Lemon Quality Detection Using CNN

J. Lekha[1], D.Suryaprabha[2], S. Saraswathi[3], Noel Mathew Thomas[4]

{lekha.j@christuniversity.in[1], spayrus@gmail.com [2], nascsaraswathi@nehrucolleges.com[3] , noelmathew8539@gmail.com[4]}

Associate Professor, Dept. of Data Science, Christ University, India[1],Assistant Professor, Dept. of IT & DCFS,Nehru Arts And Science College India [2], Dean, Academic Affairs ,Nehru Arts And Science College India[3]

Abstract. The traditional method of visually inspecting each lemon and sorting them according to their quality has been done by human inspections. Human inspections can result in some misconceptions while grading the quality of lemons. Lemons are important crops that has many health benefits and plays an important role in the daily lives of poor to rich people. It is not only considered as an important crop within India but all over the globe. It is an essential ingredient in a variety of applications like food industries, medicines, cosmetics etc.,Hence identifying the quality of lemons plays a vital role results in ensuring that the beneficiaries are getting the good quality lemons. The role of deep learning algorithms is high in classification of the lemons based on its quality parameters. The CNN algorithm has been proven the better method when compared with the traditional machine learning algorithms like Suppeor Vector Machine(SVM) and Random Forest. The CNN model proved its accuracy to be 98.75% , which is better results when compared with the traditional methods. This research has proved that the deep learning algorithms are efficient in grading the lemon based on quality parameters.

Keywords: Lemon, CNN, SVM, RandomForest.

1 Introduction

India is a country which has rich set of land and other natural resources. Many historical crops like mango, banana, lemon, water melon are some of the important agricultural crop varieties of India. The agricultural technology is also developing considerably in recent years. Lemon is one of the important agricultural crop that plays a vital role in healthcare industries to fertilizer manufacturing company leading to economical, ecological and industrial growth of the country. Due to its utility in a wide variety of applications it is one of the valuable crops all over the world. It is one of the economical crop for the farmers as it consumes less amount to yield the crop. The quality assessment of lemon is a major concern to focus upon in the current era due to its supply in large quantity to various industries. Many researchers have made an intensive research on how to classify the lemons based on its appearance and disease type. Tominaga et al.in the year 2003 used image processing technique for classifying the lemon based on two

IACIDS 2023, November 23-25, Lavasa, India

DOI 10.4108/eai.23-11-2023.2343188

parameters colour and shape. AlZamily and Abu Naser made an extensive study and improvised by implementing deep learning algorithms for classifying lemon based on colour and shape.

A country's economic growth involves a good amount of export to other countries. Lemons are a high in demand crop which has its top rating in export goods category. Visual assessment of large quantity of lemons may lead to human errors in classification. This may result in rejection of goods by the buyers and leading to less percentage of export of Lemon to other countries. To overcome all these limitations, automatic this research proposes automatic classification of Lemon quality based on machine learning and deep learning techniques. The research design proposes a novel method that uses CNN for effective classification of lemon quality. This model is inputted with a labelled training dataset with intensive set of features such as textures, shape, colour and patterns etc., The model is trained with a good set labelled lemon images which the model utilizes to classify if the Lemon is good or bad based on the trained images. This research provides a good insight on the best choice of features that helps in accurate classification of Lemon based on quality that benefits the industries that involves mass usage of Lemons.

2 Literature Review

An algorithm was created by Sharath et al. to identify three diseases in pomegranates: cercosporin, borer, and bacterial blight. [1] Plant disease detection uses image processing software. Segmentation, feature extraction, and disease classification are performed after pre-processing mobile device-taken photos. Based on the colour characteristics and edge information, diseases are categorised. The efficiency in detecting diseases was found to be 85%An accuracy of 85% in detecting diseases was discovered.

Iqbal et al. [2] reviewed methods for recognizing and categorizing diseases in citrus plant leaves, providing a taxonomy of citrus leaf diseases, and evaluating pre-processing, segmentation, feature extraction, and classification techniques. They found that pre-processing can improve segmentation accuracy and that K-means clustering is often used for segmenting diseased plants. However, they noted the need for new tools to automate detection and classification fully. Texture characteristics were highlighted as important for representing disease in images, with SVM and NN commonly using them for disease detection and classification.

The approach put forward by Rohit Ranjan,[3] will help farmer to boost lemon fruit farming. Pictures of lemon fruit that have black spots and anthracnose on them were taken. K-mean clustering is used to segment and group test photographs. The classification process, which employs support vector machines to identify the images. They initialised cluster k, used in the clustering procedure, as 3. The name of the ailment is displayed in the proposed system's output.

Blasco et al., [4] developed a fruit-sorting algorithm that employs multispectral computer vision to classify different types of exterior faults in citrus. The algorithm integrates morphological traits, visible and non-visible information, and determines spectral and morphological parameters like colour, area, length, width, and the fast Fourier transform (FFT) of the perimeter's radius signature. The algorithm uses some important parameters to identify

Al-Hiary et al. made a study and classified plant diseases such as early scorch, cottony mold and late scorch using the colour information of the plants by applying texture statistic method on SDGM matrix [5]. The algorithm achieved an accuracy and precision rate of 83% and 94% in detecting the plant diseases using green pixel masking.

Fuentes et al. (2019),[6] used long short-term memory with CNN to classify the disease on sugarcane images and detected the type of diseases outperforming the previous methods with an accuracy of 92.7%. Singh et al. (2019) applied a deep belief network(DBN) to classify the sugarcane leaf images into diseased and non-diseased with an accuracy of 91.7%.

Kukreja and Dhiman [7] incorporated DL algorithms for detecting diseases diseased affecting the citrus fruits. The system was trained with 150 distinct images and the images were augmented to 1200 images. Each image was analyzed using nine features. The dataset was then pre-processed to improve the features of the image and attained an overall accuracy of 89.1%.

Gavhale et al. demonstrated a early disease detection method based on image processing on the leaves of citrus fruits [8]. The system was tested with 50 images belonging to each disease sample, and the system was trained using a dataset containing 200 images of two major diseases namely canker and anthracnose. The researchers used SF-CES for improving the color enhancement of the image so that the portion of the disease can be easily identified by applying K-mean clustering.

The study by Shuxiang Fan et al., [9] used deep learning approach to detect faulty apples on a 4-line fruit automation system. They combined CNN and an inexpensive computer vision module consisting of a commercial camera and self-developed LED line lights. Their CNN-based classification architecture outperformed a conventional image processing technique in terms of accuracy, recall, and specificity for the testing set, achieving results of 96.5%, 100%, and 92.9%, respectively. The study suggested that the proposed CNN-based classification model coupled with the reasonably priced computer vision system could have tremendous potential for commercial fruit packing lines.

Chongke Bi et al., [10] proposed a low-cost, reliable, and high-precision approach for identifying apple leaf diseases in their work. The approach can be easily implemented on mobile devices and does not require the aid of specialists. The researchers evaluated the effectiveness and accuracy of two well-known CNN models, ResNet152 and InceptionV3, and compared them to the MobileNet model. They found that the MobileNet model had the highest efficiency, with a faster detection time and lower cost compared to the other models.

S. Ashwinkumar [11] implemented the OMNCNN architecture which combines the layers liks pre-processing, segmentation, feature extraction, and classification to classify plant diseases. This model applied Kapur's thresholding-based segmentation to analyze the infected portion of the leaf images using bilateral filtering. The system achieved 0.985 precision and a recall of 0.9892.

3 Convolutional Neural Network

Convolutional Neural Networks (CNNs) is a deeplearning algorithm frequently applied in the field of computer vision that enables image classification with promising accuracy. This research focuses on deeper understanding of the layers undrewent by CNN algorithms. The components of CNN architecture includes input layer, convolution layers, pooling layers, fully connected layers, activation functions and the output layer and is stated in fig.1.

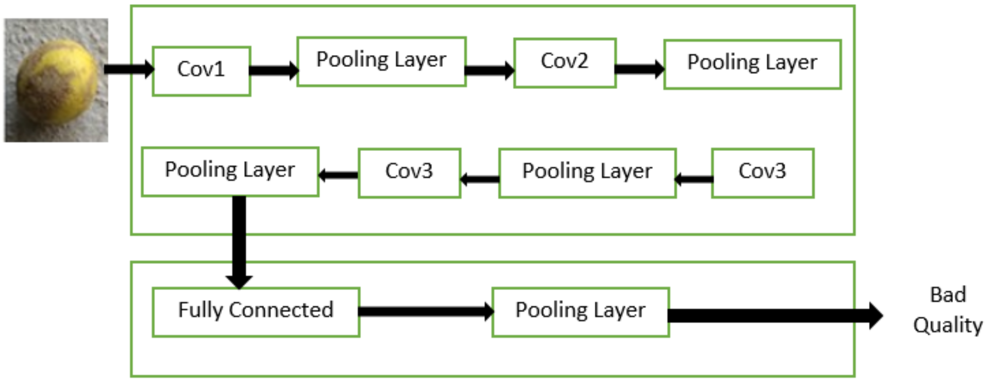

Fig. 1. Prediction using CNN architecture

3.1 Convolution Layer:

The role of this layer is to receive the input images and extract the useful features by passing the input images to a self-learnable filters named as kernels. The filter works by extracting the edges, corners and blobs of each input image. The feature mapping is done for each feature and is produced as output.

Based on the complexity of the image and image size the size and size of the filter is fixed. The size of the filter increases as the complexity of the image increases. In this research, 9 different size of filters are used producing different set of outputs.

3.2 Activation Layer::

The non-linearity of the network is applied using the activation layer .Since most of the real-time data is non linear in nature , application of activation function is an important task in CNN architecture. The activation function operates by transforming on each element from linear form to non-linear form.ReLU, tanh ReLU and sigmoid are the set of frequently used activation functions used along with CNN model.

3.3 Pooling Layer:

The spatial dimensions of the produced feature map is reduced in this layer. The inputted feature map is inputted into non-overlapping layers and two categories of pooling is recorded for each pool: maximum and average of each pool.

The pooling is done in two layers: (i) The first layer works by reducing the complexity of computation through reducing the number of parameters and (2) by translating the degree of the input image. Each layer performs max-pooling layer that performs spatial dimensionality reduction by a factor of 2.

3.4 Fully Connected Layer:

This is the third layer which connects the neurons from the pooling layer to the current layer. The output from the pooling layer is received in a one dimensional tensor and inputted into this fully connected layers. This layer is responsible in learning the non-linear combinations of most relevant features. In this research the convolution model is provided with 256 elements of tensor nodes which is then sent to two layers one with 64 neurons and one with 3 neurons respectively. The output of this layer is a probability distribution of 3 classes.

From the discussion of the above three layers, it is understood that CNN is more suitable for tasks involving computer vision and image processing tasks. It outperforms many other algorithm in tasks such as feature extraction, conversion of linearity into non-linearity, spatial dimensionality reduction of feature maps etc.

4 Methodology

4.1 Image Dataset Acquisition :

The input dataset contains 2,076 images with a proportion of 951 diseased images and 1,125 good quality lemon images. The data set was generated by capturing images using a digital camera under suitable day light conditions to maintain good detection ration of the lemon quality. The system is trained with good quality lemon images as positive class and diseased images as negaive class. The fig2 shows the sample classification of diseased and googd quality lemon images.

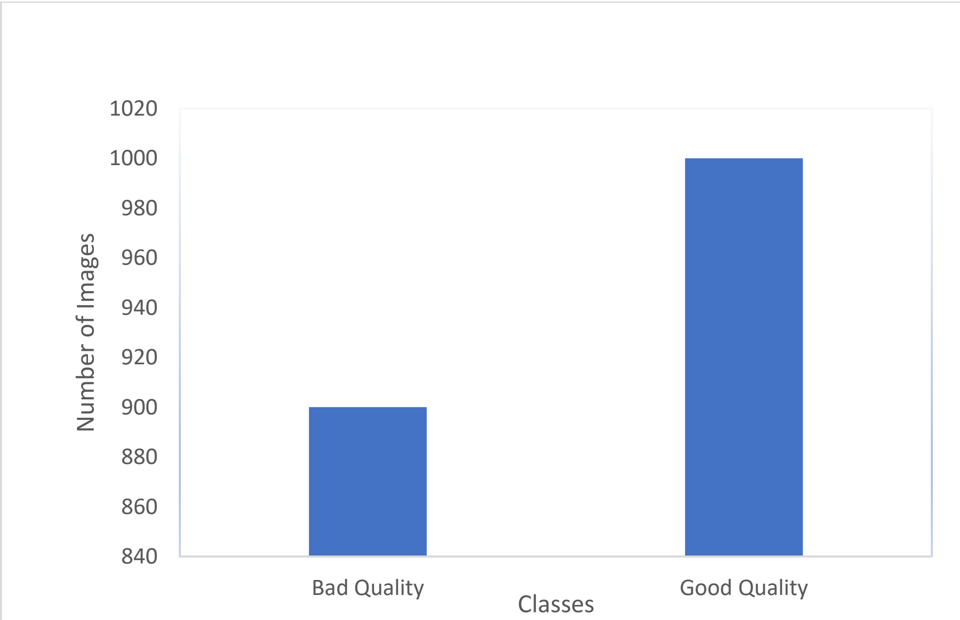

Fig. 2. Bar plot of number of images in each class

4.2 Image Pre-processing :

Preprocessing is an important step in any image classification problem. It helps the classification model to provide high accuracy in training and testing models. There are many types of image preprocessing which may chosen based on the relevance and objective of the research. In this research the image enhancement pre-processing is done to maximize the quality of the image so as the model can extract the classification features and accurately classify the images.

The pre-processing is done step by step. The initial step is to convert the image from color ranges to grayscale images as shown in fig 3. This conversion is achieved by reducing the intensity of the red, green and blue colors in the image to gray color. The reason behind this conversion is, gray scale images express high contrast level that helps the model to easily extract the important features in the image. The next step is to reduce the size of the image that helps the model to quickly process the image. This is done by reducing the resolution of the image so that the model will be able to process more images in less time. The next step of pre-processing is to do equalization by applying histogram equalization. This method helps in equalization of the image contrast level into even intensity images across all the gray scale values so that the model will be able to learn the dull portions of the image due to lack of lightings during image capture. This method also helps in increasing the accuracy of the image.

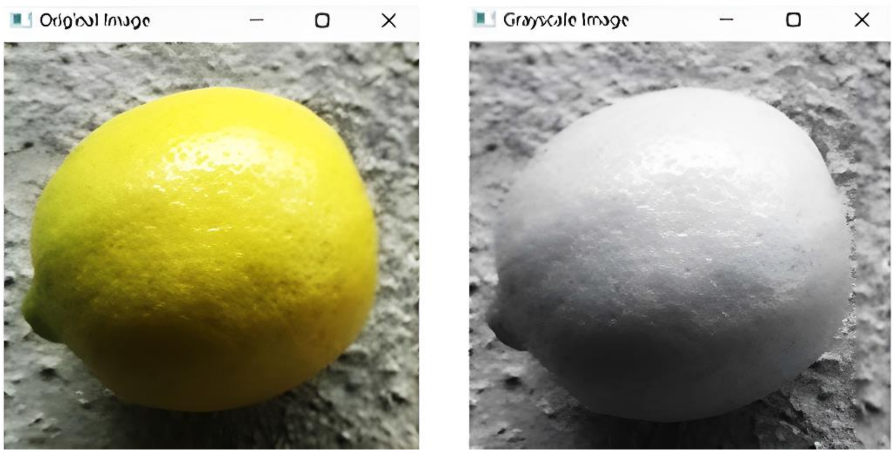

Fig. 3. Image after converting into greyscale.

The next step in pre-processing is to normalise the pixel values in the equalised gray toned image. This helps in adjusting the values of pixels in a selected range to scale and distribute the image with uniformity. This type of pre-processing can be done in 2 ways. First is to scale the pixel values to a particular range and the second way is to scale the pixel values from zero to mean of all the pixel values. This also helps to balance the dull parts of the image due to poor lighting conditions during image acquisition in order to improve the accuracy of the model.In conclusion, image pre-processing is proven to be an important step in image classification problems.

4.3 Feature Extraction

The CNN is used to extract useful features from the pre-processed images. Feature extraction is the method used to fetch[12] the usefule portions of the input image and transform them into meaningful features representing the data for classification. This helps the model to provide effecient results with less complexity in data analysis.

In this research the feature extraction is done using Convolution Neural Network. This method does the extraction by passing the input through three layers : convolution layes, pooling layer dropout layers. The first layer passes the input data through a set of learnable filters for extracting the edges, textures and shapes in the input image. The second layer then creates a feature map resulting in dimensionality reduction of feature set so that the least improtant features are eliminated and the important features are preserved. The third layer takes care of overfitting by dropping out certain least important features during training the model and the output is passed to the final convolution layer and classified as positive or negative class.

CNNs architecture is very much effective for feature extraction because of it ability to automatically learn and extract features from the input images without the need for traditional feature engineering.The total number of parameters extracted by the the model is 156,067. All these parameters are used for training and improving the model’s performance. The parameters are update to the training model through an optimisation algorithm to minimize the loss function and improve the trainig accuracy of the model.This step is also important to ensure the training time and the operational complexity of the model is less.

The model after passing all the steps of pre-processing, extraction and training have resulted in an accuracy of 95.96% on the training data and 97.40% on the testing data, with a loss of 0.1112 on the training data and 0.0848 on the testing data as shown in the given fig. 4. The model had outperformed the existing methods with a high accuracy on both training and testing dataproving that the model is not overfitting to the training data.

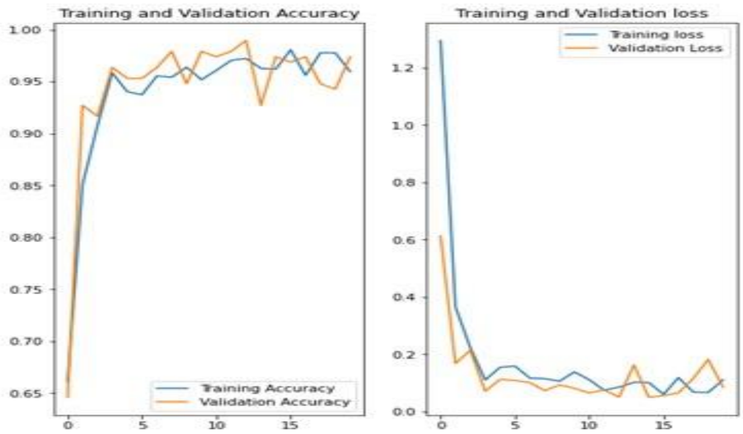

Fig. 4. Graph showing training and validation accuracy and loss.

4.4 Classification

Classification is done in lemon quality dataset to categorize lemons into different classes based on their quality. This helped in classification of lemons images of low quality and good quality for industrial and daily consumption as shown in fig 5.

Fig. 5. Categories trained by the classifier.

CNN is used to classify the images of lemons. CNN has multiple convolutional and pooling layers followed by fully connected layers. The model is trained with an epochs rate of 20 for which it achieves a high accuracy of 98.75% on the test data. The results are also compared with the traditional models like SVM and random forest. From the results it is proven that CNN had an proven an accuracy of almost 98.75% with is higher than SVM with 87% accuracy and Random Forest 71% accuracy.

5 Conclusion

This research had made an intensive study on the traditional methods used for disease prediction on important crops and its metrics used for evaluation. The literature review stated that two of the existing methods SVM and Random forest have proven a reasonable results in the disease prediction. Based on this findings the proposed work implemented a CNN based deep learning model for improving the existing results through pre-processing, feature extraction and model building for training and testing. The proposed model had outperformed the esiting SVM and Random forest with an improved accuracy and loss values. This proposed work can be further tested for disease prediction on other important crops[13][14] used in daily lives and in demand for exporting.

References

[1] Sharath, D. M., et al. "Disease Detection in Pomegranate using Image Processing." 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI) (48184). IEEE, (2020)

[2] Iqbal, Zahid, et al. "An automated detection and classification of citrus plant diseases using image processing techniques: A review." Computers and electronics in agriculture , 153 ,pp.12-32, (2018)

[3] Usha P , et al., "A Survey make out of Lemon Fruit and Leaf Disease Utilizing Ways of Image Processing Techniques", International Journal of Engineering Trends and Applications (IJETA), 8(3), pp. 41-45 (2021)
[4] Blasco, José, et al. "Recognition and classification of external skin damage in citrus fruits using multispectral data and morphological features." Biosystems engineering 103.2 , pp. 137-145, (2009)
[5] Al-Hiary, Heba, et al. "Fast and accurate detection and classification of plant diseases." International Journal of Computer Applications 17.1, pp. 31-38, (2011)
[6] Fuentes, A., Yoon, S., Kim, S., & Park, D. S. Sugarcane detection using convolutional neural network with long short-term memory. Sensors, 19(22), pp. 39-49, (2019)
[7] V. Kukreja and P. Dhiman, "A Deep Neural Network based disease detection scheme for Citrus fruits," 2020 International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, pp. 97-101, doi: 10.1109/ICOSEC49089.2020.9215359, (2020)
[8] Gavhale, Kiran R., Ujwalla Gawande, and Kamal O. Hajari. "Unhealthy region of citrus leaf detection using image processing techniques." International Conference for Convergence for Technology-2014. IEEE (2014)
[9] Fan S, Li J, Zhang Y, Tian X, Wang Q, He X, Zhang C, Huang W. "On line detection of defective apples using computer vision system combined with deep learning methods". Journal of Food Engineering. 286:110102, (2020).
[10] Bi, C., Wang, J., Duan, Y., Fu, B., Kang, J. R., & Shi, Y. MobileNet based apple leaf diseases identification. Mobile Networks and Applications, pp. 1-9, (2022)
[11] Ashwinkumar, S., Rajagopal, S., Manimaran, V., & Jegajothi, B. Automated plant leaf disease detection and classification using optimal MobileNet-based convolutional neural networks. Materials Today: Proceedings, 51, pp. 480-487,(2022)
[12] Bhagwat, R., & Dandawate, Y. Comprehensive Multilayer Convolutional Neural Network for Plant Disease Detection. International Journal of Advanced Computer Science and Applications, 12(1), (2021)
[13] S. C. K., J. C. D. and N. Patil, "Cardamom Plant Disease Detection Approach Using EfficientNetV2," in IEEE Access, vol. 10, pp. 789- 804, DOI: 10.1109/ACCESS.2021.3138920. (2022)
[14] Rath, A. K., & Meher, J. K. Disease detection in an infected plant leaf by computational method. Archives of phytopathology and plant protection, 52(19-20), pp.1348-1358. (2019)