

# Analysis of Convolutional Neural Network based Image Classification Techniques

**Milan Tripathi**

Student, Department of Computer Engineering,  
Advanced College of Engineering and Management,  
Institute of Engineering,  
Tribhuvan University,  
Nepal.  
Email: [tripathimilan28@gmail.com](mailto:tripathimilan28@gmail.com)

**Abstract:** With the rapid urbanization and people moving from rural areas to urban time has become a very huge commodity. As a result of this change in people's lifestyles, there is a growing need for speed and efficiency. In the supermarket industry, item identification and billing are generally done manually, which takes a lot of time and effort. The lack of a bar code on the fruit products slows down the processing time. Before beginning the billing process, the seller may need to weigh the items in order to update the barcode, or the biller may need to input the item's name manually. This doubles the effort and also consumes a significant amount of time. As a result, several convolutional neural network-based classifiers are proposed to identify the fruits by visualizing via the camera for establishing a quick billing procedure in order to overcome this difficulty. The best model among the suggested models is capable of classifying pictures with start-of-art accuracy, which is superior than that of previously published studies.

**Keywords:** Automatic billing system; Convolutional neural network; Computer vision; Deep Learning

## 1. INTRODUCTION

Artificial intelligence refers to programmes that enable a computer to complete a task without the assistance of a person. All essential instructions to be given to the system are programmed at the time of installation. This saves a significant amount of time and energy in the

workplace. Among the various techniques used in Artificial Intelligence, deep learning-based processes have become very popular. It is used to solve various problems in computer vision and natural language processing domains. With the advancement in technology and the availability of huge datasets, deep learning has become a very promising technology. Various models based on convolutional neural networks are employed to identify fruit pictures in this research.. This system can be used in industries and factories to detect fruits without any human intervention. This will greatly reduce the time by automating the identification, billing, and other processes. Figure 1 provides the basic architecture of deep learning-based computer vision model.

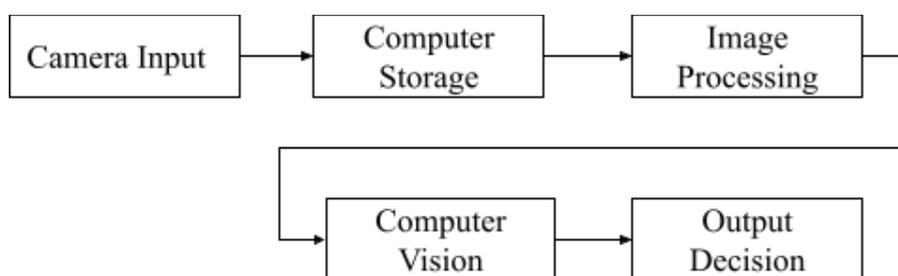


Figure 1. Basic Architecture of Proposed Computer Vision System

The system is attached to a camera to capture an image, which can be read by a system-programmed algorithm. Aside from smartphone cameras, computer system screen images are used as visual scenes for computer vision based systems [1].

Computer Vision has become more prominent among medical researchers. A study [2] uses thermal pictures to propose a computer-aided diagnosis method based on convolutional neural networks (CNN). When compared to other techniques, CNNs are faster, more reliable and robust. The proposed models have outperformed various state-of-the-art architectures, including ResNet50, SeResNet50, and Inception, in terms of accuracy (92%) and F1-score (92%). Researchers describe a fully autonomous brain tumor segmentation and classification model based on a Deep Convolutional Neural Network with a multiscale approach in this publication [3]. The suggested neural model can evaluate MRI scans with three types of tumors: meningioma, glioma, and pituitary tumor, in sagittal, coronal, and axial perspectives,

and does not require pre-processing of input images to remove skull or vertebral column components. In the study, their method had a tumor classification accuracy of 0.973, which was much greater than the other approaches by utilizing the same database.

Other than medical purposes, computer vision is used for navigation purpose. To enable FPV pedestrian navigation, researchers suggested a hybrid structure with a convolutional neural network (CNN) and local image characteristics in this paper [4]. AlphaMEX, a revolutionary end-to-end trainable global pooling operator was created to increase the CNN scene categorization accuracy. The efficiency of proposed method is demonstrated by an experimental data. On the ImageNet dataset, the proposed AlphaMEX-ResNet outperforms the original ResNet (k = 12) by 1.7 percent in terms of top-1 error rate. In this research [5], a fine-grained classification model based on deep learning called RMA (ResNet-Multiscale-Attention) is suggested to analyze the subtle and local distinctions among navigation mark kinds for performing the navigation mark recognition. The RMA has an accuracy of nearly 96% in classifying 42 types of navigation marks according to experimental results on a dataset by comprising 10260 navigation mark images and it outperforms the ResNet-50 model, which has an accuracy of about 94%.

Computer Vision has also been employed in biometric authentication. This article [6] has presented a facial recognition system, which facilitates a simple and quick searching for offenders by saving more time, and also it assists the police and administration more effectively. Face identification from the video has been performed by using a pre-trained model called FaceNet (FN) in this work. FN can gain 98.47 percent accuracy. This research [7] discusses about the facial biometrics system, in which multiple classifiers are utilized within the framework of face recognition. Multiple algorithms are employed in the process. All of them performed quite well, with an average accuracy of above 90%, while PCA+LDA+1N had the greatest average accuracy of 98%. The goal of this study [8] is to improve face recognition accuracy using the local phase quantization (LPQ) descriptor. To carry out the research, we suggest using the difference of Gaussians (DoG) to normalize facial pictures before encoding them with LPQ and classifying them using support vector machines. The suggested technique improved from 0.89% to 17.50% additional descriptors

and a combination of them, according to experimental findings obtained from three databases. This paper [9] provides a realistic alternative that uses the bare minimum of model design while matching the performance of deep learning techniques at the cutting edge. Their technique surpasses deep learning-based methods in various benchmark datasets and even excels those based on large data. The origins of ear modeling are review in this article [10]. The classification approach of deep learning utilizing CNN was compared to the frequently used machine-learning techniques. The average ear recognition rate for both the left and right ear was 92% in this article, thanks to deep learning utilizing Convolution Neural Network [CNN].

In this period of pandemic, deep learning-based solutions are heavily researched to overcome COVID-19. Researchers suggest CoroNet, a deep CNN model for automatically detecting COVID-19 infection from chest X-ray images, in this work [11]. The suggested model is built on the Xception architecture, which has been pre-trained on the ImageNet dataset, as well as the COVID-19 dataset and additional chest pneumonia X-ray imaging datasets. For 4-class instances, the suggested model had an overall accuracy of 89.6%, while for 3-class cases; it had an accuracy of 95%. This article [12] describes an automated COVID-19 detection method based on chest X-ray images that may be used in conjunction with the RT-PCR test to enhance diagnostic rates. Textural characteristics are retrieved from chest X-ray pictures and local binary pattern (LBP) based images in the suggested method. To test the robustness of the proposed technique, 2905 chest X-ray pictures of normal, pneumonia and COVID-19 infected people were analyzed on various class combinations. The created technique performs at a high level. Using two datasets comprising normal and COVID-19 positive pictures, Haque [13] presented a unique convolutional neural network model to detect COVID-19 patients. Using a second dataset, the suggested model has an accuracy of 98.3%. However, the model only considers binary categorization; it is unable to distinguish between COVID and non-COVID pneumonia cases.

In other domains also, deep learning and computer vision is used. Capsule networks using structured data exhibit optimal performance in visual inference areas. The categorization of hierarchical multi-label text is accomplished in this study [14] by using a basic capsule

network method. To illustrate its better performance, it is compared with the existing machine learning and deep learning algorithms. Extreme Learning Machine (ELM) is a new type of learning algorithm that may give a high recognition rate in a short amount of time. To complete the classification job, the ELM technique was created with the existence of a sigmoidal function of biases in the hidden nodes in this study article [15]. In addition, the suggested enhanced version of ELM improves accuracy and effectiveness in classification and regression issues. In this paper [16], a unique and consistent oversampling technique is presented to improve the classification performance particularly on binary unbalanced datasets. It is called as NMOTe, and it is considered as an improved and superior alternative to the existing approaches. In this paper, the performance of NMOTe on various typical datasets that are examined to get a statistical understanding of why it has surpassed the current state-of-the-art as the most resilient approach for tackling the two-class data imbalance problem. Deep Neural Networks (DNN) have recently shown a wide range of capabilities in the pattern recognition paradigm. Depth layer networks, filters, training, and testing datasets become as a part of this study [17] on DNN. When compared to current discoveries in neuroscience research, the suggested study outperforms them. For image analysis, image fusion has acquired a lot of traction in the medical and satellite imaging fields. Researchers have attempted to enhance the WOA algorithm by changing the WOA algorithm in this study [18]. Other metaheuristic optimization techniques develop a comparison based on the simulation and synthesis results. This study [19] incorporates the Support Vector Machine-based Cuckoo Search Algorithm to develop an accurate monitoring and reduce maintenance costs in the sector of energy generation by utilizing windmills. The results demonstrate that, the SVM-based CSO is more accurate than other current models that are involved in predicting the fault models.

Finally, Computer Vision is also used to detect different types of fruits. To improve the accuracy of automatic vegetable recognition and classification, this paper [20] presents a deep learning-based method for recognizing and classifying the vegetable images. The improved VGG network model was used to train the vegetable image dataset by using the open-source Caffe deep learning framework. The output characteristic of the first two completely linked layers has been proposed to be combined (VGG-M). The experimental verification indicated that the proposed method has achieved a recognition accuracy rate of

96.5% in the test dataset, which is much higher than the VGG network (92.1%) and AlexNet network (86.3%). Researchers suggested a realistic approach for performing fruit recognition in this study [21] by utilizing two newly developed classifiers, which are successful and efficient. They used EfficientNet and MixNet from two deep neural network families, which intend to create an expert system to reliably and quickly recognize the fruits. The results of the experiments demonstrate that, by using EfficientNet and MixNet on the dataset in question, it significantly increases the overall prediction accuracy when compared to a well-established baseline. This study [22] offers a support vector machine (SVM) classifier, which uses deep features derived from the fully connected layer of the convolutional neural network (CNN) model to classify 40 different Indian fruits. Six of the most powerful deep learning architectures are used in the trials. The findings of the assessment demonstrate that, the SVM classifier with a deep learning feature outperforms its transfer learning competitors. The VGG16 and SVM deep learning features produce the best results. Four deep learning recognition models were used in this study [23] to perform recognition trials on red and green apples under three lighting and two picture sharpness conditions along with transfer learning in order to speed up the training process. The findings revealed that, among the four detection models, the enhanced YOLOv3 model had obtained the best recognition impact.

## 2. PROPOSED WORK

The purposed deep learning-based image classification techniques' performance is analyzed with several classifiers for achieving its better outcome. The images utilized in this work are download from the Kaggle Fruits 360 dataset, which consists of 3334 images in the training folder with a group of Avocado, Banana, Cherry, Cocos, Kiwi, Mango, and Orange images. Similarly, 1121 images are available in test folders. Figure 2 shows few sample images that are used for training.



Figure 2. Sample Dataset Images

The proposed method classifies the fruit by detecting the most important features of the images by applying filters or feature detectors to the input image in order to generate the feature maps or the activation maps by using the activation function. Feature detectors or filters aid in the identification of various features present in a picture, such as edges, vertical lines, horizontal lines, bends, and so on.

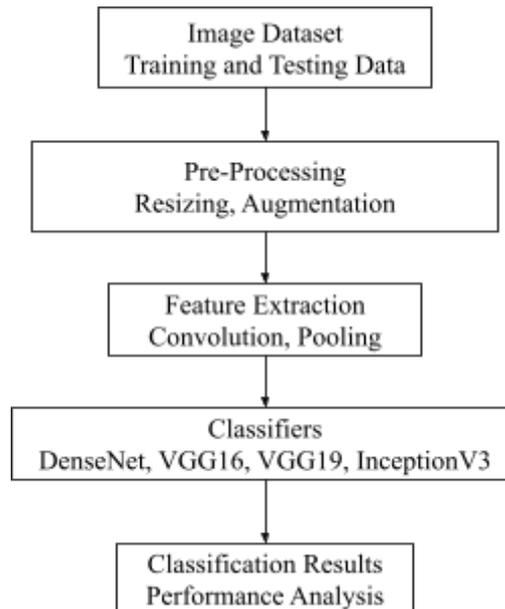


Figure 3. Overview of the Proposed System

Figure 3 represents the overview of the proposed system. Kaggle Fruits 360 dataset provides both training and testing datasets. The images from both datasets are read and resized to a height and width of 100 and 100 respectively. After that, the image data augmentation technique is used to expand the training dataset in order to improve the performance and ability to generalize the model. To extract the important features, convolutional and pooling layers are used.

A convolution is a combined integration of two functions that demonstrates how one modifies the other. Equation (1) and (2) is the mathematical representation of the operation. Figure 4 shows an example of a convolution operation.

$$(f * g)(t) = \int_{-\infty}^{\infty} f(\tau)g(t - \tau)d \tau \quad \text{Eqn. 1}$$

$$= \int_{-\infty}^{\infty} f(t - \tau)g(\tau)d\tau \quad \text{Eqn. 2}$$

Three major elements are involved in this operation: input image, feature detector, and feature map. The matrix representation of the input image is multiplied element-wise with the feature detector in order to gain a feature map. Another thing is stride, which is the shift of the number of pixels over the input image. Activation function is also used in conjunction with convolution. It is also considered as a critical component of the neural network since it introduces non-linear properties. This enables a neural network to learn complex and non-linear mapping between inputs and outputs. ReLU is the abbreviation for Rectified Linear Unit. If x is positive, it outputs x, otherwise, it outputs zero. It can be mathematically summarised as in equation (3).

$$A(x) = \max(0, x) \quad \text{Eqn. 3}$$

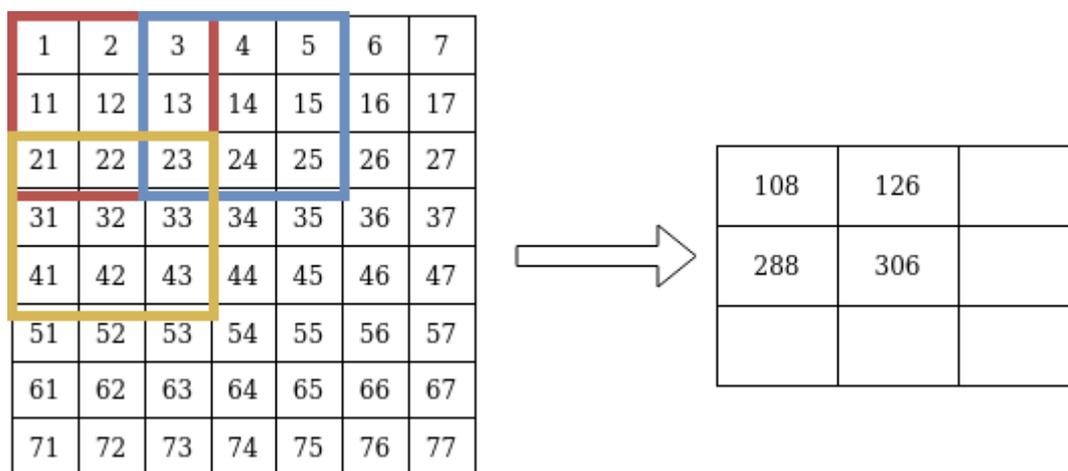


Figure 4. Convolution Operation with filter 3\*3 with stride 2

Pooling Layers will have reduced dimensions by combining the layers. As a result, the number of parameters to learn and the quantity of computation done in the network are reduced. The characteristics included in an area of the feature map created by a convolution layer are summarized by the pooling layer.

These layers are stacked with each other to form a single convolution layer. Different arrangements of these layers define various architectures along with the addition of the other components. Other useful components are batch normalization, dropout, etc.

VGG16 [24] is a convolutional neural network (CNN) architecture that won the 2014 ILSVR (Imagenet) competition. The most distinguishing feature of VGG16 is that, rather than having a huge number of hyper-parameters; it focuses on having 3x3 convolution layers with a stride 1 and always uses the same padding and maxpool layer of the 2x2 stride with 2 filters. Throughout the design, the convolution and max pool layers are arranged in the same way. It features two FC (completely connected layers) in the end, followed by a softmax for output. The 16 in VGG16 alludes to the fact that it contains 16 layers with different weights. As a result, the sole distinction between VGG16 and VGG19 represents the number of layers. Figure 5 and Figure 6 represents the architecture of the VGG16 and VGG19 that are used in this paper.

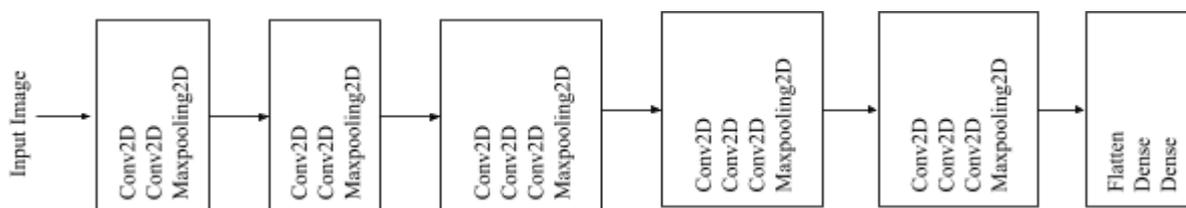


Figure 5. VGG16 architecture Used

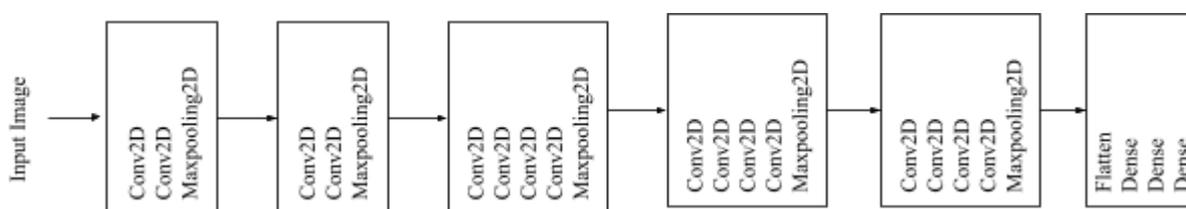


Figure 6. VGG19 architecture Used

Inception-v3 [25] is a convolutional neural network design obtained from the Inception family which uses Label Smoothing, Factorized 7 x 7 convolutions, and an auxiliary classifier to further transmit the label information down to the network, when compared to

other enhancements (along with the use of batch normalization for layers in the side head). Since the utilized Inception-V3 model architecture is too large, it is not included in this paper. Totally, it includes 159 layers, which consists of convolution, batch normalization, max pooling, concatenation, average pooling, and dense layer. The concatenation transfers information from one layer to another.

The name DenseNet [26] comes from the fact that each layer in a DenseNet design is linked to every other layer.  $L(L+1)/2$  direct connections exist for  $L$  layers. The feature maps of all previous layers are utilized as inputs for each layer, and their feature maps are used as inputs for subsequent layers. DenseNets simply connect every layer to every other layer, as easy as it may appear. This is the centralized concept, and it is also incredibly strong. The concatenation of feature maps from preceding layers is considered as the input of a layer within DenseNet. From Figure 7, it can be seen that each layer is interconnected with the other. There are a total of 121 layers in the proposed DenseNet architecture.

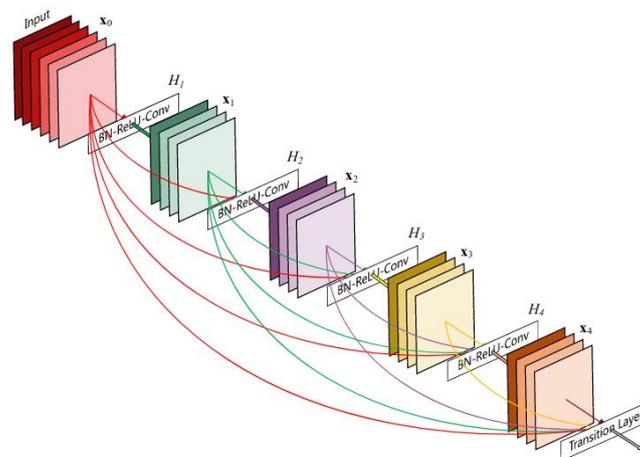


Figure 7. A 5-layer dense block with a growth rate of  $k=4$ . Each layer takes all preceding feature-maps as input (Huang G, 2017)

The proposed models are developed by using all these layers. Since our task is multi-class classification, the final model should be able to output the probable class of the fruit. To achieve this, the final output of the model is converted to a list of probabilities by using the softmax activation function. Softmax returns a vector, which includes the probability

distribution of a set of possible outcomes. The mathematical representation of the softmax is given in equation (4).

$$S(y_i) = \frac{e^{y_i}}{\sum_j e^{y_j}} \quad \text{Eqn. 4}$$

The main objective of the proposed model is to learn patterns from the given dataset. This will finally enable the model to correctly classify the new images into true class. To do that, the model losses associated with classifying should be minimized; this is controlled by employing an optimizer. Optimizers are algorithms or methods that are used for changing the neural network characteristics, such as weights and learning rate in order to minimize losses. Different types of optimizers are available; they are Gradient Descent, Momentum, Adagrad, RMSProp, etc. In this paper, an Adam optimizer is used.

The trained model is evaluated by using various metrics like Accuracy, Precision, Recall, F1-Score, and Confusion Matrix. One of the most significant measures used for assessing the model performance in classification tasks is accuracy. The correct observation rate of the proposed model is expressed as a percentage. Equation (5) represents the equation of the accuracy.

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total Number of Predictions}} \quad \text{Eqn. 5}$$

This equation can be further breakdown as represented in equation (6).

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad \text{Eqn. 6}$$

Where,

- TP represents True Positive, where both the actual and predicted class are true
- TN represents True Negative, where the model predicts the actual value, but the actual is Negative

- FP represents False Positive, where the model predicts the false value and the actual value is also negative
- FN represents False Negative, where the model predicts false value, but the actual value is positive

Precision is considered as the proportion of relevant outcomes that are properly classified by the model, whereas recall is the percentage of relevant results that are correctly classified by the model. Equation 5 and 6 represents the mathematical representation of the Precision and Recall.

$$Precision = \frac{TP}{TP+FP} \quad \text{Eqn. 7}$$

$$Recall = \frac{TP}{TP+FN} \quad \text{Eqn. 8}$$

Confusion Matrix is the visual representation of the performance of a statistical classification model. Figure 8 is the basic overview of the confusion matrix.

		Actual Class	
		Positive	Negative
Predicted Class	Positive	TP	FP
	Negative	FN	TN

Figure 8. Confusion Matrix

Finally, F1-Score considers both precision and recall to determine the model's accuracy. The F1 score attempts to adjust it by assigning more weight to obtain false negatives and false positives, while ignoring the large number of genuine negatives. Equation (9) represents the mathematical representation of F1-Score.

$$F1 - Score = 2 * \frac{precision*recall}{precision+recall} \quad \text{Eqn. 9}$$

#### 4. RESULTS AND DISCUSSION

The experiment has been performed by using Kaggle, which is a cloud-based system that includes Intel(R) Xeon(R) CPU with processor base frequency of 2.30GHz and 25.3GB RAM, wherein deep learning framework like Keras and Python 3.5 has been used. The images are gained from Kaggle Fruits 360 dataset. Table 1-4 specifies the confusion matrix for DenseNet, VGG16, VGG19, and InceptionV3 respectively. The performance of classifiers are analysed with 1121 images, which consist of 143 avocados, 166 bananas, 164 cherries, 166 cocos, 156 kiwis, 166 mangoes, and 160 oranges. Table 5 points out the performance of the classifiers in terms of precision, recall, and f1-score. Similarly, figure 9 shows the visual performance comparison of different models.

Table 1. Confusion Matrix of DenseNet

Apple	143	0	0	0	0	0	0
Avocado	0	166	0	0	0	0	0
Banana	0	0	164	0	0	0	0
Cherry	0	0	0	166	0	0	0
Cocos	0	0	0	0	156	0	0
Kiwi	0	0	0	0	0	166	0
Lemon	0	0	0	0	0	0	160
	Apple	Avocado	Banana	Cherry	Cocos	Kiwi	Lemon

Table 2. Confusion Matrix of VGG16

Apple	125	0	0	0	0	18	0
Avocado	0	135	0	0	31	0	0
Banana	0	0	164	0	0	0	0
Cherry	0	0	0	166	0	0	0
Cocos	0	45	0	0	111	0	0
Kiwi	0	0	0	0	0	166	0
Lemon	0	0	0	0	0	0	160
	Apple	Avocado	Banana	Cherry	Cocos	Kiwi	Lemon

Table 3. Confusion Matrix of VGG19

Apple	74	0	0	0	0	0	69
Avocado	0	154	0	0	12	0	0
Banana	0	0	164	0	0	0	0
Cherry	0	0	0	148	1	17	0
Cocos	0	42	0	0	114	0	0
Kiwi	0	0	0	0	0	166	0
Lemon	0	0	0	0	0	0	160
	Apple	Avocado	Banana	Cherry	Cocos	Kiwi	Lemon

Table 4. Confusion Matrix of InceptionV3

Apple	110	0	0	0	0	0	33	0
Avocado	0	104	3	0	53	3	3	0
Banana	0	0	164	0	0	0	0	0
Cherry	0	0	0	162	4	0	0	0
Cocos	0	0	0	20	136	0	0	0
Kiwi	1	0	0	0	0	0	165	0
Lemon	0	0	0	0	0	0	0	160
	Apple	Avocado	Banana	Cherry	Cocos	Kiwi	Lemon	

Table 5. Performance analysis of classifiers

Classifier	Accuracy	Precision	Recall	F1-Score
DenseNet	100	100	100	100
VGG16	97.30	98	98	98
VGG19	87.81	91	88	86
InceptionV3	93.55	94	94	93

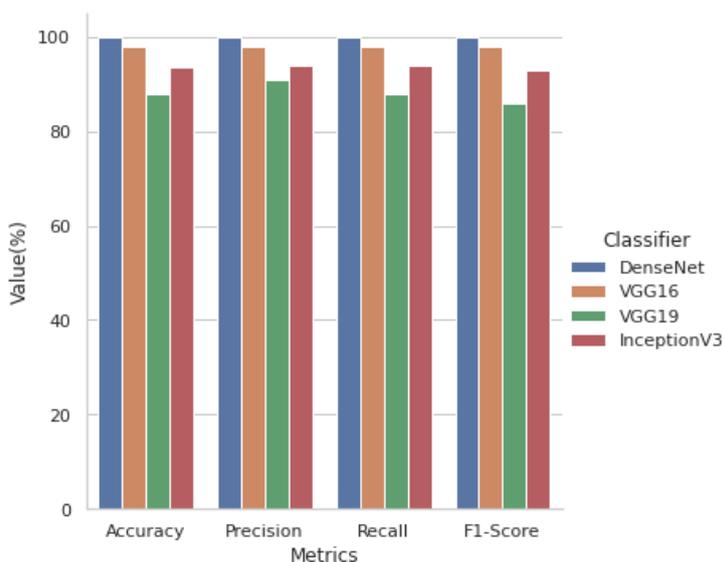


Figure 9. Performance comparison of Classifiers

From Table 5 and Figure 9, it can be observed that, DenseNet has classified the image with the highest metrics. The VGG19 based model performs poorly in all metrics along with the accuracy, precision, recall, and F1-Score value of 87.81, 91, 88, and 86 respectively. The confusion matrix also clearly shows the perfectness of the DenseNet in classifying the testing dataset. Other models also have the difficulty in classifying the multiple classes properly. Figure 9 provides a visual representation for the comparison of the classifiers by concerning various metrics. The comparison clearly shows that, the DenseNet and VGG16 delivers best performance while VGG19 performs poorly. Thus, among all the models, DenseNet and VGG16 perform optimally by resulting in a high value in all four metrics and VGG19 performs the worst.

DenseNet outperforms the previous models by eliminating the vanishing-gradient problem, which improves feature propagation, encourages feature reuse, and significantly reduces the number of parameters. The model can also be trained with fewer parameters. The smooth flow of data is ensured via dynamic feature propagation.

## 5. CONCLUSION

The proposed deep learning-based image classification system suggests DenseNet based model to classify the images effectively than the other classifiers by obtaining a training and testing accuracy of 99.25% and 100% respectively. The proposed methodology can be merged with a camera system in industries and factories to detect fruit types based on the gained image. The model performance can be optimized by employing a huge number of training and testing images. Nevertheless, the DenseNet based model has delivered a start-of-art performance, which can be employed in both industries and factories.

## References

- [1] Mohamed, Ashraf A, Ahmed A, Shalaby.(2019). “Digital imaging devices as sensors for iron determination”, *Food chemistry*, 360-367.

- [2] J. P. Zuluaga-Gomez, Z. A. Masry, K. Benagoune, S. Meraghni, N. Zerhouni.(2021).”A CNN-based methodology for breast cancer diagnosis using thermal images”, *Computer Vision and Pattern Recognition*.
- [3] Díaz-Pernas F.J, Martínez-Zarzuela M, Antón-Rodríguez M, González-Ortega D.(2021).”A Deep Learning Approach for Brain Tumor Classification and Segmentation Using a Multiscale Convolutional Neural Network”, *Healthcare*.
- [4] Zhao Q, Zhang BX, Lu SC, Zhang H, Sun D, Li GQ, Feng WQ.(2018).”A CNN-SIFT Hybrid Pedestrian Navigation Method Based on First-Person Vision”, *REMOTE SENSING*.
- [5] Pan, Mingyang, et al.(2020).”Visual recognition based on deep learning for navigation mark classification”, *IEEE Access*, 32767-32775.
- [6] Saibal Manna, Sushil Ghildiyal, Kishan kumar Bhimani.(2020).”Face Recognition from Video using Deep Learning”, *2020 5th International Conference on Communication and Electronics Systems (ICCES)*.
- [7] Nisar Ahmed, Farhan Ajmal Khan, Zain Ullah, Hasnain Ahmed, Taimur Shahzad.(2021).”Face Recognition Comparative Analysis Using Different Machine Learning Approaches”, *Advances in Science and Technology*, 265-272.
- [8] Chi-Kien Tran, Ngo Thi Thanh Hoa, Nguyen Thi Cam Ngoan, Lan-Anh Nguyen,(2021). ”SVM-Based Face Recognition through Difference of Gaussians and Local Phase Quantization”, *International Journal of Computer Theory and Engineering*, 1-10.
- [9] Yufei Zhao, Jinxin Yang, Jiangtao Du, Zhen Chen, Wen-Chi Yang.(2021).”A Lightweight Classifier for Facial Expression Recognition based on Evolutionary SVM Ensembles”, *2021 6th International Conference for Convergence in Technology (I2CT)*.
- [10] Aimee Booyens, Serestina Viriri.(2020). ”Exploration of Ear Biometrics with Deep Learning”. *Computer Vision and Graphics*”, *International Conference, ICCVG*, 14-16.
- [11] Asif Iqbal Khan, Junaid Latief Shah Mohammad Mudasir Bhat.(2020). ”CoroNet: A Deep Neural Network for Detection and Diagnosis of COVID-19 from Chest X-ray Images”, *Computer Methods and Programs in Biomedicine*.
- [12] Shishir Maheshwari, Rishi Raj Sharma, Mohit Kumar.(2021). ”LBP-based information assisted intelligent system for COVID-19 identification”, *Computers in Biology and Medicine*.

- [13] K. F. Haque, A. Abdelgawad.(2020). “A deep learning approach to detect COVID-19 patients from chest x-ray images”, *Artificial Intelligence* , 418-435.
- [14] Shakya, Subarna.(2020). "Performance analysis of wind turbine monitoring mechanism using integrated classification and optimization techniques", *Journal of Artificial Intelligence 2*, no. 01,31-41.
- [15] Dutta, Sayantan, and Ayan Banerjee. (2020). "Highly Precise Modified Blue Whale Method Framed by Blending Bat and Local Search Algorithm for the Optimality of Image Fusion Algorithm", *Journal of Soft Computing Paradigm (JSCP) 2*, no. 04,195-208.
- [16] Haoxiang, Wang, and S. Smys. (2021). "Overview of Configuring Adaptive Activation Functions for Deep Neural Networks-A Comparative Study", *Journal of Ubiquitous Computing and Communication Technologies (UCCT) 3*, no. 01, 10-22.
- [17] Chakrabarty, Navoneel, and Sanket Biswas.(2020). "Navo Minority Over-sampling Technique (NMOTe): A Consistent Performance Booster on Imbalanced Datasets", *Journal of Electronics 2*, no. 02, 96-136.
- [18] Mugunthan, S. R., and T. Vijayakumar. (2021)."Design of Improved Version of Sigmoidal Function with Biases for Classification Task in ELM Domain", *Journal of Soft Computing Paradigm (JSCP) 3*, no. 02, 70-82.
- [19] Manoharan, J. Samuel. (2021). "Capsule Network Algorithm for Performance Optimization of Text Classification", *Journal of Soft Computing Paradigm (JSCP) 3* , no. 01, 1-9.
- [20] H.W. Tao, L. Zhao, J. Xi, et al.(2014). “Fruits and vegetables recognition based on color and texture features”, *Trans. Chin. Soc. Agric. Eng*, 305-311.
- [21] L. T. Duong.(2020). “Automated fruit recognition using EfficientNet and MixNet”, *Computers and Electronics in Agriculture*.
- [22] Behera, Santi Kumari Rath, Amiya Kumar Sethy, Prabira Kumar.(2020). “Fruit Recognition using Support Vector Machine based on Deep Features”, *Karbala International Journal of Modern Science*.
- [23] Guantao Xuan, Chong Gao, Yuanyuan Shao, Meng Zhang, Yongxian Wang, Jingrun Zhong, Qingguo Li, Hongxing Peng.(2020). “Apple Detection in Natural Environment Using Deep Learning Algorithms”, *IEEE Access*, 216772-216780.

- [24] Karen Simonyan, Andrew Zisserman.(2015). “Very deep convolutional networks for large-scale image recognition”, *ICLR*.
- [25] Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z.(2016). “Rethinking the inception architecture for computer vision”, *2016 IEEE Conference on Computer Vision and Pattern Recognition*, 2818-26.
- [26] Huang G, Liu Z, Maaten L. V. D, Weinberger K. Q.(2017). “Densely connected convolutional networks”, *2017 In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

### Authors Biography



Milan Tripathi received a B.S. degree in computer engineering from Advanced College of Engineering and Management, Tribhuvan University, Kathmandu, Nepal, in 2019. He is currently working as an AI Researcher. He is guiding bachelor's and master's students in their project and thesis papers. To date, he has guided two master's students in their thesis projects. His research interests are computer vision, deep learning and image processing.