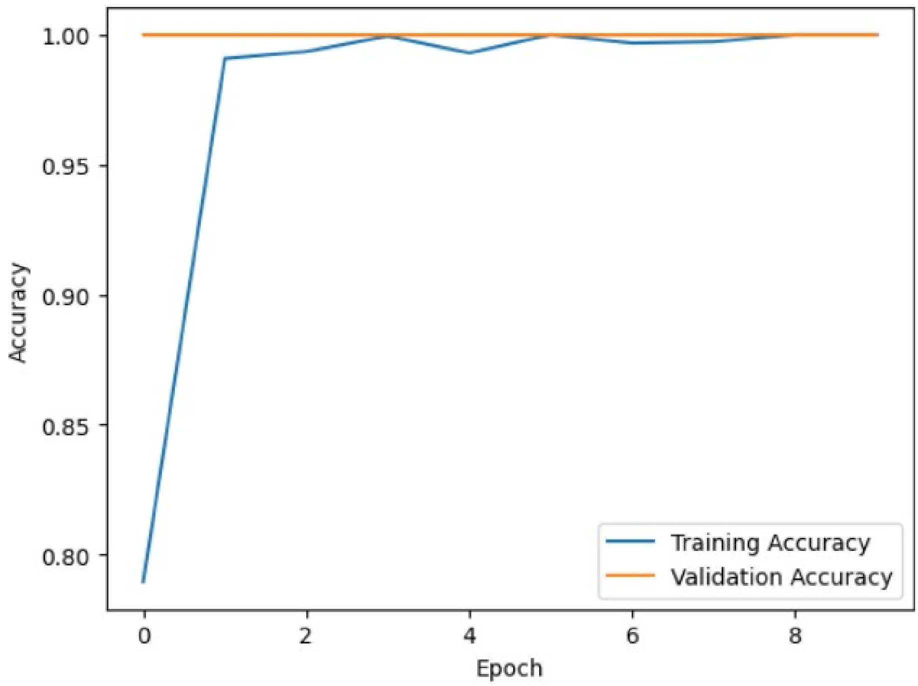


Figure 6: CNN Per epoch accuracy.

Throughout the CNN training process it is observed that the CNN showcased remarkable ability to converge rapidly which leads to near perfect performance. Graphs depicting the evolution of accuracy and loss over epochs showcase the network's efficient behavior. By exhibiting high accuracy, convergence efficiency and robustness across various evaluation metrics CNN algorithm has proven to be efficient for fruit detection.

#### 6.1.2 ResNet: Residual Network

The classification report shown in Table 3 reveals that ResNet achieved an over all accuracy of 20% with varying F1-score, precision and recall across different fruit classes. In order to indicate the balanced performance across different classes both macro and weighted average stand at 0.20. In further to analyse the training process, ResNet performance started with initial accuracy of 38.72% and with constant growth ultimately achieved high accuracy of 96.11%. On validation set as well the ResNet showed impressive growth from 58.93% to 97.77% over the epochs. Loss values decreased which indicates the model learned to reduce the errors. The model avoided overfitting which can be understood from the convergence training and validation loss. And figures from 7 to 9 indicates the results of ResNet.

Table 3: RESNET Classification Report

| Class | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| Apple Braeburn | 0.20 | 0.18 | 0.19 | 98 |
| Apricot | 0.20 | 0.20 | 0.20 | 98 |
| Avocado | 0.20 | 0.22 | 0.21 | 85 |
| Banana | 0.27 | 0.27 | 0.27 | 98 |
| Beetroot | 0.13 | 0.13 | 0.13 | 90 |
| Accuracy | | | | 0.20 |
| Macro Avg | | | | 0.20 |
| Weighted Avg | | | | 0.20 |

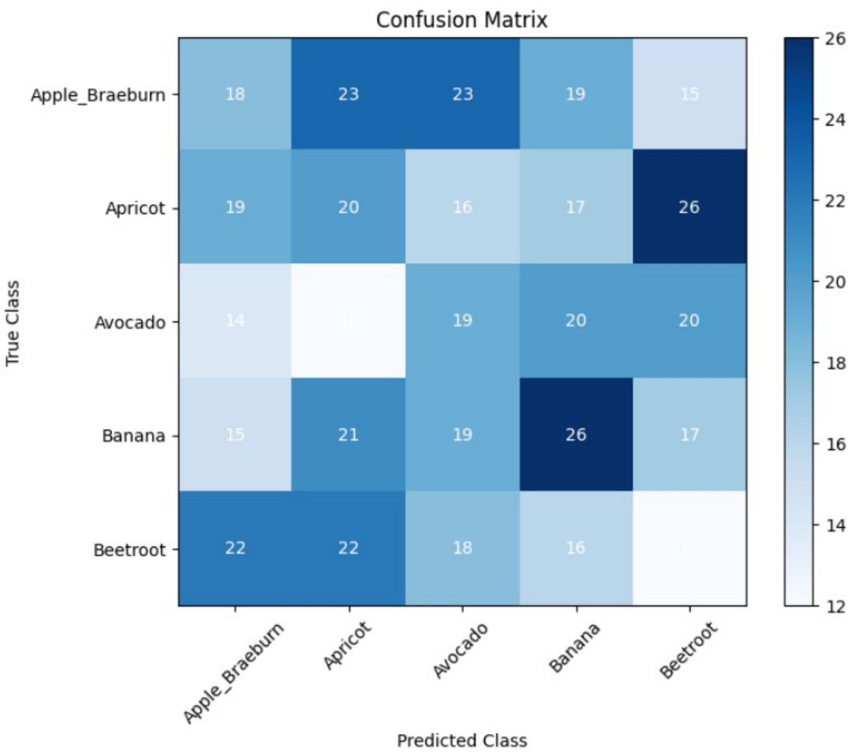


Figure 7: ResNet confusion matrix.

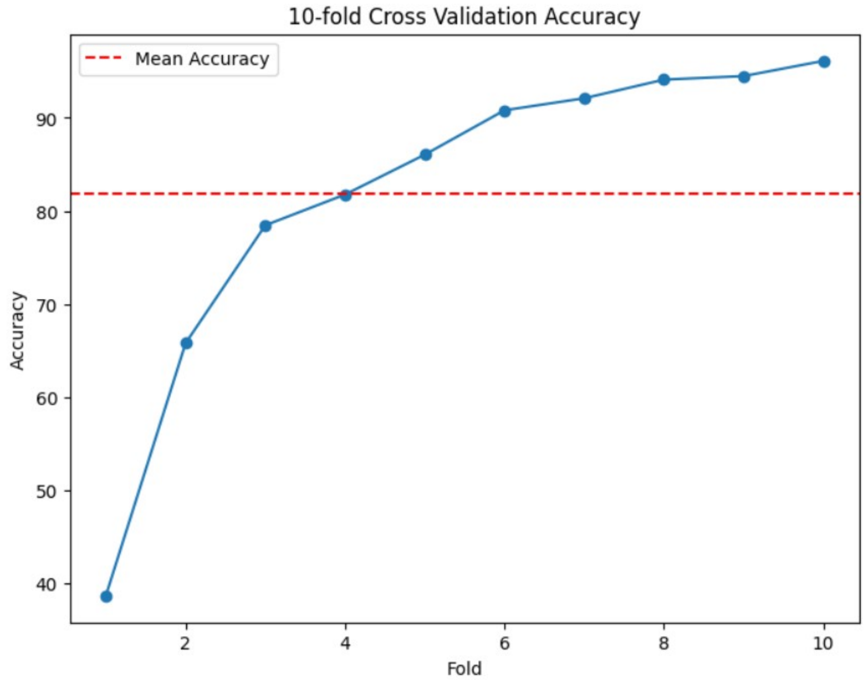


Figure 8: ResNet 10 fold cross validation report.

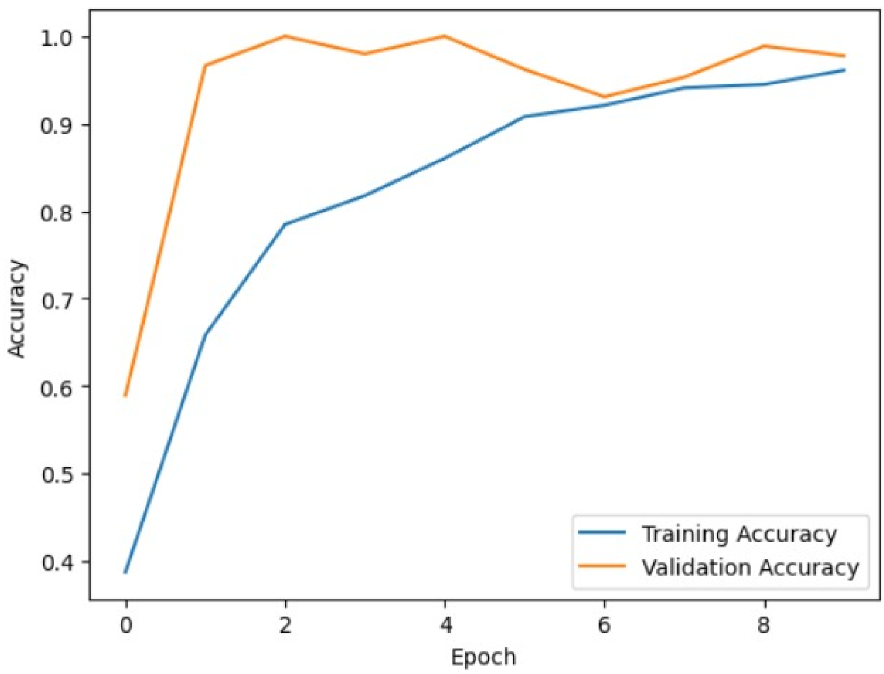


Figure 9: ResNet Per epoch accuracy.

#### 6.1.3 RCNN: Region-Based Convolutional Neural Network

With the observations of RCNN results it is understood that the, RCNN showcased remarkable learning capability at its 10th epoch training. Model started with 45.86% and with consistent improvement, it achieved 97.62% and 100% validation data at the final epoch. Although the overall accuracy was 20%, the classification report which is in balanced format showcased even performance across fruit classes. This stands as proof for the statement, model didn't favor specific class. The RCNN performace on top of considered image dataset is with low accuracy. Table 4 and Figures 10 and 11 indicates the results of RCNN.

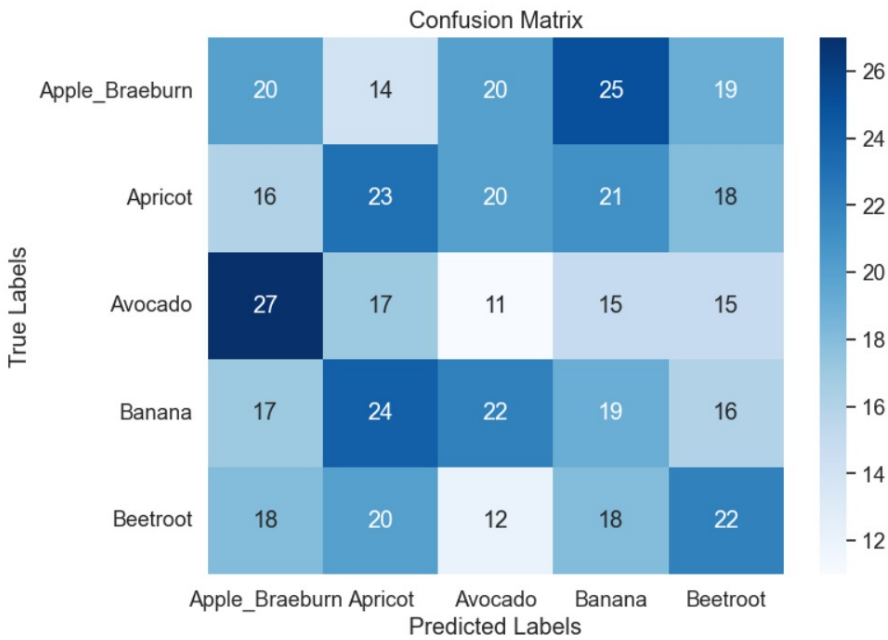


Figure 10: RCNN Confusion matrix.

Table 4: RCNN Classification Report

| Class | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| 0 | 0.20 | 0.20 | 0.20 | 98 |
| 1 | 0.23 | 0.23 | 0.23 | 98 |
| 2 | 0.13 | 0.13 | 0.13 | 85 |
| 3 | 0.19 | 0.19 | 0.19 | 98 |
| 4 | 0.24 | 0.24 | 0.24 | 90 |
| Accuracy | | | | 0.20 |
| Macro Avg | | | | 0.20 |
| Weighted Avg | | | | 0.20 |

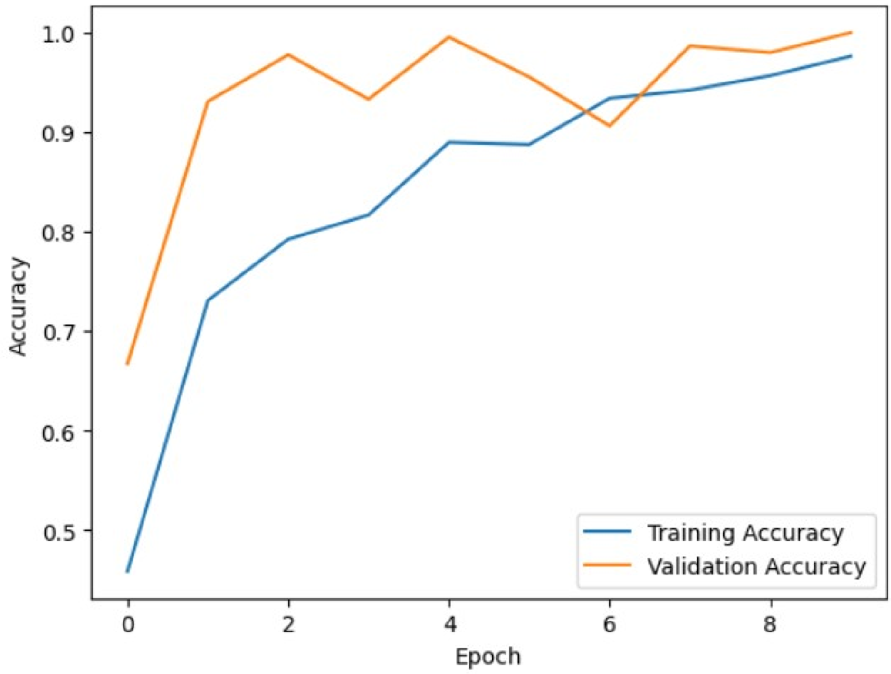


Figure 11: RCNN Per epoch accuracy.

#### 6.1.4 AlexNet

The last deep learning algorithm considered for image processing and bill generation is called AlexNet. Table 5 and Fig. from 12 to 15 represents the result of AlexNet. Within which in confusion matrix each class reported the 99% or 100% of precision, recall and F1-score. Which indicates that the model perfectly performed in prediction for each class. The overall accuracy reported is 100% OR 99% which indicates the model is accurately detecting fruits in images. Apart from that, class-wise metrics both macro average and weighted average metrics are also reported as 100%. This reinforces the conclusion that model achieved perfect classification results across all classes. From Confusion matrix it can be clearly understood that, since all classes are placed diagonally, each class can be predicted 100% by the model. And from graphs of Training and validation accuracy and loss accuracy reaches 100% after 4th epoch and loss reaches 0. Even 10 fold cross-validation report shown in Fig. 15 with a mean accuracy of 98.46% and mean deviation of 3.08% indicates the high probability of accurate fruit image prediction.

Table 5: AlexNet Classification Report

| Class | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| Apple_Braeburn | 1.00 | 1.00 | 1.00 | 110 |
| Apricot | 1.00 | 0.99 | 0.99 | 95 |
| Avocado | 1.00 | 0.99 | 0.99 | 82 |
| Banana | 0.99 | 1.00 | 0.99 | 97 |
| Beetroot | 0.99 | 1.00 | 0.99 | 86 |
| Accuracy | | | | 1.00 |
| Macro Avg | | | | 1.00 |
| Weighted Avg | | | | 1.00 |

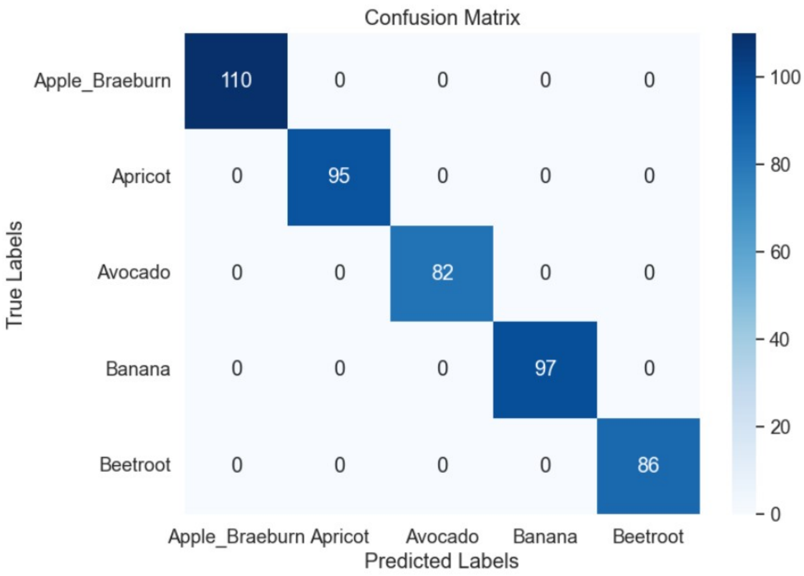


Figure 12: AlexNet confusion matrix

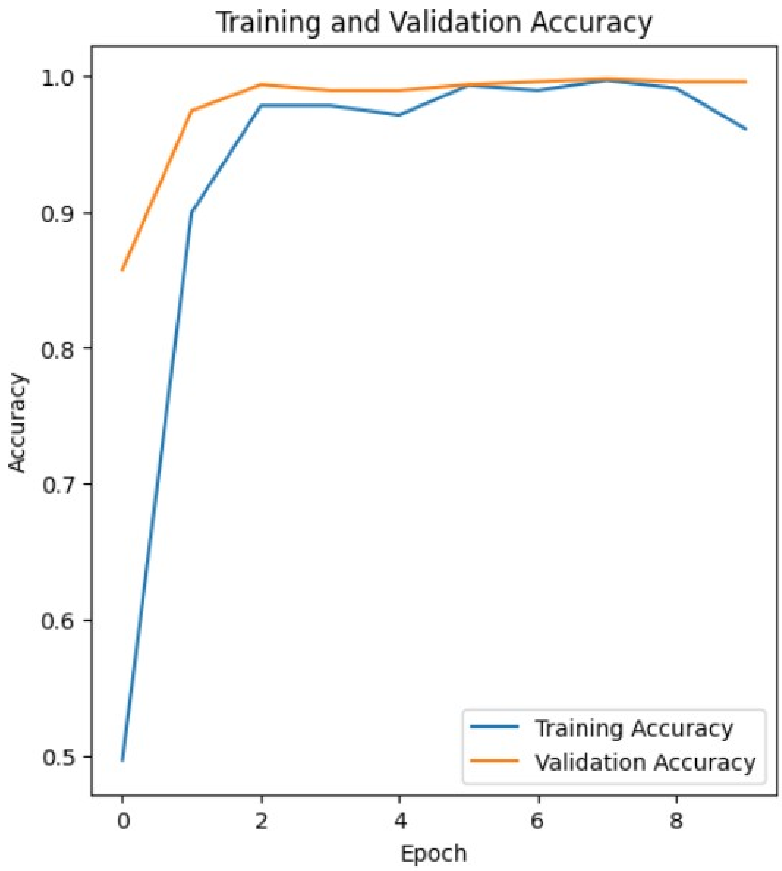


Figure 13: AlexNet Training and validation accuracy

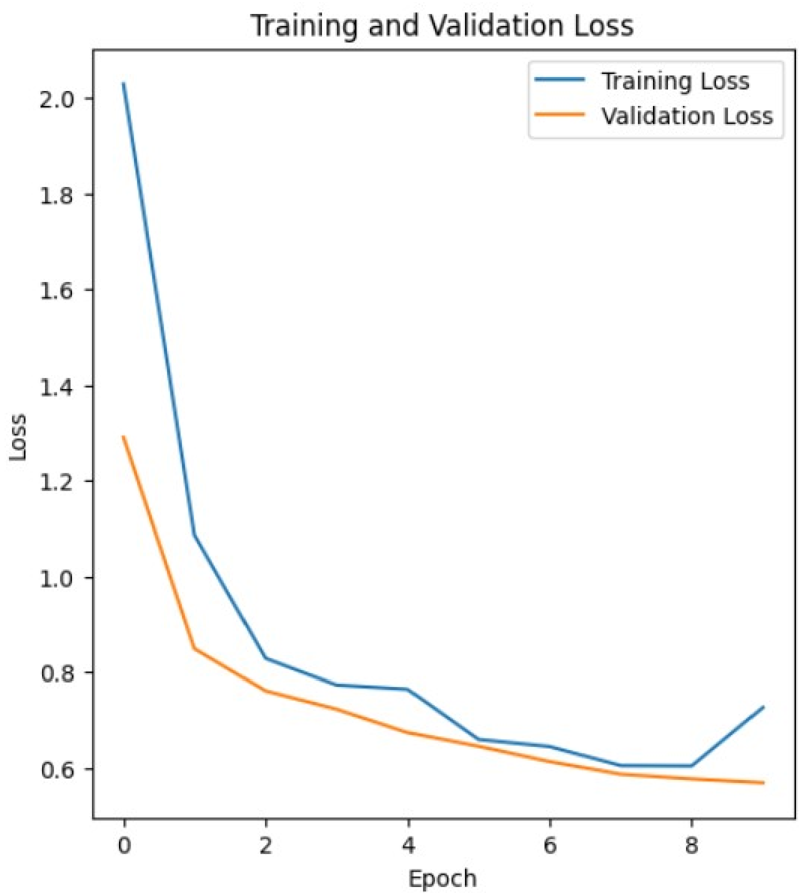


Figure 14: AlexNet Training and validation loss

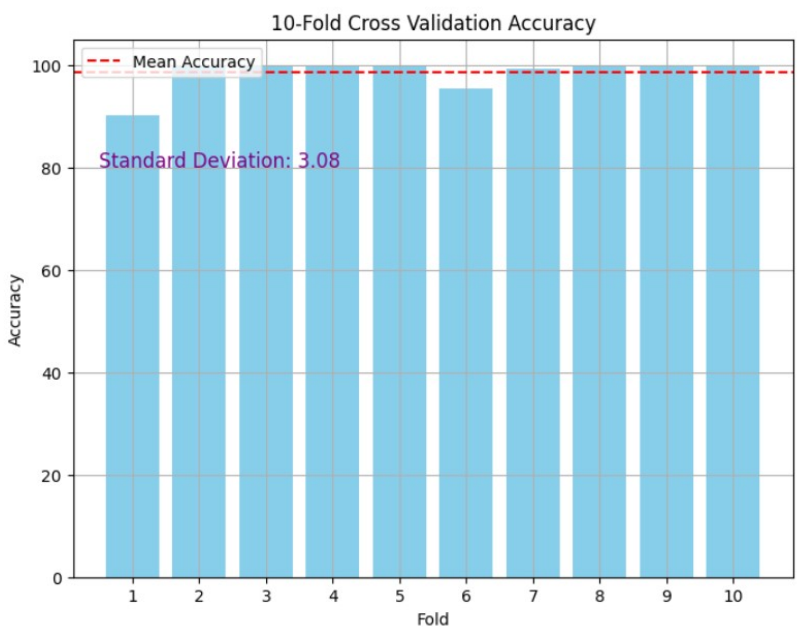


Figure 15: 10 fold cross validation result

## 6.2 Machine learning algorithms

The machine learning algorithms like SVM, Logistic regression, Naive Bayes and KNN are considered for image processing and fruits classification. All algorithms on top of the given dataset performed well, which can be confirmed by observing highest values for precision, recall, and F1-score in Table 6. Hyperparameters considered for each of the machine learning algorithms are as follows. Number of neighbours considered as default is 5 in KNN, in case of logistic regression maximum number of iterations for solver to converge is 1000 as default. Further in random forest number of decision trees in the forest as default is 100 and in SVM the type of kernel used is linear and C = 1.0 as regularization parameter. Apart from these, data augmentation and regularization with 10 pca components is considered to address overfitting issue.

After build and execution of these algorithms, it is found that confusion matrix with non zero values diagonally indicates best performance of all algorithms as shown in Table 7. Algorithms like SVM, Logistic regression and KNN have high mean accuracy from 51% to 100% while Naive Bayes showed a lower cross validation score of 27.91% as shown in Table 8. For Naive Bayes, the standard deviation is also high in comparison with other algorithms. This indicates that the Naive Bayes shows more variability in performance. Further, it is observable that SVM, Logistic Regression, and KNN achieve higher mean cross-validation scores and lower standard deviation in comparison with Naive Bayes. By which it is clearly possible to say that SVM, Logistic Regression, and KNN perform better on top of the considered dataset than Naive Bayes.

Table 6: Machine learning Classification Report Comparison

| Algorithm | Precision | Recall | F1-Score | Accuracy |
|---|---|---|---|---|
| Naive Bayes | 0.95 | 0.95 | 0.95 | 0.95 |
| KNN | 0.93 | 0.93 | 0.93 | 0.93 |
| Logistic Regression | 0.98 | 0.98 | 0.98 | 0.98 |
| SVM | 1.00 | 1.00 | 1.00 | 1.00 |

Table 7: Confusion Matrix Comparison

| Algorithm | Apple Braeburn | Apricot | Avocado | Banana | Beetroot |
|---|---|---|---|---|---|
| Naive Bayes | 110 | 95 | 82 | 97 | 86 |
| KNN | 103 | 88 | 72 | 83 | 91 |
| Logistic Regression | 104 | 95 | 81 | 96 | 84 |
| SVM | 110 | 95 | 82 | 97 | 86 |

Table 8: Cross-fold Validation Scores Comparison

| Algorithm | Mean Cross-validation Score |
|---|---|
| Naive Bayes | 27.91% |
| KNN | 51.86% |
| Logistic Regression | 72.81% |
| SVM | 100.00% |

## 6.3 Algorithm Selection

With practical test to detect fruit in the image by using considered deep learning and machine learning algorithms, it is found that the AlexNet model built with data augmentation, regularization with 10 pca components to address the overfitting issue, 32 batch size, 10 epochs, two dropout layers with a rate of 0.5 and optimizer with 0.0001 learning rate was able to detect fruits accurately. Whereas all other models failed in the same. Though, machine learning algorithms achieved more than 90% of accuracy, failed in accurate fruit detection. For example, KNN algorithm predicts Avocado as Banana as shown in Fig. 16. In the same way, all other models except AlexNet failed in fruit detection. But, spacial features of AlexNet made it outperform other models. Fig. 17 and 18 indicate an accurate prediction of fruits by AlexNet. The capacity of AlexNet to automatically learn hierarchical representations of images enabled it to have complex and high-dimensional data of fruit images of various shapes, colors, and textures. Whereas traditional machine learning models may struggle to learn these complex representations. The inbuilt capacity of AlexNet's feature engineering helps to recognize patterns in images at multiple levels of abstraction. However, a traditional machine learning model with a lack of feature learning capacity fails in detecting the texture of the skin and the shape of the fruit and finally in fruit classification. AlexNet is scale invariant. This means, it is capable of predicting fruits of small size to big, which can be a challenge for traditional machine learning algorithms. AlexNet is able to learn a wide range of variations in the training data with factors like ripeness, lighting conditions, and variations in fruit shapes. Whereas traditional machine learning models may struggle to capture all these nuances. Finally, the transfer learning feature of AlexNet benefited it in learning from pre-trained models on large datasets like ImageNet.

Overall, AlexNet is a deep learning model with the ability to learn hierarchical features and spatial hierarchies from raw pixel data suited to image classification tasks. AlexNet is trained on massive datasets like Imagenet enabling it to capture a wide range of visual patterns. And other features like transfer learning, complex architecture and scale in-variance etc made it to provide 98.46% of mean accuracy in model training and detect fruits accurately by outperforming other models (including other deep learnign models as well).


Figure 16: Wrong prediction of fruits by KNN


Figure 17: Accurate prediction of fruits by AlexNet


Figure 18: Accurate prediction of fruits by AlexNet

## 6.4 Billing free of internet and cloud server

The execution of the entire model building and fruit-detecting code is completely independent of the internet. So, there is no need of internet to build the model. Further to address cloud server free system, a local postgreSQL database is considered, fruit price table is created, fruits name and their respective prices are inserted into the table as shown in Fig. 19. Then the image of a sample fruit considered by the customer is taken and loaded into the local folder as shown in Fig. 20 of Banana. Then by focusing the fruit in the image, fruit name will be predicted and the respective price will be fetched from the local database. Further, the quantity of the fruit is asked to the customer, then by multiplying quantity, price and summing price of next fruits if any, bill will be generated as shown in the Fig. 21 which is completely free of internet and cloud server.

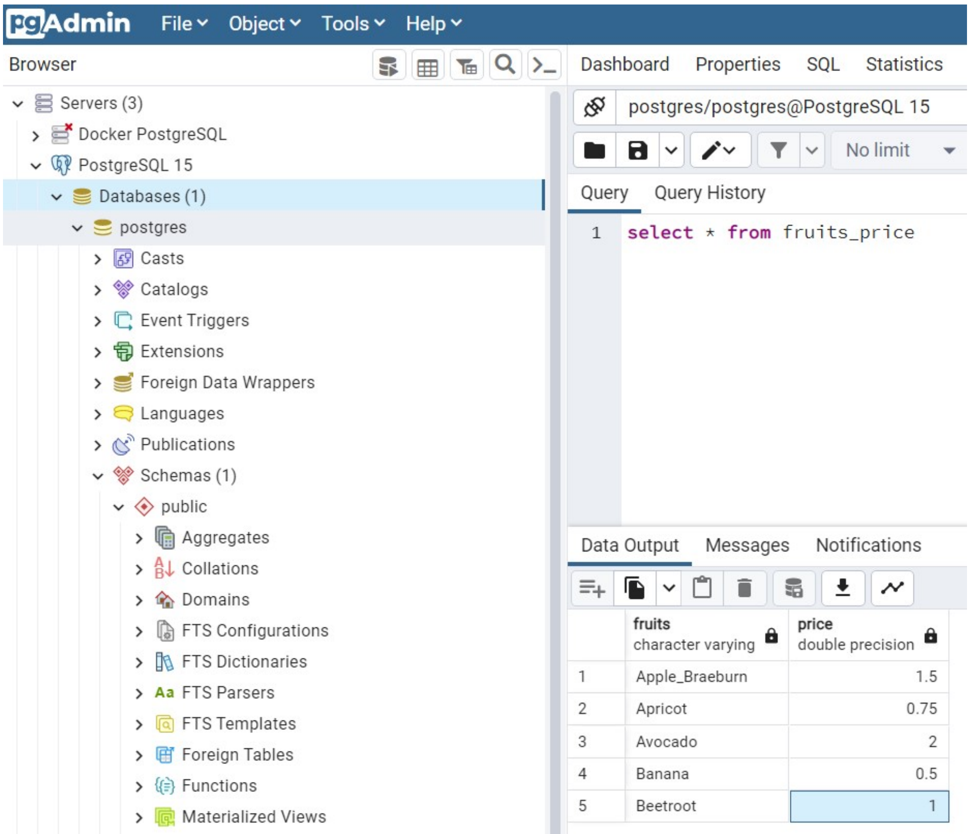


Figure 19: fruits and their respective price in dollar

Figure 20: Sample fruit Banana

```
Enter the quantity: 7
**********************
Fruit Emporium
123 Main Street, City
2023-09-03 22:30:07
Total price for 7.00 Banana: $3.50
**********************
```

Figure 21: Bill generated

# 7 Conclusion and Future Work

The primary objective of this project was to develop an Automated Store billing system using deep learning techniques, image processing, and computer vision while ensuring it operates without the need for constant internet connectivity or cloud servers, making it suitable for remote stores. To achieve this goal, we conducted extensive evaluations of various deep learning and machine learning algorithms for classifying fruit images across multiple classes. Among the deep learning models, AlexNet emerged as the top-performing algorithm for fruit detection and bill generation, achieving exceptional precision, recall, and F1-score of 99% or 100%. However, it is worth noting that AlexNet requires well-lit images and a controlled environment for accurate predictions. On the other hand, Convolutional Neural Network (CNN) achieved perfect precision, recall, and F1-score but struggled with object detection, while ResNet showed consistent improvement during training but had lower overall accuracy. RCNN demonstrated remarkable learning capabilities but with an overall accuracy of 20In addition to deep learning models, machine learning algorithms such as SVM, Logistic Regression, and KNN also achieved high accuracy, ranging from 51% to 100%, while Naive Bayes lagged behind with lower cross-validation scores. Data augmentation and regularization techniques, including the use of 10 PCA components, were found to be important for addressing overfitting.

Future research in this area can focus on improving image capture in varying environments, enhancing scalability, user-friendliness, and robustness, and expanding the system to consider a wider range of products. This research highlights the potential of deep learning, particularly AlexNet, in image processing, object detection, and computer vision applications, with implications for operational and cost efficiency.

# References


(n.d.).

Barchard, K. A. (2011). Preventing human error: The impact of data entry methods on data accuracy and statistical results. Available online 4 May 2011.
**URL:** *https://www.sciencedirect.com/science/article/abs/pii/S0747563211000707*

Cai, Z. and Vasconcelo, C. (2018). Cascade r-cnn: Delving into high quality object detection, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* .

Chen, W., Wang, L., Li, Z., Zhang, L. and Liu, Y. (2018). Efficient fruit detection using deep learning models, *IEEE Transactions on Industrial Informatics* **14**(12): 5496–5505.

Chetti, P., Balasubramanian, R. and Palaniappan, K. (2017). Image processing techniques for fruit detection and recognition: A review, *Computers and Electronics in Agriculture* **133**: 30–48.

Duan, Y., Qi, H., Wang, Y., Shi, J. and Wang, X. (2019). Centernet: Keypoint triplets for object detection, *arXiv preprint arXiv:1904.08189* .

Fang, Z., Z. X. L. Y. Y. J. and Yu, H. (2021). Deep learning for precise harvest: Identifying ripe and unripe capsicums in plantations, *Computers and Electronics in Agriculture* **178**: 105892.

Federation, N. R. (2020). The 2020 national retail security survey. Accessed on September 1, 2023.
**URL:** *https://cdn.nrf.com/sites/default/files/2020-07/RS-105905*$_2$020$_N$*ationalRetailSecuritySurvey.pdf*

Gautam, S. and Kaur, H. (2019). Machine learning and computer vision-based techniques for fruit detection and classification: A review, *Computers and Electronics in Agriculture* **155**: 228–244.

Gebremedhin, S., Kebede, A. and Gebremariam, S. (2019). An automatic vision-based detection system for fruit diseases using image processing techniques, *Sensors* **19**(14): 3117.

Ghosal, S., G. A. D. A. C. S. (n.d.). Fruit detection from images using convolutional neural networks, *arXiv preprint arXiv:* .

Girshick, R., Donahue, J., Darrell, T. and Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation.

He, K., Zhang, X., Ren, S. and Sun, J. (2015a). Deep residual learning for image recognition, *arXiv preprint arXiv:1512.03385* .

He, K., Zhang, X., Ren, S. and Sun, J. (2015b). Deep residual learning for image recognition. Accessed August 14, 2023.
**URL:** *https://encyclopedia.pub/entry/27415*

Johnson, Z., Feng, J. and Zhou, Z.-H. (2019). Deepforest: Towards an alternative to deep neural networks, *arXiv preprint arXiv:1902.08835* .

Kalkundre, R., Sahu, V., Bandi, P. Y. and Sawant, P. T. (2022). Machine learning based automated product billing and inventory.

Krizhevsky, A., Sutskever, I. and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks, *Communications of the ACM* **60**(6): 84–90.

Law, H. and Deng, J. (2018). Cornernet: Detecting objects as paired keypoints, *arXiv preprint arXiv:1808.01244* .

LeCun (2015). Deep learning: A comprehensive overview on techniques, taxonomy, applications and research directions.
**URL:** *https://link.springer.com/article/10.1007/s42979-021-00815-1*

Lee, J., Park, E., Kim, B., Kim, H. and Lee, S. (2019). A comparative study of deep learning models for fruit detection, *Sensors* **19**(10): 2427.

Lottes, J., van Henten, J. and Oerlemans, B. (2018). Low-cost smart camera for real-time fruit detection and tracking in precision agriculture, *Computers and Electronics in Agriculture* **143**: 305–315.

Mishra, A., S. A. K. A. K. V. (2017). A machine learning approach for fruit detection and localization in unconstrained environment, *arXiv preprint arXiv:1409.1556* .

Nascimento, R. A., Silva, B. B., de Almeida, R. S. and dos Santos, M. S. (2020). Fruit detection based on deep learning techniques for agroecological analysis, *Sensors* **20**(12): 3461.

Rahman, M. Z., Islam, M. A., Uddin, M. J. and Rashid, M. M. (2017). A novel approach for fruit detection and localization using alexnet, *Sensors* **17**(12): 2920.

Redmon, J., Divvala, S., Girshick, R. and Farhadi, A. (2016). You only look once: Unified, real-time object detection, *arXiv preprint arXiv:1506.02640* .
**URL:** *https://arxiv.org/abs/1506.02640*

Ren, S., He, K., Girshick, R. B. and Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks, *arXiv preprint arXiv:1506.01497* .

Safuan, Amir, A. and Aziz (2019). Fruit detection and counting from multiple images using image processing, *Sensors* **19**(15): 3122.

Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition, *arXiv preprint arXiv:1409.1556* .

Sohail, M. T., Islam, M. N. U., Rahman, M. M. and Uddin, M. A. (2021). Fruit detection and recognition from images using convolutional neural network, *Sensors* **21**(22): 7598.

Tan, M., Pang, R. and Le, Q. V. (2020). Efficientdet: Scalable and efficient object detection, *arXiv preprint arXiv:1911.09070* .
**URL:** *https://arxiv.org/abs/1911.09070*

Wang, J., Liu, Q. and Song, X. (2016). A deep convolutional neural network for fruit recognition, *IEEE Transactions on Image Processing* **25**(5): 2320–2330.

Wu, W.-H., Yang, C.-P., Chang, W.-C. and Wen, H.-K. (2019). A comparative study of machine learning techniques for fruit detection, *IEEE Transactions on Image Processing* **28**(4): 2015–2028.

Zhan, H., Wang, J., Wang, Y. and Cao, X. (2018). Computer vision techniques for ripeness classification and size estimation of fruits: A review, *Computers and Electronics in Agriculture* **142**: 143–157.

Zhang, Z., Yang, Y. and Zhang, L. (2019). An efficient vision-based fruit detection and localization system for robotic harvesting, *IEEE Transactions on Automation Science and Engineering* .

Zhu, e. a. (n.d.). A lightweight convolutional neural network for fruit detection using low-cost cameras, *arXiv preprint arXiv:* .