http://communications.science.ankara.edu.tr

Commun.Fac.Sci.Univ.Ank.Series A2-A3
Volume 66, Number 2, Pages 162-175 (2024)
DOI: 10.33769/aupse.1411294
ISSN 1303-6009 E-ISSN 2618-6462

Research Article; Received: December 28, 2023; Accepted: April 03, 2024

AUTOMATIC RECOGNITION OF COFFEE BEAN VARIETIES BASED ON PRE-TRAINED CNN ARCHITECTURES

Aynur YONAR[1] and Öznur ÖZALTIN[2]

[1]Department of Statistics, Selcuk University, Konya, TÜRKİYE
[2]Department of Statistics, Atatürk University, Erzurum, TÜRKİYE

Abstract. Coffee is an agricultural commodity of fundamental and considerable economic importance on the global market. In this study, the coffee bean varieties were examined from images via artificial intelligence due to their quality and value on the market. This study aims to create an automated system that can efficiently identify coffee beans without requiring a significant amount of time. In this study, five pre-trained Convolutional Neural Network (CNN) architectures were performed to detect four varieties of coffee beans through images. Extracting features from images is a challenging and specialized task. However, CNN possesses the ability to extract features automatically. Therefore, these architectures were employed as both deep feature extractors and classifiers. Primarily, 1600 coffee beans' images were split into 75:25 training and testing sets. Next, 5-fold cross-validation was applied during the training process. This study presented both validation and testing results. Eventually, ShuffleNet achieved the best classification performance with 99.33% and 99.75% accuracy rates in identifying types of coffee beans for the training and testing sets, respectively. As a result, this study has demonstrated that deep learning technologies can automatically recognize the different types of coffee beans.

1. Introduction

Coffee holds a crucial role as an economic crop, significantly influencing global trade and agriculture, and it ranks among the most widely consumed beverages across the globe. Since coffee prices are directly influenced the quality and type of coffee, separating coffee beans is very important for world markets. Assessing the

Keywords. Artificial intelligence, agricultural economics, classification, coffee beans, convolutional neural networks, cross-validation, pre-trained architectures.
aynursahin@selcuk.edu.tr-Corresponding author; 0000-0003-1681-9398
doanoznur09@gmail.com; 0000-0001-9841-1702.

color of coffee beans plays a vital role in establishing their quality and market value [1, 2].

This evaluation is typically carried out through visual examination or conventional tools, yet these approaches come with challenges like inconsistency, time intensity, and subjectivity. Using computer vision systems instead of traditional methods is an excellent alternative to eliminate these adverse situations. Computer vision systems provide more accurate, impartial, and sensitive classification results [3, 4].

Computer vision is the field of computer science that deals with technologies that enable computers to identify and manipulate objects they see like humans. It covers the subjects of acquiring the image, processing, analyzing, understanding, extracting numerical data from the images, and making decisions.

In recent times, deep learning methods have found extensive application in the field of computer vision and can extract more detailed information than machine learning methods. In deep learning, generally considered a black-box approach, features are automatically determined with input given as images [5]. Then, classification is performed using these features.

The wide range of colors in coffee beans poses a challenge for their classification through visual inspection. Hence, to tackle the classification of coffee beans, various deep learning algorithms, including Convolutional Neural Networks (CNN) architectures, have been suggested in the literature. Some of these are presented below.

Unal *et al*. [6] created a specialized data set containing 1554 images of 3 unique coffee types: Espresso, Kenya, and Starbucks Pike Place coffee beans, and classified them via 4 different CNN-based models: SqueezeNet, Inception V3, VGG16, and VGG19. The findings indicated that SqueezeNet emerged as the most successful model, achieving the highest average classification success rate of 87.3%.

De Oliveira *et al*. [1] proposed a computer vision system based on artificial neural network (ANN) and Bayes classifier to analyze and categorize four green coffee bean types: whitish, cane green, green, and bluish green. The results indicated that the system achieved a 100% accuracy rate in categorizing variations in the color of green coffee beans.

Jumarlis *et al*. [7] provided a website to detect the level of the coffee beans utilizing image input through a web-based program used GLCM (gray-level co-occurrence matrix) and the K-NN (k-Nearest Neighbor) methods, and the system provides 90% accuracy.

Arboleda [8] classified green coffee beans using 22 data mining algorithms consisting of decision tree, discriminant analysis, support vector machine (SVM), K-NN, and ensembles families and obtained the highest classification accuracy with 94.1% by Coarse Tree Algorithm.

Fukai *et al.* [9] developed an automatic coffee bean sorting system for coffee bean producers using deep CNN to detect the type of coffee beans and trained the ANN to implement into Raspberry Pi compute module with the camera module. They compared the results of conventional linear SVM. CNN gives better performance than SVM.

Huang *et al*. [10] designed an automated system for identifying green coffee beans, employing image processing and data augmentation technologies to handle the data, and utilizing CNN to analyze image information. Following the research, a classification accuracy of 94.63% was achieved.

Gope and Fukai [2] classified green coffee bean images through CNN and SVM. Firstly, they created four trained CNN models corresponding to different image sizes, including 32×32 pixels, 64×64 pixels, 128×128 pixels, and 256×256 pixels, to compare the classification accuracies of CNN and traditional linear SVM for normal and pea berry coffee beans. The CNN yielded a notable accuracy rate of 96.71%.

Santos *et al.* [11] used SVM, Deep Neural Network (DNN) and Random Forest (RF), to assess defects in coffee beans. They concluded that all classification models performed similarly. In addition to these studies, state-of-the-art studies were also examined, as detailed in Table 1.

The research is focused on efficiently identifying the type of coffee beans through the utilization of deep learning algorithms. In this context, five different CNN-based pre-trained architectures, AlexNet, Inception-v1, MobileNet-v2, ShuffleNet, and SqueezeNet, have been used to classify coffee beans. The advantages and contributions of this study is as follows:

(i) This study automatically shows recognition of coffee bean types via pre-trained CNN.
(ii) Expert opinion is not needed to extract features.
(iii) To find the optimal architecture, AlexNet, Inception-v1, MobileNet-v2, ShuffleNet, and SqueezeNet are compared by using distinct performance metrics on training and testing sets.
(iv) ShuffleNet has the highest performance on training and testing set to determine coffee beans.
(v) This study may shed light on the determination of the quality and diseases of agricultural products.

The pipeline of this study is shown in Figure 1. The performance of the techniques has been compared using AUC, accuracy, sensitivity, specificity, precision, F1-score, G-Mean metrics, and ROC curves. Finally, the model having the highest success is determined.

TABLE 1. State-of-the-art studies.

Study	Coffee Classes	Methods	Metrics
Tsai *et al.* [12]	2 classes	Mass spectrometry (MS) analysis + ANN	Accuracy 0.9958 Sensitivity 0.9875 Specificity 1.0000
Arboleda [13]	2 classes	Feature extraction + K-NN	Accuracy 0.9700
Raveena and Surendran [14]	6 classes	ResNet50 VGG16	Accuracy 0.9897 Sensitivity 0.9844 F1-score 0.9864 Accuracy 0.9638 Sensitivity 0.9523 F1-score 0.9563
Kim et al. (2024) [15]	2 classes	CNN Based Model	Accuracy 0.9927
Chang and Liu (2024) [16]	8 classes	CNN Based Model VGG-16 ResNet-18 AlexNet GoogleNet	Accuracy 0.9600 Kappa 0.9500 Accuracy 0.8100 Kappa 0.7900 Accuracy 0.8900 Kappa 0.8490 Accuracy 0.8900 Kappa 0.8000 Accuracy 0.9200 Kappa 0.9000

The remainder of this paper is organized as follows: In the "Material and Methods" section, details about the coffee beans dataset, CNN and pre-trained architectures, cross-validation, confusion matrix, and performance measures are provided. The "Experimental Results" section presents the outcomes, while the "Conclusion" section offers concluding remarks.

2. Materials and Methods

This section provides concise information about the coffee beans dataset, CNN, pre-trained architectures, cross-validation, and performance measures utilized in the study.

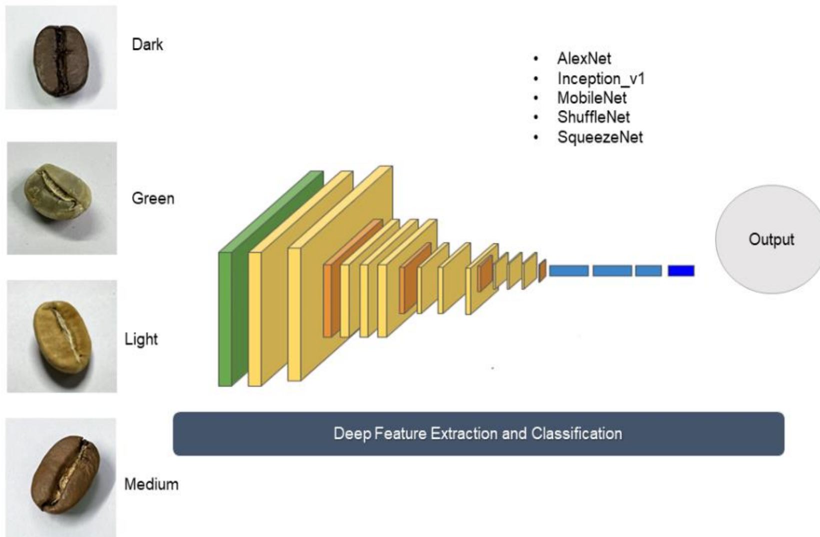


FIGURE 1. Pipeline of this study.

2.1. **Coffee Beans Dataset.** The coffee beans dataset used in this study were from study of [17]. In addition, the dataset is publicly available on the Kaggle platform. There are four different coffee classifications: dark, green, light, and medium. There are 1600 images in total, 400 in each class. Each example bean's image is 224x224x3 pixels in size, and images of four different coffee beans were used to recognize the coffee type via deep learning. This study's dataset is split into 75% training and 25% testing sets.

2.2. **Convolutional neural network (CNN).** The convolutional neural network (CNN) stands out as a prominent deep learning method characterized by its intricate structure composed of multiple layers. CNN are commonly applied to address image processing challenges due to their capability to conduct feature extraction, learning, and classification based on these extracted features. Moreover, CNN overcomes the computational complexity problem that other classification algorithms have in real-time data and provides very good classification results in studies involving both large and small datasets.

CNN processes an image in different layers and then extracts all its features. The most used layers are given below [6, 18, 19]:

Convolution Layer: This layer serves as the fundamental building block of CNNs, extracting features by systematically applying different filters to the image. To effectively determine the increasing number of features, it is essential to augment both the number of steps and filters in conventional layers at a proportional rate. However, as the number of features increases, learning becomes more difficult for the network, so this number must be determined optimally.

Pooling Layer: This layer simplifies big data from the convolution layer by preserving their existing properties to reduce programming complexity and improve learning.

Activation Layer: This layer, also known as the non-linear layer, activates the system with non-linear functions and prevents values from falling outside the valid data range.

Fully Connected Layer: This layer is the crucial artificial neural network layer within CNNs, playing a pivotal role in the learning processes and feature extraction.

SoftMax Layer: The distribution of classes becomes apparent, and an output is generated through the labeling process within this layer.

In the context of this study, five distinct pre-trained CNN architectures—AlexNet, Inception_v1, MobileNet_v2, ShuffleNet, and SqueezeNet were employed for the classification of coffee bean varieties.

AlexNet was proposed by [20] for image classification. The network won the ImageNet Large-Scale Visual Recognition Competition (ILSVRC) 2012 with more than 26% accuracy over contemporary models. It comprises eight trainable layers, including five convolutional layers and three fully connected layers. The last layer of the fully connected layer is associated with a SoftMax classifier configured for N classes, where N denotes the number of classes. The network employs multiple convolutional kernels for the extraction of features from the image. Additionally, it incorporates dropout for regularization and a rectified linear unit (ReLU) activation function to expedite training convergence [21].

Inception_v1 model, which was the winner of ILSVRC in 2015, has a significant success in the development of CNN classifiers. Convolution occurs with three sizes of filters (1×1, 3×3, 5×5) at the same level with maximum pooling. The outputs from this more comprehensive layer are combined and fed as input to the next layer.

MobileNet_v2 was proposed by Sandler et al. in 2018. MobileNetv2 comprises one convolutional layer, succeeded by 19 residual bottleneck modules, and subsequently, two convolutional layers. The bottleneck module incorporates a shortcut connection exclusively when the stride is set to 1. Shortcut is not used for higher pitch due to size difference. They also used ReLU as a nonlinear function instead of simple ReLU to limit the calculations.

ShuffleNet, introduced by [22], is primarily composed of a standard convolution and a series of ShuffleNet units organized into three stages. The ShuffleNet unit bears resemblance to the ResNet block, employing depth convolution on 3x3 layers and substituting the 1x1 layer with point group convolution. The depth convolution layer is preceded by a channel blending process.

SqueezeNet stands out as a more compact and innovative CNN architecture, characterized by fewer parameters compared to other CNN models. SqueezeNet consists of fifteen layers, including two convolution layers, three max-pooling layers, eight fire layers, a global average pooling layer, and a SoftMax layer with an output layer. Fire layers create compression and expansion between convolution layers. SqueezeNet is an excellent candidate to improve the hardware efficiency of neural network architectures. Details of pre-trained architectures used in this study are given in Table 2.

2.3. **Cross-Validation.** During the separation of the data set into the training and test sets, irregular distribution of the data set may negatively affect the model's performance. This problem can be solved with the k-fold cross-validation method. The dataset is partitioned into segments represented as *k* folds in cross-validation. Subsequently, k-1 folds are trained in the framework and tested on the remaining folds at each step. The critical point here is to use the previously untested part as the test set in each step [24]. We used 5-fold cross-validation for train data in this study.

TABLE 2. Details of pre-trained architectures used in this study [23].

Pre-trained architectures	Features		
	Parameters (millions)	**Input Image Size**	**Depth**
AlexNet	61	227x227	8
Inception_v1	7	224x224	22
MobileNet_v2	3.5	224x224	53
ShuffleNet	1.4	224x224	50
SqueezeNet	1.24	227x227	18

2.4. **Confusion matrix and Performance measures.** The confusion matrix illustrates the current state of the dataset, presenting the count of both accurate and inaccurate predictions made by the classification model in a tabular format. For the evaluation of classification model performance in this study, a four-class confusion matrix was utilized. Moreover, six performance metrics derived from the confusion matrix were utilized to analyze the results of the experimental study, as detailed in Table 3. Additionally, the distinctiveness of the results was assessed using the values of the receiver operating characteristic (ROC) curves and AUC (area under the curve). In the ROC curve, the false positive rate is represented on the x-axis, while the true positive rate is depicted on the y-axis. An AUC value approaching 1 signifies high classification performance for the method [25].

TABLE 3. Performance metrics formulas. TP : True Positive, TN : True Negative, FP : False Positive, FN : False Negative.

Metric	Formula
Accuracy	$ACC = \frac{TP+TN}{TP+TN+FP+FN}$
Sensitivity	$TPR = \frac{TP}{TP+FN}$
Specificity	$TNR = \frac{TN}{TN+FP}$
Precision	$PPV = \frac{TP}{TP+FP}$
F1-Score	$F1\ score = \frac{2}{\frac{1}{TPR}+\frac{1}{PPV}}$
G-Mean	$\sqrt{TPR \times TNR}$

3. Experimental Results

Section-Times New Roman 11, Figure -illustrated in Figure 1. This study focuses on detecting coffee bean types via artificial intelligence. The acquired images of coffee beans were categorized employing five distinct pre-trained CNN models: AlexNet, Interception v1, MobileNet_v2, ShuffleNet, and SqueezeNet. The models were trained in the MATLAB environment, utilizing an Intel Core i7-7500U CPU, NVIDIA GeForce GTX 950M, 16 GB RAM, and a 64-bit Operating System.

The dataset was divided into 75% train and 25% test sets. 5-fold cross-validation was applied to the training set to obtain confident results.

The confusion matrix of each CNN model is presented in Figure 2. There are correct and incorrect classification numbers for each algorithm. For example, the AlexNet model correctly classified 92, 46, 97, and 95 images of Dark, Green, Light, and Medium coffee beans, respectively. The category with the highest misclassification of coffee bean images is the green coffee bean class.

The performance of the models was evaluated using metrics including accuracy, sensitivity, specificity, precision, recall, and F-1 score, along with G-Mean. Table 4 and 5 show 5-fold cross-validation and test results for each algorithm, respectively. All performance matrices achieved the highest value with ShuffleNet for validation. However, when the architectures were tested, the best optimal performance was obtained with MobileNetv2 and ShuffleNet to detect coffee bean varieties. As a result, ShuffleNet was found to be the most efficient net in terms of validation and testing data results. Because the difference between the validation and test results is as small as possible, the algorithm does not learn excessively. Therefore, it was the best one. In Table 4 and Table 5, bold values display the highest performance metrics.

To have information about the distinctiveness of the models, ROC curves are drawn and given in Figure 3. It is observed that the ShuffleNet model exhibits the highest level of distinctiveness.

TABLE 4. Average performance metrics of all models for validation data.

CNN Algorithm	**AUC**	**Accuracy**	**Sensitivity**	**Specificity**	**Precision**	**F1-score**	**G-Mean**
AlexNet	0.9717	0.8025	0.8025	0.9342	0.8270	0.8043	0.8658
Inception_v1	0.9992	0.9892	0.9892	0.9964	0.9894	0.9892	0.9928
MobileNet_v2	0.9995	0.9842	0.9842	0.9947	0.9847	0.9842	0.9894
ShuffleNet	**0.9999**	**0.9933**	**0.9933**	**0.9978**	**0.9933**	**0.9933**	**0.9956**
SqueezeNet	0.9972	0.9592	0.9592	0.9864	0.9593	0.9591	0.9727

TABLE 5. Average performance metrics of all models for test data.

CNN Algorithm	**AUC**	**Accuracy**	**Sensitivity**	**Specificity**	**Precision**	**F1-score**	**G-Mean**
AlexNet	0.9918	0.8250	0.8250	0.9417	0.8796	0.8179	0.8814
Inception_v1	**1.0000**	0.9950	0.9950	0.9983	0.9951	0.9950	0.9967
MobileNet_v2	**1.0000**	**0.9975**	**0.9975**	**0.9992**	**0.9975**	**0.9975**	**0.9983**
ShuffleNet	**1.0000**	**0.9975**	**0.9975**	**0.9992**	**0.9975**	**0.9975**	**0.9983**
SqueezeNet	0.9998	0.9675	0.9675	0.9892	0.9706	0.9679	0.9783

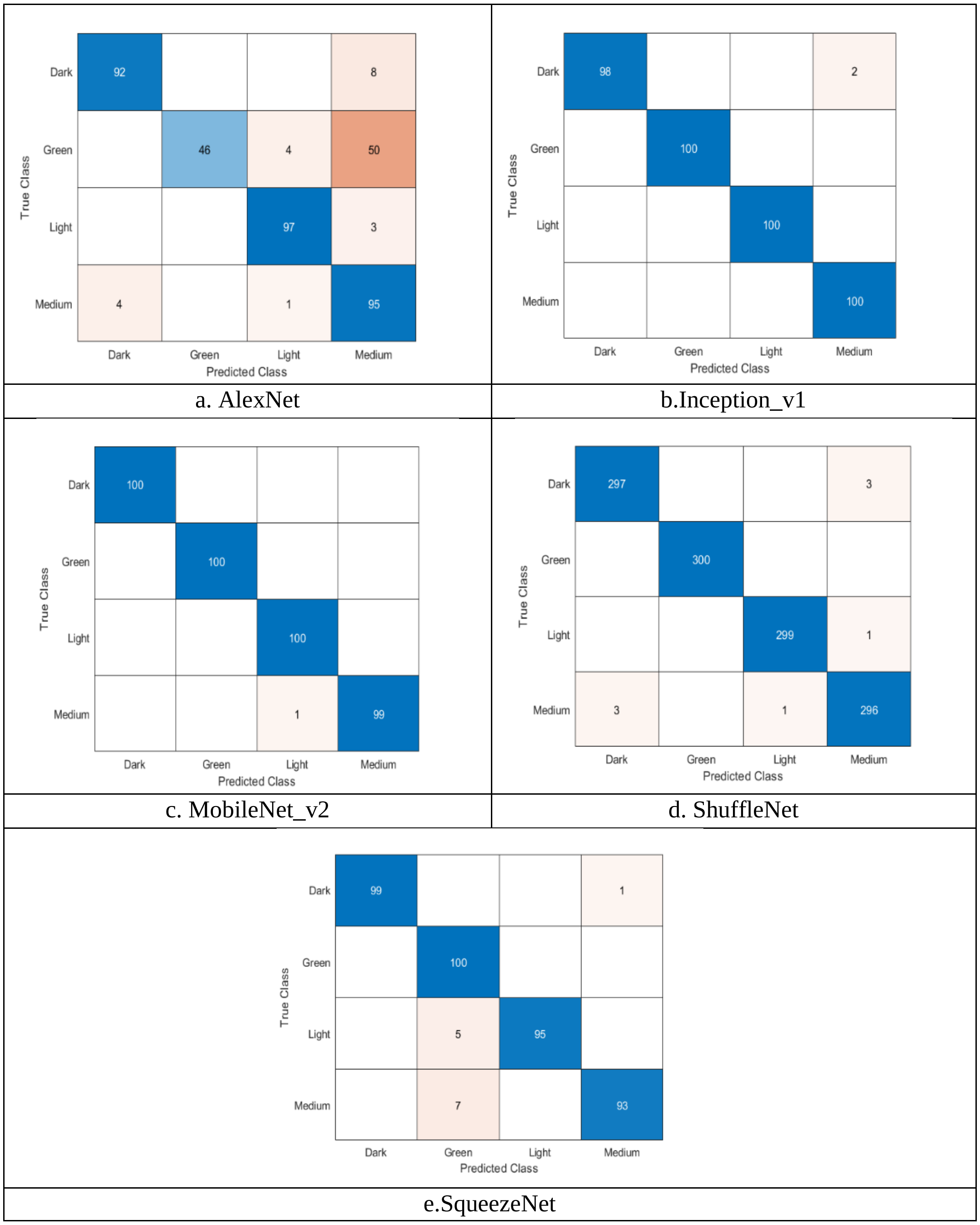
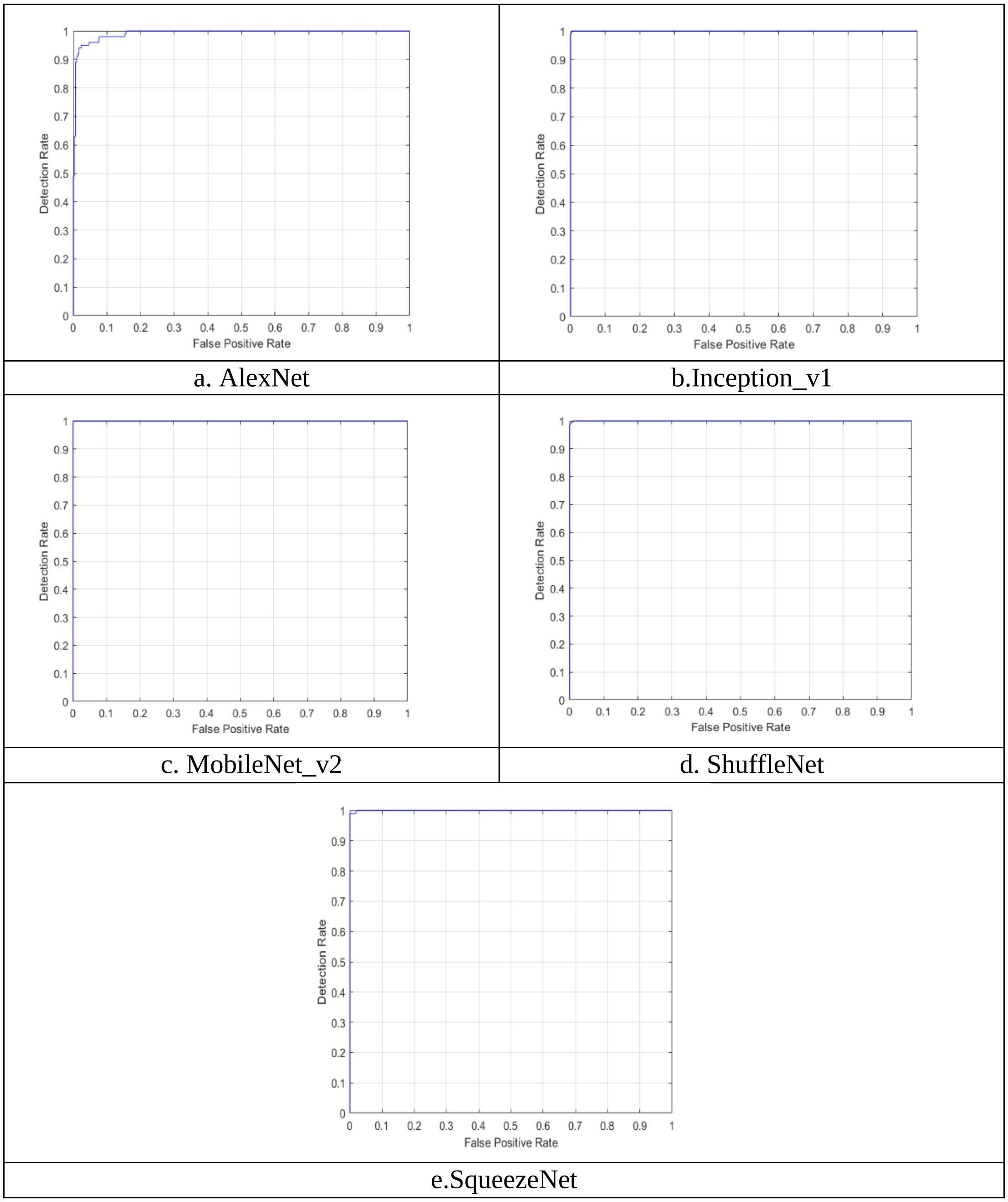


FIGURE 2. Confusion matrixes of CNN algorithms.



a. AlexNet

b.Inception_v1

c. MobileNet_v2

d. ShuffleNet

e.SqueezeNet

FIGURE 3. ROC Curves of CNN algorithms.

5. Conclusions

In this study, five distinct CNN architectures—AlexNet, Inception_v1, MobileNet_v2, ShuffleNet, and SqueezeNet—were employed for the classification of coffee beans. The coffee bean dataset was initially split into a training set comprising 75% of the data and a testing set consisting of the remaining 25%. Additionally, a training set and validation set were created to prevent overfitting by applying 5-fold cross-validation to the training set. For the validation set, the best architecture, according to all performance metrics, was determined to be ShuffleNet. Among the testing set utilized for assessing model performance, MobileNet_v2 and ShuffleNet demonstrated the highest success in classifying the test set. However, it is expected that the difference between the validation or training result and the test results will be minimal. In this study, Shufflenet was identified with the smallest difference as the top choice. Thus, the utilization of ShuffleNet architecture in detecting coffee beans can streamline quality control processes and minimize decision-making errors.

This study has some constraints:

(i) Types of coffee bean is limited. In addition to the identification of species, diseases could also be identified.

(ii) If the same images had features extracted by experts, comparisons could be made with other artificial intelligence algorithms.

In future studies, different existing or newly proposed architectures and other coffee bean or agricultural image datasets can be used for comparison purposes.

Author Contribution Statements The authors contributed equally to this work.

Declaration of Competing Interests The authors declare that there is no conflict of interest regarding the publication of this paper.

References

[1] De Oliveira, E. M., Leme, D. S., Barbosa, B. H. G., Rodarte, M. P., Pereira, R. G. F. A., A computer vision system for coffee beans classification based on computational intelligence techniques, *J. Food Eng.,* 171 (2016), 22-27.

[2] Gope, H. L., Fukai, H., Normal and pea berry coffee beans classification from green coffee bean images using convolutional neural networks and support vector machine, *Int. J. Comput. Inf. Eng.*, 14 (6) (2020), 189-196.

[3] Adiwijaya, N. O., Romadhon, H. I., Putra, J. A., Kuswanto, D. P., The quality of coffee bean classification system based on color by using k-nearest neighbor method, *J. Phys.: Conference Series*, 2157 (2022).

[4] Vogt, M. A. B., Developing stronger association between market value of coffee and functional biodiversity, *J. Environ. Manage.*, 269 (2020).

[5] Buhrmester, V., Münch, D., Arens, M., Analysis of explainers of black box deep neural networks for computer vision: *A survey, Mach. Learn. Knowl. Extr.*, 3 (4) (2021), 966-989.

[6] Unal, Y., Taspinar, Y. S., Cinar, I., Kursun, R., Koklu, M., Application of pre-trained deep convolutional neural networks for coffee beans species detection, *J. Food Anal. Method*, 15 (12) (2022), 3232-3243.

[7] Jumarlis, M., Mirfan, M., Manga, A. R., Classification of coffee bean defects using gray-level co-occurrence matrix and k-nearest neighbor, *ILKOM J. Ilmiah*, 14 (1) (2022), 1-9.

[8] Arboleda, E. R., Comparing performances of data mining algorithms for classification of green coffee beans, *J Int. J. Eng. Adv. Technol*, 8 (5) (2019), 1563-1567.

[9] Fukai, H., Furukawa, J., Katsuragawa, H., Pinto, C., Afonso, C., Classification of green coffee beans by convolutional neural network and its implementation on raspberry Pi and Camera Module, *Timor. Acad. J. Sci.*, 1 (2018), 1-10.

[10] Huang, N. F., Chou, D. L., Lee, C. A., Wu, F. P., Chuang, A. C., Chen, Y. H., Tsai, Y. C., Smart agriculture: real-time classification of green coffee beans by using a convolutional neural network, *JIET Smart Cities*, 2 (4) (2020), 167-172.

[11] Santos, F. F. L. d., Rosas, J. T. F., Martins, R. N., Araújo, G. d. M., Viana, L. d. A., Gonçalves, J. d. P., Quality assessment of coffee beans through computer vision and machine learning algorithms, *Coff. Sci.*, (2020).

[12] Tsai, J.-J., Chang, C.-C., Huang, D.-Y., Lin, T.-S., Chen, Y.-C., Analysis and classification of coffee beans using single coffee bean mass spectrometry with machine learning strategy, *Food Chem.*, (2023), 426, https://doi.org/10.1016/j.foodchem.2023.136610.

[13] Arboleda, E., Classification of immature and mature coffee beans using texture Features and medium K nearest neighbor, *J. Artif. Intell. Technol*, 3 (3) (2022), 114-118, https://doi.org/10.37965/jait.2023.0203.

[14] Raveena, S., Surendran, R., ResNet50-based classification of coffee cherry maturity using deep-CNN, *2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT)*, Tirunelveli, India, (2023), 1275-1281, https://doi.org/10.1109/ICSSIT55814.2023.10061006.

[15] Kim, Y., Lee, J., Kim, S., Study of active food processing technology using computer vision and AI in coffee roasting, *Food Sci. Biotechnol.*, (2024), 1-8, https://doi.org/10.1007/s10068-023-01507-7.

[16] Chang, S.-J., Liu, K.-H., Multiscale defect extraction neural network for green coffee bean defects detection, *IEEE Access*, 12 (2024), 15856-15866, https://doi.org/10.1109/ACCESS.2024.3356596.

[17] Ontoum, S., Khemanantakul, T., Sroison, P., Triyason, T., Watanapa, B., Coffee roast intelligence, arXiv: 2206.01841, (2022).

[18] Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., Lew, M. S., Deep learning for visual understanding: A review, *Neurocomputing*, 187 (2016), 27-48.

[19] Ozaltin, O., Yeniay, O., A novel proposed CNN–SVM architecture for ECG scalograms classification, *Soft Comput.*, 27 (8) (2023), 4639-4658.

[20] Krizhevsky, A., Sutskever, I., Hinton, G. E., Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.*, 25 (2012).

[21] Zaidi, S. S. A., Ansari, M. S., Aslam, A., Kanwal, N., Asghar, M., Lee, B., A survey of modern deep learning-based object detection models, *Dig. Signal Process.*, 126, (2022).

[22] Zhang, X., Zhou, X., Lin, M., Sun, J. Shufflenet: An extremely efficient convolutional neural network for mobile devices, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2018).

[23] Mathworks, 2023, https://www.mathworks.com/help/deeplearning/ug/pretrained-convolutional-neural-networks.html.

[24] Yadav, S., Shukla, S., Analysis of k-fold cross-validation over hold-out validation on colossal datasets for quality classification, *2016 IEEE 6th International Conference on Advanced Computing (IACC)*, (2016).

[25] Gorunescu, F., Gorunescu, F., Classification performance evaluation, *Data Mining: Concepts, Models Techniques*, (2011), 319-330.