Check for
updates

Determining the freshness of fruits in the food industry by image classification using transfer learning

Aafreen Kazi[1] · Siba Prasada Panda[2]

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract
Convolutional neural networks (CNNs) play a significant role in a number of computer vision applications like image classification and object detection. Past studies have focused on the application of deep neural networks and convolutional neural networks for detecting the freshness of perishable goods like fruits and vegetables. However, this study tries to explore the potential of transfer learning with respect to the CNN models in image classification of fruits rather than implementing the traditional CNN architectures. In this study, the determination of three different types of fruits and its relative freshness was classified using various architectures of classical convolutional neural networks and a residual convolutional neural network. The performance of each model of the convolutional neural network was evaluated on the given set of data. The study tried to evaluate the best performing model on image data that consisted of six different types of fruits based on their test scores. The results suggest that the freshness of fruits could be determined with high accuracy by using traditional and residual convolutional neural networks. However, the residual networks perform extremely well on the fruits dataset which was considered with a test accuracy score greater than 99%. A glimpse on the implementation of the given algorithm in the food industry was also explored. Thus, this study attempts to select the CNN model that could be applied in the food industry in complementation with other computer vision techniques like background reduction, image augmentation methods.

✉ Aafreen Kazi
  kaziaafreen.ak@gmail.com

[1] MPSTME, NMIMS University, Mumbai, India

[2] Department of data Science, MPSTME, NMIMS University, Mumbai, India

🖄 Springer

1 Introduction

Convolutional Neural Networks (CNNs) are a type of deep learning networks that have the potential to extract important features from a given set of raw inputs that could be used for classification problems. They work on the basis of two important layers that include a convolutional filter layer and a pooling layer. A convolution layer that simply applies a filter operation on the given set of inputs to extract only relevant pixels. It should be noted that a feature map is developed after a repetitive application of the convolution filter on an input. A pooling layer is responsible for the dimensionality reduction of the feature map by considering only the relevant pixels [15]. Nowadays, the CNN models are being implemented in a number of computer vision tasks in a variety of fields that include medical, automobile, marketing, finance and education sectors due to the potential of the CNN architectures to provide exceptional performance in predicting image classes. They constitute of techniques that could be used to classify images by training a lesser number of parameters as compared to the traditional deep learning models. Different architectures of the CNNs that have been developed over the past few years have shown great performance for image classification. Transfer learning is the method of applying previously built architecture on different set of problems in order to get highly accurate results. Perishable products form a major component in the process of manufacturing in the food industry. However, the most common problem dealt with manufacturing the products that include perishable goods like fruits and vegetables is determining and handling their freshness. In this study, a transfer learning approach that deals with the application of a residual network like ResNet50 and different classical convolutional architectures like AlexNet and VGG-16 to an image dataset was used to predict the class and freshness for a given image of fruit.

Image Classification could be performed using techniques of machine learning as well as deep learning. However, for RGB images, the number of parameters trained in traditional deep learning and machine learning model result in inefficiency. Apart from that, machine learning models have poor transfer learning ability which avoids the re-usability of models.

Automated grading systems using the techniques of deep learning could play a significant role in evaluating the quality of perishable raw materials without a lot of human intervention [2, 11]. The roles of automated grading systems include efficient sorting and labor saving, uniformization of fruit quality, enhancing market value of products, fair payment to producers based not only on quantity but on quality of each product, farming guidance from grading results and GIS (Geographical Information System) and contribution to the traceability system for food safety and security [23, 27]. Associated benefits also include speed operation, and greater product stability [18]. Apart from that, the process of automation could increase the efficiency of the overall process of sorting and grading of the perishable goods like fruits and vegetables [20, 25].

The rest of the paper is structured as follows section 2 deals with past work. Section 3 explains the Materials, Methods, and Evaluation. Experimental results along with the comparison of different methods for detecting the freshness of perishable goods in Section. 4. Section 5 examines Industrial Implementation of the study. Lastly, the conclusion has been drawn in Section 6.

2 Past studies

Quality inspection is an important part of the food industry. It is responsible for an efficient sorting and grading of fruits and vegetables that are required in the process of manufacturing [5,

6]. Automating the process of quality inspection through different techniques of computer vision could improve the overall manufacturing process by not just improving its efficiency but also assigning more important tasks to the human laborers [4, 13].

A wide variety of machine learning methods like k-nearest neighbors (KNN), Support Vector Machine (SVM), Artificial Neural Network (ANN), Deep Learning/Convolutional Neural Network (CNN) could be used to build classification models that could help to evaluate the freshness of perishable products. Computer vision techniques have been used to recognize mould colonies on un-hulled paddy. The accuracy rates were found to be higher in CNN-based models [17]. An experimental machine vision system was used to identify surface defects on apples including bruises which suggests that out of the classification of apple disease done through texture, color and shape feature could help in determining the surface defects on apples and then the rotten or infected part of the apples were detected using K-means clustering whereas multi-class SVM was used to classify apples into fresh and rotten types [26]. Even though the study suggests that a technique which utilizes feature extraction methodology could prove to be vital for determining the surface defects of an apple, it fails to use the cutting-technologies for feature extraction techniques that would have better performance. The techniques of deep learning could be applied for the detection of fungus in mangoes and it has been concluded that the deep learning classification provides a key step to produce reliable outcomes in detecting the presence of fungus in mangoes for grayscale image [1].

The integration of the machine learning algorithms could also be used to classify the perishable products based on their relative freshness. KNN and SVM could be used to classify the type of a perishable product along with detecting the freshness of that particular product. A study generated an algorithm that classified four types of fruits and then detected the freshness of each one. This study utilized the KNN algorithm for classification of the type of fruit whereas the SVM algorithm was used to identify the freshness of each fruit classified. Linear Fischer Linear Discriminant Analysis (LDA) followed by linear SVM could be used to detect external defects of fruits [3]. However, since the algorithm was tested on just 46 images, its results could not be considered at an industrial level. Even though it has been suggested that the RBF-SVM tends to outperform for grading and discrimination of tomatoes, its performance decrease significantly as the number of classes increase and therefore SVM model would not work well for multi-class classification [10].

The Back Propagation Neural Network (BPNN) could be used to classify and recognize the fruit image samples, using three different types of feature sets, viz., colour, texture, combination of both colour and texture features [24]. However, the error propagation technique in the BPNN model which results in a reduced functional approximation suggests that the model cannot function properly in the presence of more than three layers which would decrease the overall performance of the model in detecting different varieties of perishable goods [17]. A simple architecture of four convolutional layers along with two fully connected layers have achieved an accuracy of around 98% on the test set for the detection of fruits which shows the potential of CNN models for detecting the type of fruit and its future potential to be used as a discriminator for detecting the freshness [21].

Past studies have suggested that the application of a number of machine learning algorithms prove to be effective in determining the freshness of various perishable products like fruits and vegetables, the application of two or more different algorithms might decrease the overall efficiency of the process. Apart from that, deep learning models focus more on the extraction of certain features rather than trying to fit data based on certain features that increase the risks of overfitting. However, testing the data on an unseen data (test set) could provide a better insight about the actual performance of the model.

The techniques of background reduction could be used to improve the accuracy of the deep learning models. Image classification and grading of mangoes could be performed through considering background reduction and feature selection with high accuracy [7, 12]. Even though the study does not focus on computer vision to improve the performance of the deep learning models, it could be applied in cases of industrial applications where the distribution of test data would be different from the distribution of the train data used in this study.

3 Materials, methods, and evaluation

This study focuses on determining the freshness of three types of fruits using the classical and residual architectures of the convolutional neural networks. A workflow was created in order to assess the CNN models (Fig. 1). After the assessment of different convolutional neural network models, the industrial implementation in the food industry was discussed. The following flow was followed in order to assess the best convolutional model on the given dataset.

3.1 Database description

The image data is generally extracted in an unstructured format from its source. The database consisted of different images in unstructured format which was converted into a structured format in order to process the data for feeding it into the deep learning models. Various techniques of data cleaning and manipulation was applied to develop a relational database that consisted of the classes and pixel values of each image so as to create a data that would be run on the deep learning models efficiently as well as effectively.

3.1.1 Dataset information

An image dataset consisting of fresh and rotten images of three kinds of fruits was considered in order to build classification models. It consisted of images in separate folders: fresh apples, fresh bananas, fresh oranges, rotten apples, rotten bananas and rotten oranges. Therefore, the total number of classes in the dataset was six, two classes (fresh and rotten) for each type of fruit. These folders were common subfolders in train and test folders [16]. The OpenCV library was used to handle the image pre-processing.

3.1.2 Pre-processing

The images were extracted from different folders of each class. It should be noted that the study does not involve the process of image segmentation and background reduction. It
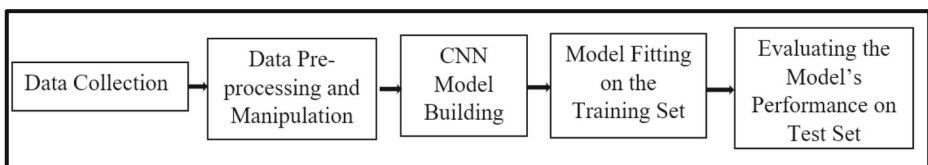

Fig. 1 Flow of the study methodology

focuses on building the model from the high-resolution images that consists of a single fruit. The following operations were performed in the image pre-processing step of the study:

- Colour Conversion: The images were converted from BGR to RGB format in order to develop uniformity in the channels of every image.
- Image Resizing: Since the images in the dataset possessed a great level of variation in their respective sizes, all of them were converted into the dimensions of 150 X 150 X 3.
- Re-scaling of Images: In order to have all the pixels in an image in the range of 0 to 255, every pixel of an image was divided by 255 and converted to float format.
- Labelling: The images in the train set and test set from each folder were labelled. The string labels were converted into numeric format so as to help the CNN models to predict the correct labels.

3.1.3 Data manipulation

Each image of the dataset was reshaped into a single size and scaled in order to maintain uniformity. The dataset was divided into train, validation and test sets. The training data was used to fit the classification models whereas the validation set was used for fitting the parameters of the models. The test set was used to evaluate the performance of the model.

The image datasets were shuffled and then split into test and train sets such that 99% of the data lies in the training set, 1% of the data lies in the validation set and the remaining 1% of The given data contributes to the test set. The dimensions of the train set were (12,239, 150, 150, 3) and the dimensions of the test data were (1360, 150, 150, 3). The dimensions of the labels present in the train set was (12,239, 6) whereas the test set was (1360, 6).

3.2 Model building

Various architectures of convolutional neural networks were used on the dataset in order to achieve high accuracy. The architecture like ResNet50, AlexNet and VGG-16 were created and implemented on the train set of the data. The CNN majorly consists of different combinations of convolutional layer and pooling layer. Along with the two layers, the CNN also consists of layers such as dropout which is used to prevent overfitting and batch normalization that allows the network to train faster. Transfer learning is a method that focuses on the application of different models built in previous research studies. It should be noted that this study trains the parameters on the image data and does not apply pre-trained parameters used in the CNN. The three networks used in this study constitutes of all the different layers mentioned above are as follows:

AlexNet model It contains eight layers with weights. The first five layers were convolutional whereas the remaining three were fully-connected layers [19]. The first layer of this architecture, takes the input image of $150 \times 150 \times 3$ dimension and applies 96 convolutional filter of size 11×11. The output of the first layer was considered as the input of the second layer and 256 convolutional filters of size 5×5. Similarly, the third and fourth layers were responsible for applying 384 convolutional filters of size 3×3. Fifth convolutional layer applies 256 kernels of size 3×3. It should be noted that all the five layers applies maximum pooling of

size 2 × 2 with a stride of four pixels along with batch normalization technique. The activation function used in the first five convolutional was a rectified linear unit (ReLU). The output from the first five layers were passed to three fully connected (FC) layers. The first two FC layers consists of 4096 units and the last FC layer consists of 1000 units. The final output layer applies the softmax activation function and consists of 6 units which is equivalent to the number of classes that were supposed to be predicted by the study. The optimizer used for training the parameters of the AlexNet model was Adam and the metric used for the evaluation was Accuracy.

VGG-16 Its architecture contains sixteen layers. The input image of 150 × 150 × 3 was subjected to a stack of convolutional filters with the smallest possible size i.e., 3 × 3. It should be noted that pooling layers are present after every two (first two blocks) or three (last three blocks) consecutive convolutional layers in the architecture. The pooling layer consisted of 2 × 2 size with a stride of 2. Some variants of this architecture also utilize a 1 × 1 convolution filter that acts as a linear transformation of the input channels which were not considered for this study. This architecture is followed by three FC layers similar to the AlexNet model. The first two FC layers consist of 4096 units whereas the last fully connected layer consists of 1000 units. The output layer consisted of the number of classes to be predicted by applying softmax activation function. The optimizer that was used for the compilation of VGG-16 model was stochastic gradient descent and the metric used for performance evaluation was accuracy.

ResNet50 A neural network with large number of layers i.e., extremely deep neural networks can decrease the overall performance of the algorithm due to the problem of vanishing and exploding gradients. A problem of gradients arises when a deep learning architecture is so huge that backpropagation becomes extremely difficult since the gradients become extremely small (vanish) or large (explode). The skip connections present in the residual networks have the potential to solve this problem. A skip connection basically feeds the activation values from a previous layer in the network to the deeper layers [14]. The ResNet50 model passes the image through 48 convolutional layers, one maximum pooling layer and one average pooling layer. The output layer consisted of a softmax activation function. The optimizer that was used for the compilation of the ResNet50 model was Adam and the metric used for performance evaluation was accuracy.

Few changes were made to the traditional architectures of AlexNet, ResNet50 and VGG-16 with respect to the dimensions to its input and output. A dropout layer is only used in the fully connected layers in order to avoid overfitting of the model. In a traditional AlexNet network, the dropout ratio was found to be 0.5 whereas in this study, the dropout used was set to 0.4 in in order to avoid reducing the capacity of the network. There were no changes made in the architecture of ResNet50. The dropout ratio used in traditional VGG-16 architectures was 0.5 whereas in this study the dropout ratio applied to the fully connected layers was 0.25 to avoid thinning of the dense layer of the network architecture. The dimensions of the output layer for all the architectures were adjusted to six since the final predictions consist of total six classes of images. Stochastic gradient descent (SGD) was used instead of batch gradient descent applied in traditional architectures to reduce time complexity for the VGG-16 architecture. The Adam optimizer was applied for optimizing ResNet50 architecture instead of the SGD optimizer with a mini-batch size of 256 that was used previously. Also, the Adam optimizer was used instead of applying SGD with momentum for AlexNet architecture. The Adam optimizer was selected

for AlexNet and ResNet50 architectures to handle gradients of the sparse and noisy data and ease of parameter configuration.

3.3 Model training and evaluation approach

All the three models were trained using the training set of the data and the parameter fit was established using the validation dataset. Early stopping with a patience level of 50 epochs was considered for improving the training efficiency of the model. The test data was predicted using the model architectures used in the corresponding train dataset in order to get the test accuracy for each model. The CNN model architecture that gave the highest test accuracy was considered as the best-performing algorithm in the study. In order to visualize the results, the predictions from best-performing model i.e. ResNet50 on the test set was captured and explained. The techniques to implement the best-performing model at an industrial level along with its benefits and limitations were discussed in the industrial implementation section. Computational outcomes for the same are discussed in the results section along with reasoning the factors responsible for the performance outcome of each model observed in the results section. After completing the entire process of classification, the test and validation images from other sources that have a different distribution than the train set could be used to further evaluate the algorithm. It should be noted that the study has considered the test and validation image datasets that have the same distribution as the training set in order to avoid discrepancies in the output of the trained model caused due to data distribution mismatch.

4 Results

Each model was trained using the training and validation set of data. After training the model parameters on the given set of data, the model predictions were considered on unseen data i.e. the test dataset which was partitioned before the training of the data train and validation set. The accuracy in predicting the freshness of the given set of images was tested on two classical CNNs and one residual network of CNN. All the three CNN architectures prove to be effective in determining the freshness of a fruit and therefore showed a test accuracy greater than 97%. During the process of model training, the loss function for the train set and the validation set were found to be exponentially decreasing for initial epochs of all the three algorithms. As the number of epochs used in model training increases, the value of the loss function shows little decrease and it becomes stable at high epoch values. The accuracy of all the three models increased exponentially for the initial epochs. However, as the number of epochs increases, the increase in accuracy decreases and it becomes more stable. It is quite important to train the algorithm on sufficient epochs in order to get a stable accuracy. For all the three models AlexNet and the VGG16, the loss function values and the accuracy values for each epoch while model training was plotted and visualized (Fig. 2).

   The two classical CNNs like VGG-16 and AlexNet showed test accuracy of around 97.74% and 99.3% respectively (Table 1). It has been observed that the residual network (ResNet50) proves to be more effective in predicting the freshness of fruits accurately. The accuracy of the ResNet50 model was found to be 99.7% on the test set (Table 1). The residual network (ResNet50) tends to work better than the classical CNN model architectures like VGG-16 and
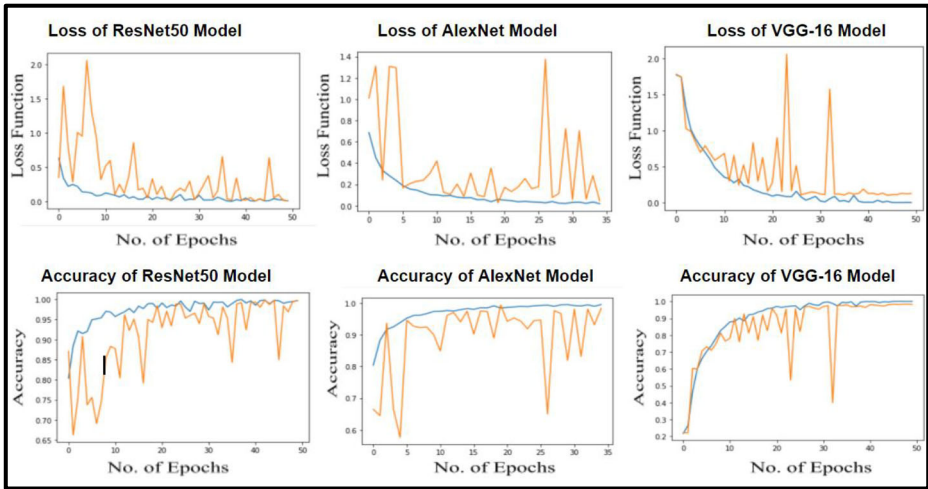
Fig. 2 Accuracy and Loss Function of the ResNet50, AlexNet and VGG-16 Model for different Epochs

AlexNet since the residual networks allow one to train on deeper layers of neural networks accurately. The performance of the ResNet50 model is the best since the residual network allows the extra or deeper layers of the network to learn an identity function that does not hurt the overall performance of deep learning architectures with a large number of layers.

Since the accuracy and loss function values outcome suggests that the best-performing model architecture is ResNet-50. The predictions from the test set were visualized (Fig. 3). The outcome results which were visualized suggests that the ResNet50 architecture successfully determines the rotten nature of the three considered fruits which included apples, oranges and bananas. The model architecture successfully determines various types of rotten spots and areas in order to classify a fruit as fresh or rotten. Also, since the training and validation sets consisted of various augmented data that captured fruits from various angles and brightness, the algorithm successfully identifies fruits and classifies them as fresh or rotten based on a given image.

5 Industrial implementation

Industrial automation is the use of control systems integrated with information technologies for handling different processes and machineries in an industry to replace human labour. It is the second step beyond mechanization in the scope of industrialization. The study could have a

Table 1 Accuracy, Loss Function Value of the Three Model Architectures

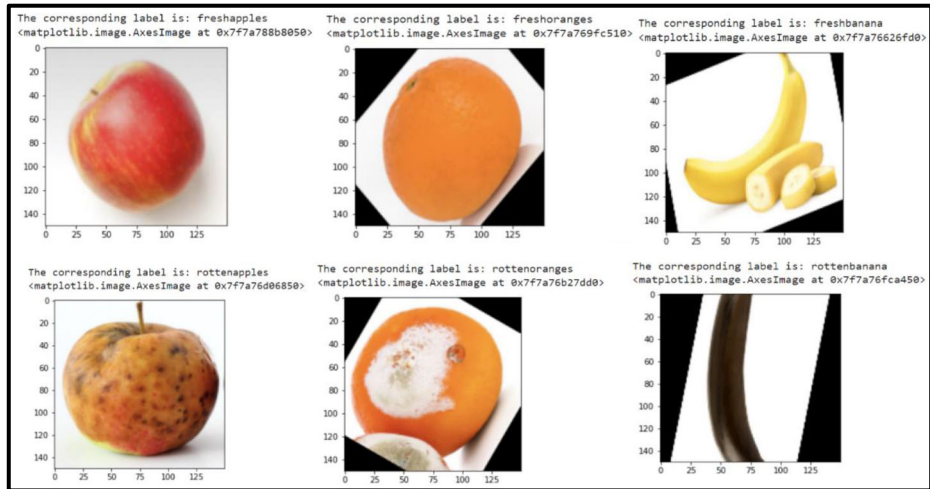| Model Architecture	No. of Layers	Test Accuracy Value	Loss Function Value
ResNet50	50	99.7%	0.0099
AlexNet	8	99.3%	0.0797
VGG-16	16	97.74%	0.3001

Fig. 3 Predictions of the Six Classes (Two Classes for Each Fruit) from the best-performing model-ResNet50

great impact for the industrial applications in the food processing industries which focuses on preparing syrups, beverages, marmalade, jams and other food products from perishable raw materials like fruits. In a pre-manufacturing setup, the cameras installed near the conveyor belt could be used to determine the freshness of a raw material before feeding them into the actual manufacturing process. A camera system that takes in image input every time a batch of raw materials or a single piece of raw material passes through the conveyor belt. It should be noted that the set-up and installation would mainly be dependent on the type of conveyor belt and the design of the manufacturing process system. Optionally, background reduction techniques could be applied to the images captured by the installed cameras and they could be split in multiple parts. A ResNet50 model could be used to determine the freshness of the raw materials through the input image data that would be coming from the cameras installed. Further, the decision of discarding the raw material or keeping them in the manufacturing process could also be performed through the algorithm (Fig. 4).

In order to further analyze the impact of the algorithm, a use-case of a food industry has been discussed. Grading and sorting the fruits based on their freshness is one of the most important and time-consuming steps in the manufacturing of jams. Basically, this step would be responsible for majorly deciding the overall quality of the final product for a jam manufacturing company. The above algorithm could help in the process of automating the process of separating fruits based on its freshness in order to decrease the overall time required for sorting and grading of fruits which is a mandatory step in not just jam industries but also other food industries that includes fruits as their main component product or side-component product. However, a major hurdle in the application of the selected algorithm i.e. ResNet50 would be that the current network has been trained and tested on images that were easier to classify. However, at industrial level, the images captured by the camera would have a different distribution than the images used for training the model. This problem could be approached by adding some industrial images into the training set and then randomly shuffling them so that the algorithm learns to classify the testing images of different distributions. Thus, as the number of images increases with time period, the performance of the algorithm should also improve.
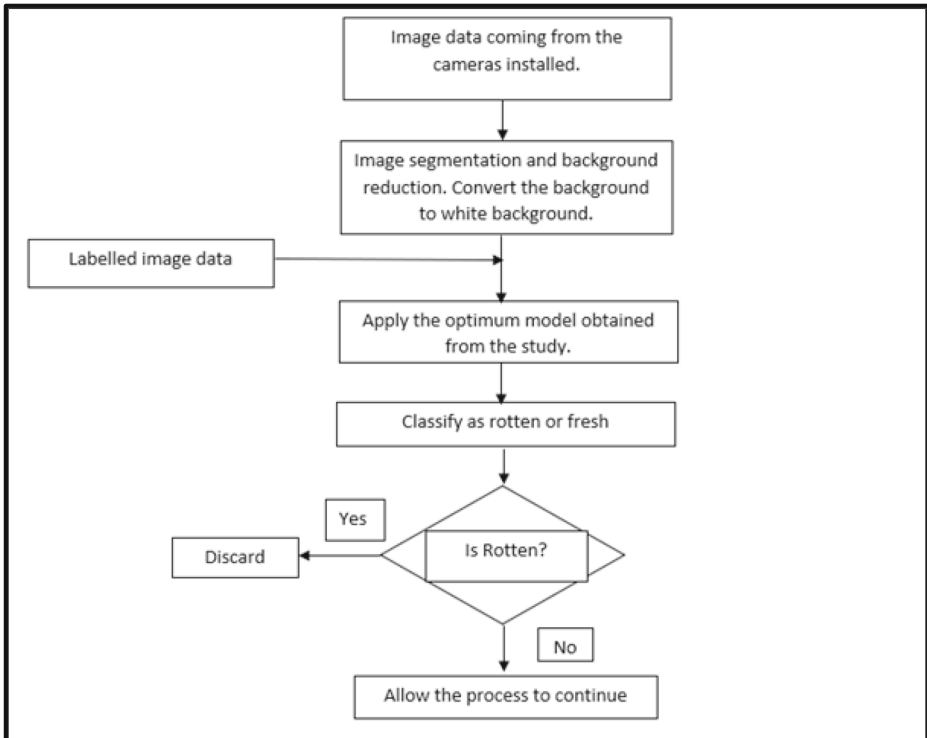
Fig. 4 Flow of the algorithm that could be used to implement the model in food industries

6 Conclusion

As the number of layers increases, the accuracy of a model increases. However, training a model with extremely deep layers of neural networks suffers from the effect of vanishing and exploding gradients. The skip connections in the residual network (ResNet50) allows the activation of a layer to be fed into another layer that lies deeper in the network that diminishes the problem of vanishing and exploding gradients and results in better performance than the VGG-16 model which is only 16 layers deep and AlexNet which is only 8 layers deep. It should be noted that the ResNet50 model provides the best accuracy and also proves to be better than the other classical convolutional networks in the overall performance. This is due to the fact that the ResNet50 has less number trainable parameters since its last layer consists of global average pooling instead of the fully-connected layer in the AlexNet and VGG-16 model. The VGG-16 has simple network architecture but has more trainable parameters as compared to the AlexNet model that increases the time complexity of the algorithm.

The study concludes that the ResNet50 model outperforms AlexNet and VGG-16 model architectures in terms of the test accuracy. However, the difference between the accuracy achieved from the implementation of the ResNet50 and AlexNet model is quite low which suggests that for the deployment of model, AlexNet could be a better choice since it has a less complex architecture and low number of trainable parameters which would prove to be computationally efficient in the industrial application. It should be noted that the different types of rotten fruits could be determined by the model if it is present in the training set

images. This study has not determined the level of rottenness in a fruit and therefore it cannot predict whether a fruit has been completely rotten or partially rotten and therefore does not consider the subjective evaluation of the fruit. However, since the study has various types of rotten fruits in the training set, the best-performing algorithm could successfully determine various types of rottenness in the given set of three different fruits at various angles and brightness levels (Fig. 3). Thus, it could be concluded that if the training set consists of fruit images with different types of rottenness then the algorithm would successfully identify them.

The test data in the model deployment process would vary than the data that was used in the study. This could be fixed by timely addition of part of the test data into the train data. Further studies could be performed on improving the performance of the models on actual industrial data by exploring different ways of computer vision. The application of the study for detecting multiple objects is beyond the scope of the study. The algorithm used in this study could be applied in complement with background reduction and splitting techniques to yield better results. Split and merge segmentation is an image processing technique used to segment an image. The image is successively split into quadrants based on a homogeneity criterion and similar regions are merged to create the segmented result. It should be noted that the splitting technique that gives pieces of images with low-resolution would not work well with these algorithms.

7 Discussion

The two major types of evaluations that play a significant role in determining and continuously monitoring the freshness of perishable goods like fruits include subjective and objective evaluations. The method of objective evaluation has proven to be inefficient and would prove to be beneficial only in certain cases. However, an objective methodology based on the saliency map features which correlates with a qualitative opinion from a human observer proves to perform better than other objective evaluation techniques [8]. The saliency map in the current study would include the various types of fruits differentiated from their corresponding backgrounds. The subjective evaluation of food includes overall assessment based on sensory response and personal opinions whereas the objective evaluation is based on measurement of a single or few attributes of a food material rather than the overall quality of the food product. In case of subjective evaluation, a metric focused on human visual characteristics such as the classes of fruits being classified as fresh, spotted but fresh, slightly rotten, completely rotten could be developed to assess the quality of fruits in a given set of images. This assessment could be done in conjunction with human observers in order to integrate a penalty and compensation function in the overall process of subjective evaluation [9]. The requirement of high costs, time and human resources for subjective evaluation have created a high demand in food industries whereas the quick response generated through objective evaluation is beneficial only for certain cases. Subjective as well as objective assessments are interrelated and are vital to determine the acceptability based on biological as well as mechanical aspects [28]. This study tries to integrate the subjective evaluation (a dataset labelled through human perception) and objective testing (statistical approach of CNN models) to produce results similar to that of humans. Basically, it has focused on developing a system that could produce results similar to that of human-like evaluation for determining the quality of three different types of fruits. In order to evaluate a conjunction of subjective and

objective evaluation, a group of human observers could be used to develop and label the data based on their perception to train the given deep learning models.

In conjunction with the subjective evaluation of the fruit image, the occurrence of fruit skin spots could make it difficult for an algorithm to predict rottenness through a given image of a fruit if spotting patterns are not present in the training set. Separation of pattern that exists in the fruit skin spots and the rotten area of a fruit would become significant in further analysis for determining the freshness of a given set of fruits. The spot patterns which exist on the fruit skin might often overlap with the rotten part of the fruit and would therefore require a large amount of additional data for the deep learning models to function thereby increasing the overall training time, costing and efficiency of the models discussed in this study at a deployment level for industries. A network that focuses on differentiating the patterns with high levels of overlapping in an image would play a significant role in such complex cases in order to achieve human-like performance in determining the freshness of fruits that consists of a spot as fresh whereas a rotten part overlapping the spot as rotten. A computational network which automates the process of pattern separation with the functionality similar to that of a section of the human brain known as the dentate gyrus of hippocampus which is responsible for the process of pattern separation in humans have shown great results for performing pattern separation of over a high-level of overhauling patterns in images. This network consists of a granule cell which is responsible for determining the output of the model. The prominent feature of the model includes two excitation steps responsible for the generation of excitation of the granule cells when an input is fed and two inhibition steps which is responsible for transferring only a few of the granule cells that helps in the determination and differentiation of overlapping patterns [22].

A number of researches done in the past were focused on rules based approaches of computer vision to identify the freshness of perishable goods like fruits and vegetables. Even though there were some studies that were focused on traditional machine learning or deep learning models that included feature extraction in order to classify images, the performance of the models were focused at research and development since their accuracy was approximately 90–95% on the test set. However, since this study uses more data for training, it has shown a higher level of accuracy as compared to the existing models. Various researches have focused on basic architectures of CNNs but there was very little focus on improving the accuracy of the model through tuning the overall parameters and hyperparameters. This study not just tries to develop and evaluate the best-performing architectures out of the state-of-art classical CNN algorithms like AlexNet, VGG-16 and residual CNNs like ResNet50 architecture but has also focuses on a roadmap through which these type of deep learning automations could be implemented at an industrial level. The research provides insights about the factors that could improve the performance of the three CNN architectures that were applied.

The following were the novel contributions of the research:

- This study provides an insight about the methodologies that could develop the best-performing deep learning architecture with respect to accuracy on fruits dataset.
- Since the ResNet50 and Alexnet model could be used to identify the rottenness of other perishable goods with higher than 99% accuracy on the given dataset, they have the potential to be used for determining the freshness of other fruits and vegetables at an industrial level.
- Even though past research has shown the potential of deep learning models in detecting the freshness of fruits, this study tries to implement a more fine tuned and optimized version of

the currently applied models to get an accuracy of 99.7% on the test set which none of the studies have achieved.

- This research provides a roadmap for food industries to implement deep learning architectures for automating the process of grading fruits since techniques that could be applied in upscaling of this research at an industrial level was discussed.

References

1. Ananthi N et al (2019) An image processing based fungus detection system for mangoes. Int J Innov Technol Explor Eng (IJITEE)
2. Aniket H et al (2020) Fruit freshness detection using CNN approach. Int Res J Modern Eng Technol Sci
3. Ann N (2019) Automatic classification for fruits' types and identification of rotten ones using k-NN and SVM. Int J Online Biomed Eng (iJOE)
4. Anuja B, Bhansal A (2018) Fruits and vegetables quality evaluation using computer vision: a review. J King Saud Univ - Comp Inform Sci
5. Ayman E (2012) Understanding color image processing by machine vision for biological materials. IntechOpen
6. Bennedsen BS, Peterson DL (2005) Identifying defects in images of rotating apples. Comput Electron Agric
7. Dameshwari S, Ravindra P (2017) Defect identification and maturity detection of mango fruits using image analysis. Am J Artif Intell
8. Dariush A, Azam B (2021) An objective method to evaluate exemplar-based inpainted images quality using Jaccard index. Multimedia Tools Appl
9. Dariush A, Azam B, "Inpainted image quality evaluation based on saliency map features", 5th Iranian conference on signal processing and intelligent systems (ICSPIS), 2019.
10. David I et al (2019) A computer vision system for defect discrimination and grading in tomatoes using machine learning and image processing. Artif Intell Agric
11. Dayanand S (2012) Identification and classification of bulk fruits images using artificial neural networks. Int J Eng Innov Technol (IJEIT)
12. Guoxiang Z (2017) Fruit and vegetables classification system using image saliency and convolutional neural network. IEEE
13. Hameed K et al (2018) A comprehensive review of fruit and vegetable classification techniques. Image and Vision Computing
14. He K et al (2012) Deep residual learning for image recognition, arXiv
15. Horea M, Mihai O (2018) Fruit recognition from images using deep learning. Acta Universitatis Sapientiae, *Informatic*a
16. Kalluri SR (2018) Fruits fresh and rotten for classification (version 1) [data set]. Zenodo https://doi.org/10.5281/ZENODO.4788774
17. Ke S et al (2016) Recognition of Mould Colony on Unhulled Paddy based on computer vision using conventional machine-learning and deep learning techniques. Scientific Reports
18. Koyama K et al (2021) Predicting sensory evaluation of spinach freshness using machine learning model and digital images. PLoS One
19. Krizhevsky A et al (2012) ImageNet classification with deep convolutional neural networks.
20. Wang Lu, et al., "Detection of fruit skin defects using machine vision system", IEEE, 2014.
21. Mahmoud H et al (2018) A study on CNN transfer learning for image classification. Adv Comput Intell Syst
22. Modhej N et al (2020) Pattern separation network based on the hippocampus activity for handwritten recognition. IEEE
23. Naoshi K (2010) Automation on fruit and vegetable grading system and food traceability. Trends Food Sci Technol
24. Ray Z et al (2017) Food supply chain management: systems, implementations, and future research. Industrial Manag Data Syst
25. Santi B et al (2020) Identification, classification & grading of fruits using machine learning & computer intelligence: a review. J Ambient Intell Human Comput
26. Shiv D, Anand J, "Apple disease classification using color, texture and shape features from images", Signal, Image and Video Processing, 2016.

27. Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale image recognition. ICLR
28. Singham P et al (2015) Importance of objective and subjective measurement of food quality and their inter-relationship. Food Process Technol

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.