Comparative Analysis of Deep Convolutional Neural Network for Accurate Identification of Foreign Objects in Rice Grains

Aji Setiawan, Member, IAENG, Kusworo Adi, and Catur Edi Widodo

Abstract— **The presence of foreign objects in rice collections is known to pose significant concerns for both food safety and product quality. Neglecting to address these issues can result in a deterioration of rice quality, a decline in its economic value, and a loss of consumer trust and confidence. Therefore, this research introduced a novel model for the recognition of natural and manufactured foreign objects through the use of Deep Convolutional Neural Network (DCNN). The model focused on analyzing images of rice for easy identification and classification. DCNN model classified the images into six groups based on different types, namely stone, paddy, fragment rice, broken, broken-yellow, and black-red. These classifications were determined using three pre-trained models, such as ResNet50, VGG16, and MobileNetV2. This research also used comparative techniques for multi-class classifications, combining multiple machine learning with digital image processing measurements and comparative performance. Despite using a cross-validation model, the experiment proved that using a pre-trained technique with transfer learning produced higher accuracy than a typical machine learning model. The most accurate prediction was made by VGG16 using a transfer learning of 97% compared to a random forest (RF) value of 94% without cross-validation with 5 K-fold and 10 K-fold cross-validations.**

Index Terms— **Deep Convolutional Neural Network (DCNN), pre-trained, foreign object, transfer learning, cross-validation.**

I. Introduction

RICE is known to play a crucial role in the diets of over half the global population, providing approximately 20 percent of their daily calorie intake. Asia leads the world in rice consumption, with a staggering rate of 90%, while Africa and Latin America are witnessing the most increasing rate as a result of population and economic growth. Given the substantial worldwide demand for commodities, the importance of quality control as one of the stages of food safety and quality assessment is gaining global attention. Public health is highly associated with security considerations, making it a crucial aspect of societal development [1]. Government and cultural stakeholders have united in agreement to prevent food contamination and damage [2]. Food safety guarantees that primary ingredients remain free from substances that could diminish their nutritional value. This effort requires a wide range of scientific models, including ensuring hygiene by preventing foreign bodies from gaining entrance during food processing. Good Agricultural and Post Harvest Management Practices have a positive and substantial effect on the quality of plant production.

The traditional technique for evaluating food safety and quality relies solely on the senses of sight and touch, making this model time-consuming and potentially leading to the alteration of the object's original form. Morphological characteristic is generally used to classify objects, describing the features of the size and shape. In agricultural industries, the size, shape, color, and texture of materials are correlated with the price. Therefore, accurate food classification necessitates the use of proper models and advanced technology capable of detecting avoidable alterations in order to meet specifications.

Characterizing the shape and size of food items can be effectively achieved through an image processing model, which includes measuring the length and the area occupied by foreign objects. The calcium ratio is derived from the rice gain area or rectangle bounding area value, and product quality is predicted using a support vector machine (SVM) with weight parameters eccentricity, chalk ratio, and elongation. Moreover, the color characteristic can be used to determine the quality of beef. For instance, when analyzing fat content in beef, converting color variation into a binary image using the Otsu threshold and applying a decision tree (DT) classification can be a valuable model for quality assessment. Otsu, mean, adaptive, and Gaussian thresholding are models that can convert binary image colors.

The presence of foreign objects in food is one of the leading causes of customer rejection and product withdrawal [3]. It is harmful, leading to a loss of consumer confidence in a brand and high product recall costs. Some examples of non-natural foreign items include small stones or pebbles, glass, metal, or rubber, whereas natural foreign objects include fruit or food skin, insects, and branches. The inadvertent inclusion of foreign objects in food or product packaging can lead to unintentional consumption and, in rare instances, necessitate surgical removal of foreign objects [4].

Convolutional Neural Network (CNN) is a sequence of layered architectures in the notion of deep learning, used particularly in the classification of images. The current

Manuscript received June 5, 2023; revised December 5, 2023. This work was funded in part by Darma Persada University, grant no 002/PEN/LP2MK/UNSADA/XI/2023.

A. Setiawan is a Doctoral candidate of Information Systems Department, Diponegoro University, Semarang, 50275, Indonesia. (corresponding author to provide phone: +62-878-8502-5203; e-mail: aji_setiawan@ft.unsada.ac.id).

K. Adi is a Professor in the Doctoral Information Systems Department, Diponegoro University, Semarang, 50275, Indonesia, (e-mail: kusworoadi@lecturer.undip.ac.id).

C. E. Widodo is a Senior Lecturer in the Doctoral Information Systems Department, Diponegoro University, Semarang, 50275, Indonesia, (e-mail: caturediwidodo@lecturer.undip.ac.id).

usage of deep learning has developed for variety and object recognition using Faster R-CNN [5][6], SSD [7], YOLO [8], and video-based classification [9]. In recent years, agricultural research has used CNN to detect foreign objects in walnuts [10], mango [11], plant disease [12], almonds [13], non-food, aeroplane trajectories [14], and coal [15]. Some research p published results on user acquisition processes with non-invasive models, such as x-rays, gamma, and infrared, to help ensure food quality and safety and prevent health hazards [16][17]. According to this research, the emphasis is placed on identifying foreign objects within the rice group, with a particular focus on deep learning-based categorization.

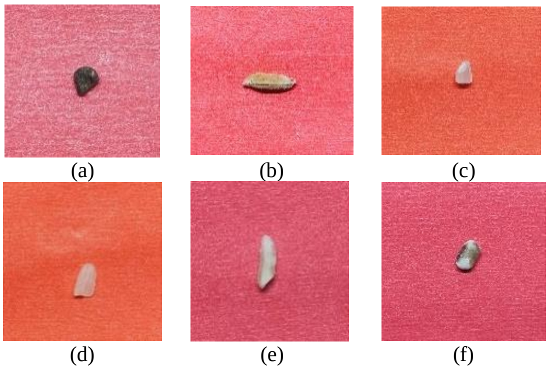


Fig. 1. Dataset of non-natural stone (a), and natural foreign objects of paddy (b), rice fragment (c), broken (d), yellow broken (e), and red-black (f)

II. MATERIAL AND METHOD

This research provides the best model reference for the classification of foreign objects based on CNN. The effectiveness of transfer learning and deep learning models was assessed by using pre-trained techniques, such as ResNet50, VGG16, and MobileNetV2. Additionally, comparisons were made with conventional machine learning, namely random forest (RF), support vector machine (SVM), DT, k-nearest neighbor (KNN), and logistic regression (LR). Cross-validation procedures with 5 K-Fold and 10 K-Fold values were used to obtain the most accurate machine learning values. The confusion matrix determined the accuracy, recall, precession, and F1-scores test. Fig. 1 shows foreign objects samples used, which included stones, rice, broken yellow, thorns, red-black, and fractures, and the contributions made were as follows:

1) A total of 1320 natural and non-natural foreign objects with six classes for the case of rice particulate.
2) Propose models in the detection of foreign objects for rice obtained from custom field datasets.
3) This research used two deep learning models with CNN and transfer learning to obtain the best accuracy. Previous testing was conducted with machine learning along with cross-validation.
4) To increase the accuracy value of the machine learning carried out using cross-validation with values of 5 to 10 K-Fold, which was not performed in the previous research.
5) Deep learning model built CNN with the transfer learning model pre-trained, including VGG16, ResNet50, and MobileNetV2.
6) Utilization of Faster RCNN for detecting foreign objects in rice based on the classification results of the best model constructed.

2.1. Data Collection

The agricultural commodities sourced from PT Hassana Boga Sejahtera were used in the acquisition of a dataset consisting of images depicting foreign things. The red background with the LEDs emitting light with a lux value ranging from 750 to 757 serves to illuminate objects of alien origin. The prototype platform is designed in a square shape, measuring 20 cm x 12 cm in diameter. The platform features a grain of rice with a height of 10 cm and is equipped with a Logitec C270 HD webcam camera.

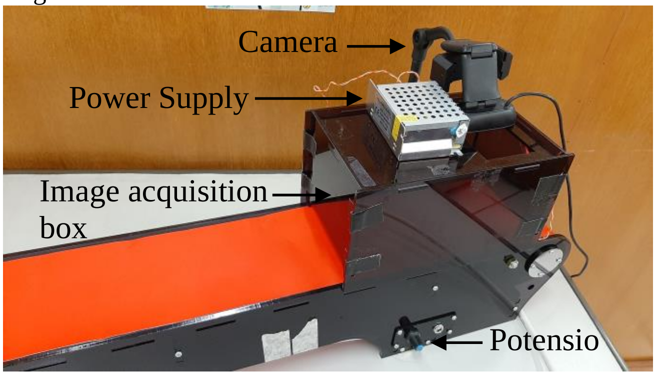


Fig 2. Image Acquisition Prototype with the box panel on the top.

LEDs are used to produce homogeneous day and nighttime lighting conditions on platforms. Fig. 2 is a prototype for the acquisition image. Variations in the angle of the image and illumination will impact the perception of depth when detecting objects in a picture. At the time of image capture, the camera is parallel to the object. The dataset was derived from the earlier investigation [18] by adding the class of breaking and turning. There were 1320 photos of foreign objects with a resolution of 72 dpi and a size of 2993 x 2993 pixels, with 220 photographs in each class. The dataset will be used in the training of 1056 and validation of 264 images.

The category of foreign objects is classified into two, namely non-natural foreign objects, which include stones, whereas natural foreign objects involve paddy, rice fragments, broken rice, yellow-broken, and red-black. The model is designed to identify foreign objects within six distinct categories, aiming to determine their status as foreign or not. Fig. 3 shows the steps of the image acquisition process for six types of foreign objects, namely stone as the non-natural class, while paddy, fragment, broken, yellow broken, and red black as the natural. The acquired results were collected into a single dataset of foreign objects, grains of rice, used for this research.

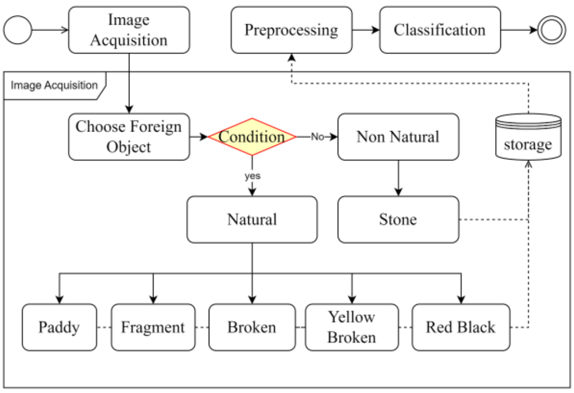


Fig. 3. Acquisition Model

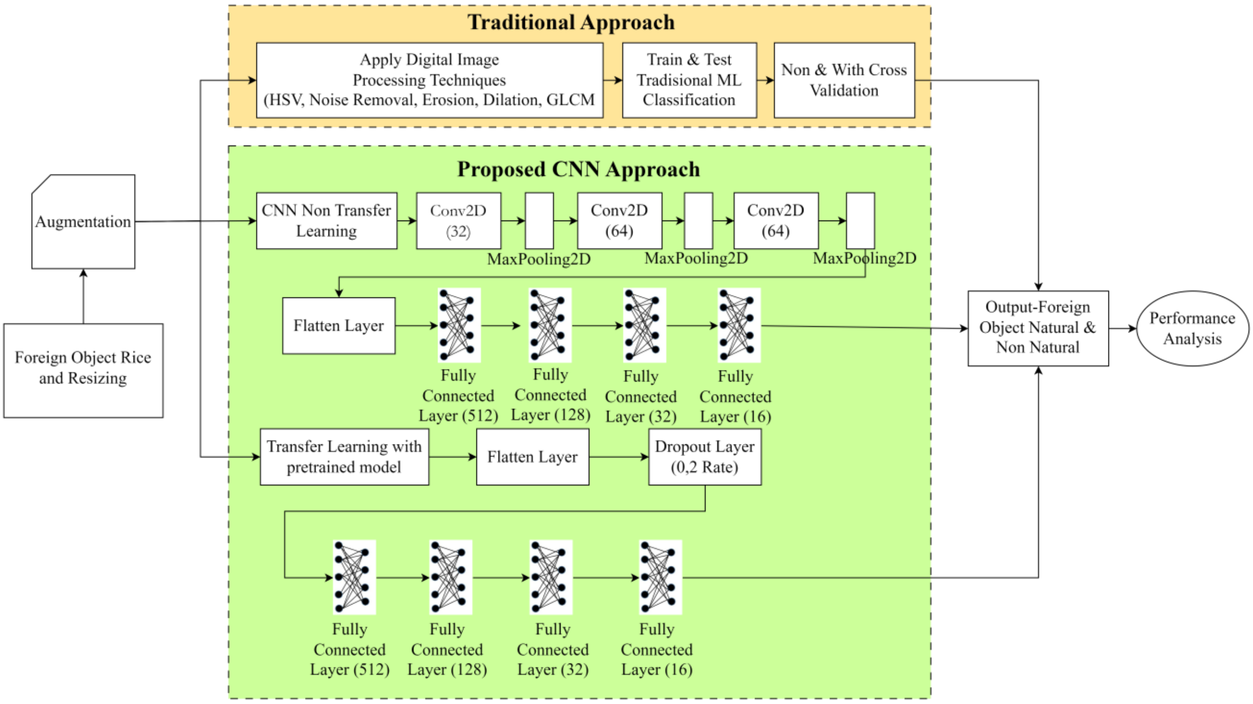


Fig. 4. Framework of research method

Data augmentation, which includes duplicating photographs with a number of alterations based on the original image samples, can be used to add more data. This strategy is effective for expanding the amount of visual data, which is necessary for deep learning and requires enormous datasets [19]. Three fundamental image changes were performed to each image to generate four additional training instances. The enhancement model consists of geometric changes (rotate, flip, and random shifting), contrast improvement, and noise removal. The dataset with and without augmentation has 1056 and 80 images, respectively, of extraterrestrial objects generated. Fig. 4 shows the research method underlying this work.

2.2. Machine Learning Model

Using machine learning with huge volumes of data frequently presents challenges due to susceptibility to noise, incomplete datasets, inconsistencies, and outliers [17]. Obtaining the weight value of a unique property of a vision for the classification process requires a preprocessing step in a digital image by separating training from testing data. Processing of digital images is performed post-acquisition and alters the image size. Efforts to improve image quality are made by reducing noise with the Gaussian filter model based on the upper and lower limits specified. The selection of the filter Gaussian because this model successfully reduces the noise, bringing the image edge closer to its original value [11][20]. In addition, morphology models are used to obtain objects on the image by completing the configuration of erosion and dilation. The erosion model is used to lower pixel values, while dilation is used to increase pixel points based on the object edge line. The segmentation process is the final step in selecting items from the preprocessing to differentiate them from the background, followed by grouping based on the attributes of each image.

In image segmentation, color is frequently used to differentiate between simple and complex objects. This feature is more reliable because it is independent of image size. The color segmentation model may be less effective due to its sensitivity to illumination and occlusion situations. In addition to RGB, it may use HSV, HIS, Lab, and other spaces to extract color information from the item being detected [13]. This research used the HSV model to generate values based on the color depth of the image, with hue and saturation and translated back into RGB. This differs from the RGB model [18], which uses only RGB as a color characteristic without HSV research [21].

This research used the GLCM model to extract texture data in order to reduce the limitations of color features. The GLCM algorithm determines the texture value by analyzing the pixels surrounding the desired region [22], producing the matrix for the entire image. The matrix is then partitioned into n × n non-overlapping blocks, with each extraction value characteristic including contrast, entropy, correlation, homogeneity, and energy [23].

After the preprocessing phase is complete, the image conversion procedure is performed in a 1-dimensional array for use in the dataset division phase, with 80% of the information used for training and 20% for testing. This research centered on the problem of multi-class categorization of six images of foreign objects. On the basis of their context, five classical categories are used to group images as non-natural stone foreign objects, natural stone, fragment rice, broken rice, yellow broken, and red black. The categorization models consist of SVM, DT, LR, RF, and KNN with K-Fold cross-validation.

SVM can perform classification in two different forms based on the optimal separation between two classes by finding a hyperlane (center line in two dimensions) with the closest points of each called support vector [24]. The hyperlane with the farthest distance to neighboring data indicates two classes having the best separation [25].

$$f(v) = \sum_{i=1}^{z} \alpha_i \omega_i . k(\upsilon, \upsilon_i) + b \qquad (1)$$

In the context of classification, the discriminant value $f(v)$ is the discriminant value for data v used for classification. Σ_i is the sum symbol from $i = 1$ to Z, representing the total number of training samples. α_i is a coefficient calculated based on certain constraints. w_i is the label (-1 or 1) of the i-th training sample. $k(v, vi)$ is a kernel function that measures how similar or different the data v and vi are in a higher feature space (possibly the space transformed by the kernel). The kernel function is the radial basis function (RBF), and the value of b is the bias or shift.

Decision tree (DT) is a widely used supervised machine learning model that is frequently used for classification and regression tasks. It expresses problems in tree structural forms, including a root node, decision nodes representing features, and leaf nodes representing outputs. This model divides the dataset based on specified criteria, selecting the best features by calculating information gain and entropy. Following the selection of rules and patterns presented by the decision nodes, DT can solve classification problems by generating final predictions from the input data. Logistic regression (LR) is a statistical model commonly used in the field of machine learning and statistics. It is a type of regression analysis that is particularly suited for Logistic regression, examining the association between a categorical outcome variable and one or more predictor factors. It accomplishes this by applying a logit transformation to the dependent variable. The logistic model estimates the logic of the dependent variable based on the independent [26].

Random forest (RF) is used as an effective data mining tool to address classification and regression problems. The use of voting in classification improves the accuracy of the classifier, while the formation of tree ensembles contributes to the improvement of accuracy. With each tree generated through a random vector, RF can provide predictions based on the selection of sample classes from the trees. The advantage of RF is its ability to avoid over-fitting when more trees are added to the ensemble [27] and to achieve higher accuracy. Maintaining low skewness and correlation in the variables is important by avoiding tree truncation and randomization of variables at each node.

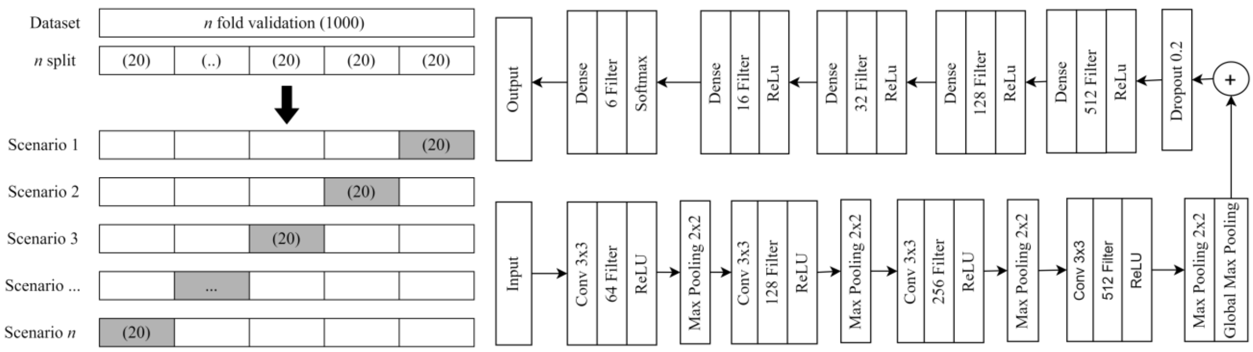
K-nearest neighbors (KNN) is a classification algorithm for determining similarities by relying on its nearest neighbors to provide information on whether it is a member of the set or not. Meanwhile, several research have been developed on how to increase the value of k by using a combination of models with fuzzy [28]. In the model known as a conventional holdout, testing with random data, there is a luck component depending on the train data employed here. In contrast, in the cross-validation approach, each time one sample is utilized as a validation/test sample while the remaining samples are used to train the model [29]. As depicted in Fig. 5, we utilized 5 and 10 K-fold validation to determine the significance of the derived testing n values fig. 5 (a).

2.3. Deep Learning Model

This research focused on the development of the CNN model and transfer learning, which is the process of applying previously acquired knowledge to solve a relative problem. On a massive image-net database, VGG16 functions as a pre-trained technique. Knowledge distillation and suggested models were used to acquire general characteristics, such as color and texture. The initial segment of the proposed model remains consistent in the learning process, as it leverages the pre-trained VGG16 to extract features from the available data. Simultaneously, another portion of the network is fine-tuned by modifying the hyperparameters to comprehend more particular and abstract data. Transfer learning reduces the need to train a model from scratch, reducing the amount of time required for computing from days to minutes.

This research made use of three different computational networks, including pre-trained Resnet50, VGG16, MobileNetV2, and transfer learning (TF) performed by adding four kernel-dense layers. The test results showed that the VGG16-TF model gave the best results, and the VGG16 Block used was depicted in Fig. 5 (b). All trained CNN models had a 4-layer fully connected after the dropout function on VGG16, Resnet50, and MobileNetV2 layers with 512, 128, 32, and 16 kernel sizes. A Global max pooling layer terminated the outputs of each pre-trained model node and a kernel layer based on the number of classes. The next step, which followed the acquisition phase, was composed of 80 % and 20 % data training and testing, respectively. CNN stage of developing models used pre-trained transfer and non-transfer learning on a "sequential" network.

The first sort of pre-training technique used was ResNet50, a residual network-based solution to the gradient loss problem on the inner nerve network. ResNet50 was composed of the remaining 50 layers of the network, which were organized into different groups of similar layers connected by identity blocks of variable sizes. The structure mostly comprised a sequence of 3×3 convolutions, 1×1 convolutions, and sigmoid activation functions [30][31]. The cross-layer connection function added the previous output to the result of the stacked layer, enabling deeper network training than would be achievable, and the model proposed resnet50 with transfer learning is shown in Table 1. In 2016, ResNet50 influenced the advancement of deep learning in academia and industry [32].



(a) (b)

Fig. 5. Diagram of k-fold forward cross-validation (a) & VGG16 block on this research (b)

TABLE I
CONFIGURATION OF PROPOSED RESNET50 AND TRANSFER LEARNING

Layer Name	Kernel	Size	Param
Resnet50 (functional)	2.048	224x224	23.564.800
Flatten	2.048	1	0
Dropout	2.048	0.2	0
Dense	512	ReLu	1.049.088
Fullyconnected_1	128	ReLu	65.664
Fullyconnected_2	32	ReLu	4.128
Fullyconnected_3	16	ReLu	528
Dense	6	Sofmax	102

The second pre-trained model used in this investigation was VGG16, which consisted of 16 convolutional layers with varying weights. The University of Oxford Visual Geometry Group (VGG) created the convolutional neural network model known as VGG16, which won the 2014 ILSVRC object identification competition [30]. It adheres to an architectural design that consists of 3x3 filters with the same padding layer and a maximum pool of 2x2 filters with a stride of 2. There are two completely connected layers at the conclusion, followed by a softmax output. This network has little over 138 million parameters, making it highly vast. This model achieves the highest accuracy of 90.1% on the Imagenet database. The setup of the VGG16 model proposed with transfer learning is shown in Table 2.

TABLE II
CONFIGURATION OF PROPOSED VGG16 AND TRANSFER LEARNING

Layer Name	Kernel	Size	Param
VGG16 (functional)	512	224x224	14.714.688
Flatten	512	1	0
Dropout	512	0.2	0
Dense	512	ReLu	262.656
Fullyconnected_1	128	ReLu	65.664
Fullyconnected_2	32	ReLu	4.128
Fullyconnected_3	16	ReLu	528
Dense	6	Sofmax	102

Pre-trained MobileNetV2 proposed by google research team in 2019 [33] has superior speeds and efficiency characteristics. MobileNetV2 used many convulsive layers (Conv) on the field 3x3 and Mobile Inverted Bottleneck Convolution (MBConv) as the primary network to increase training and accuracy compared to the first generation. Tan et al. employed MobileNetV2 to create accurate models because of its depth, width, and balanced resolution. In this research, 2,257,984 parameters were used for MobileNetV2, as shown in Table 3.

TABLE III
CONFIGURATION OF PROPOSED MOBILENETV2 AND TRANSFER LEARNING

Layer Name	Kernel	Size	Param
MobileNetV2 (functional)	1280	224x224	2.257.984
Flatten	1280	1	0
Dropout	1280	0.2	0
Dense	512	ReLu	655.872
Fullyconnected_1	128	ReLu	65.664
Fullyconnected_2	32	ReLu	4.128
Fullyconnected_3	16	ReLu	528
Dense	6	Sofmax	102

Following the pre-training stages of ResNet50, VGG16, and MobileNetV2, the weight parameters of each pre-trained model will be transferred to the fully connected layer. The process of flattening the layer includes converting the multidimensional output of the previous surface into a one-dimensional array. This flattened representation was subsequently used in classification. The input layer of a neural network is constructed using a one-dimensional array, where each element corresponds to the components of a neuron. The layer in question served as a connection between the convolutional and dense layers. The second surface, in a similar manner to the prior layer, extracts the most pertinent aspects of the image.

The procedure was continued with the dropout layer, and this regularization model drastically decreases overfitting and accelerates learning by transforming input data to 0 to ignore multiple nodes at the frequency level selected at every phase of training [34]. CNN model dropout layer has been regularized by setting the rate to 0.2. The detail of the hyperparameter is shown in Table 4.

TABLE IV
HYPERPARAMETERS USED

Hyper-parameter	Feature extraction value	Fine tunning value
Optimizer	SGD	SGD
Learning rate	0.001	0.0001
Batch Size	16	16
Epoch	50	50

CNN final classification is the Fully Connected Layer, often known as the dense. This layer is positioned at the bottom of the CNN model, with each neuron in it connected to others on the layer before and after using typical multilayer perceptron and neural networking feed-forward models [35]. Due to the six classes, the final layer of the dense web would comprise six nodes to make classification predictions, one for each potential outcome. Four layers of 512, 128, 32, and 16 neurons, respectively, were used. The "softmax" activation function was used to select the neuron output with the highest probability, and Equation 2 was selected for the activation of softmax.

$$Soft\max\ Activation, S\left(a_i\right) = \frac{\exp\left(a_i\right)}{\sum_{j=1}^{n} \exp\left(a_j\right)} \tag{2}$$

Optimizers, metrics, and losses are the three parameters that comprise the CNN model. The stochastic gradient descent (SGD) optimizer is used to regulate the learning rate. SGD was selected because, despite its simplicity, it provides superior performance to other optimizers, such as Adam, during deep learning training [23]. Equation 3 was used to calculate the weight renewal of the pre-trained layer.

$$w = w - learning\_rate * g \tag{3}$$

Where w represents the initial weight and g is the gradient. The learning rate can be understood as the leap speed in the weight update. While too-low values lengthen the training period, too-high values can lead to suboptimal training results. In addition, the SDG used momentum, which aided in accelerating the gradient vector in the correct direction by facilitating training convergence. The used learning rate was 0.001, and the momentum value was 0.91. Based on Equation 4, a weight renewal formula was derived as follows.

$$velocity = momentum*velocity - learning\_rate*g$$
$$w = w + velocity \tag{4}$$

The "accuracy" parameter was used as a model metric to assess training success and analyze losses using the "categorical cross-entropy" function due to scenarios involving multi-class classification issues. Reduced losses indicated improved performance compared to the model specifications presented in Table 5. CNN with no pre-trained model was also constructed.

TABLE V
CONFIGURATION OF PROPOSED CNN WITH NO TRANSFER LEARNING

Layer Name	Kernel	Size	Param
Conv2D	(3,3)		32
MaxPooling2D	(2,2)		-
Conv2D	(3,3)		64
MaxPooling2D	(2,2)		-
Conv2D	(3,3)		128
MaxPooling2D	(2,2)		-
Flatten	86528	1	0
Dropout	512	0.2	0
Dense	512	ReLu	262.656
Fullyconnected_1	128	ReLu	65.664
Fullyconnected_2	32	ReLu	4.128
Fullyconnected_3	16	ReLu	528
Dense	6	Sofmax	102

CNN was trained with datasets and hyperparameter optimization using three distinct models, including (i) without transfer learning, (ii) combining a pre-trained ResNet50, (iii) incorporating VGG16, (iv) adding MobileNetV2 with transfer learning. After 50 iterations of each training type, the model was assessed with the test dataset.

To prevent overfitting, a callback was imposed when training the model, including stopping the training process for certain conditions, as shown in Algorithm 1. In this case, when the model performance did not improve within 10 epochs after no significant improvement, a value of delta=0.001 was considered significant in the accuracy graph.

Algorithm 1 Algorithm for EarlyStopping Train Model

```
Require: ModelVGG16                      ▷ Deep learning model
Require: train_generator                 ▷ Training data generator
Require: val_generator                   ▷ Validation data generator
Ensure: history                          ▷ Training history
Import necessary libraries
Procedure EarlyStopping (monitor= 'val_accuracy',
   patience=10, min_delta=0.001, mode= 'max',
   restore_best_weights=True)
if min_delta not increases 0.001 and patience > 10 epoch then
        EearlyStopping(break) <= matrix(mode=max)
end if
return restore_best_weight
```

In addition to the callback function to stop training, the learning rate automatically adjusted where the ReduceLRonPlateau function reduced the learning rate when the model performance did not improve based on the accuracy matrix as in Algorithm 2. The learning rate reduction was 30%, with the smallest learning rate value limit of 0.001. However, the learning rate reduction was executed only after 8 epochs of waiting.

Algorithm 2 Algorithm for ReduceLROnPlateau Train Model

```
Require: ModelVGG16                      ▷ Deep learning model
Require: train_generator                 ▷ Training data generator
Require: val_generator                   ▷ Validation data generator
Ensure: history                          ▷ Training history
Import necessary libraries
Procedure ReduceLROnPlateau (monitor= 'val_accuracy',
   factor=0.3, patience=8,  min_lr=0.001, mode='auto',
   verbose=1)
if model != validation and patience > 8  then
        learning_rate = learning_rate - 30% and min_lr=0.001
print(learning_rate)
end if
Train the model:
history = NULL
for each epoch in range(1, 51) do
if Training data generator = train_generator and
        Number of epochs = 1 and
        Validation data generator = val_generator then
        Callbacks = [EarlyStopping, ReduceLROnPlateau]
        history.append(model)
end if
end for
Return history
```

2.4. Matrix Evaluations

The confusion matrix was required to verify that the constructed model can investigate categorization problems, including natural and non-natural foreign objects, with details of paddy, rice fragment, yellow broken, and red-black. Fig 6 shows the model evaluation using F1 score values, recall, precision, and accuracy.

		Predicted Class	
		1	0
Actual Class	1	True Positive	False Negative
	0	False Positive	True Negative

Fig. 6. Confusion matrix to classify foreign objects

Informational specifics:

TP: True Positive, predicted, and existing.
FP: False Positive; infraction is anticipated but not existing.
FN: False Negative, although it is anticipated that the violation will not occur.
TN: True Negative, correct test or prediction result indicating the absence of a violation or negative condition.

The calculation formulas for the indexes are as follows:

$$accuracy = \frac{TP + TN}{TP + FP + TN + FN} * 100\% \tag{5}$$

$$precision = \frac{TP}{TP + FP} * 100\% \quad (6)$$

$$recall = \frac{TP}{TP + FN} * 100\% \quad (7)$$

$$F1score = \frac{2 * precision * recall}{precision * recall} \quad (8)$$

The term "TP" refers to true positives, indicating photographs containing the characteristics of foreign rice were accurately categorized. The abbreviation "FN" denoted the scenario in which an image, initially classified as non-foreign rice, was ultimately determined to be foreign. The term "FP" refers to the image being incorrectly identified as a positive instance of foreign rice when, in reality, it does not belong to the category of foreign.

III. Result and Discussion

Based on comparative research with instances using relatively limited datasets, this research examined the influence of drawing inferences from a large number of datasets, including up to 1320 images of foreign objects classified into six. In the first stage of the typical machine learning model, the full image on the dataset was previously preprocessed using noise removal, the morphology of erosion, and dilatation in order to differentiate objects from the background, as depicted in Fig. 6. Images that have passed the preprocessing stage will be retained and used to classify machine learning as training and testing data. The 1320-image datasets will be used to train, test, and evaluate six types of machine learning classifications using the preprocessing dataset.

In prior research [18], the application of K-Fold cross-validation was hampered by the use of relatively modest datasets. Efforts were made to address these regulations through the augmentation of dataset quantity and the expansion of foreign object categories. This was achieved by using cross-validation with a 10 K-Fold value and integrating supplementary machine learning models, including RF, KNN, and LR.

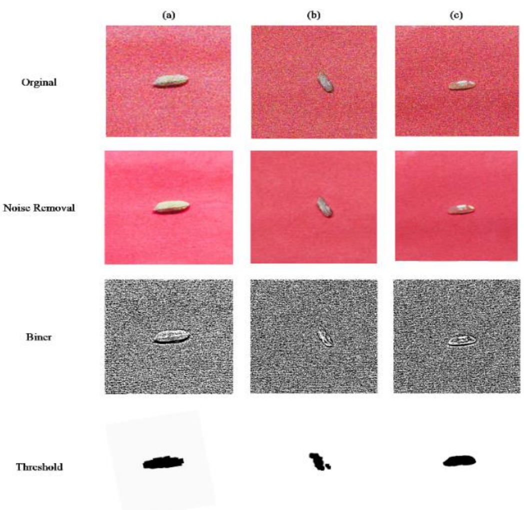


Fig. 6. Findings from Image Processing Stages

Table 6 shows the performance of five types of machine learning classification using specific model performance. Also, cross-validation improved the performance of the classification model. For instance, the accuracy of RF classification without cross-validation was 0.94 or 94%, whereas using the model resulted in a value of 0.95 or 95%. The value of DT increased dramatically as the K-Fold value augmented from 73% to 89% to 91% with 10 K-Fold validation. According to the findings, cross-validation on training data and testing on the accuracy, precision, recall, and F1 scores boosted the prediction performance of machine learning models.

TABLE VI
PERFORMANCE ANALYSIS RESULTS OBTAINED FROM TRADITIONAL MACHINE LEARNING MODEL WITH CROSS-VALIDATION

Evaluation	Machine Learning Model				
	RF	SVM	DT	KNN	LR
Non Cross-Validation					
Accuracy	**0.94**	0.82	0.73	0.67	0.88
Precision	0.95	0.83	0.74	0.67	0.88
Recall	0.94	0.82	0.73	0.67	0.88
F1 Score	0.95	0.83	0.73	0.67	0.88
5 K-Fold Cross-Validation					
Accuracy	**0.95**	0.83	0.89	0.79	0.93
Precision	0.95	0.84	0.90	0.79	0.93
Recall	0.95	0.83	0.89	0.78	0.93
F1 Score	0.95	0.83	0.90	0.79	0.93
10 K-Fold Cross-Validation					
Accuracy	**0.95**	0.83	0.91	0.79	0.94
Precision	0.95	0.85	0.91	0.79	0.94
Recall	0.95	0.83	0.91	0.79	0.94
F1 Score	0.95	0.84	0.91	0.79	0.94

As a comparison of the results of the machine learning model, it was proposed to develop CNN with a deep and transfer learning model. This was incorporated with slight changes in hyperparameter values to determine the best model for predicting the depth of foreign objects after each model was trained for 50 epochs with a callback optimization procedure. Table 7 indicates four strategies presented by CNN, including one non-transfer and three transfer learning models. According to the table, the CNN strategy has the highest accuracy among other models at 97 %, followed by ResNet50 at 93%, MobileNetV2 at 90%. CNN had a minimum accuracy of 50%, placing them in the low group.

TABLE VII
PERFORMANCE ANALYSIS RESULTS USING THE PROPOSED CNN MACHINE LEARNING MODEL

CNN Models	Acc	Pre	Rec	F1
CNN	0.47	0.39	0.47	0.42
ResNet50 + TF	0.93	0.95	0.93	0.93
VGG16 + TF	**0.97**	**0.97**	**0.97**	**0.97**
MobileNetV2 + TF	0,90	0,94	0,90	0,90

Note : acc(accuracy), pre (precision), rec(recall), f1(f1-score)

Figure 7 provides a comprehensive visualization of the accuracy per training epoch, wherein the number of epochs is capped at a maximum of 50, as stipulated by the callback function. Subfigures (a), (b), (c), and (d) respectively depict CNN accuracy without transfer learning, utilizing VGG16, incorporating ResNet50, and employing MobileNetV2. This concise representation encapsulates diverse accuracy trends across various CNN configurations within the specified epoch constraints.

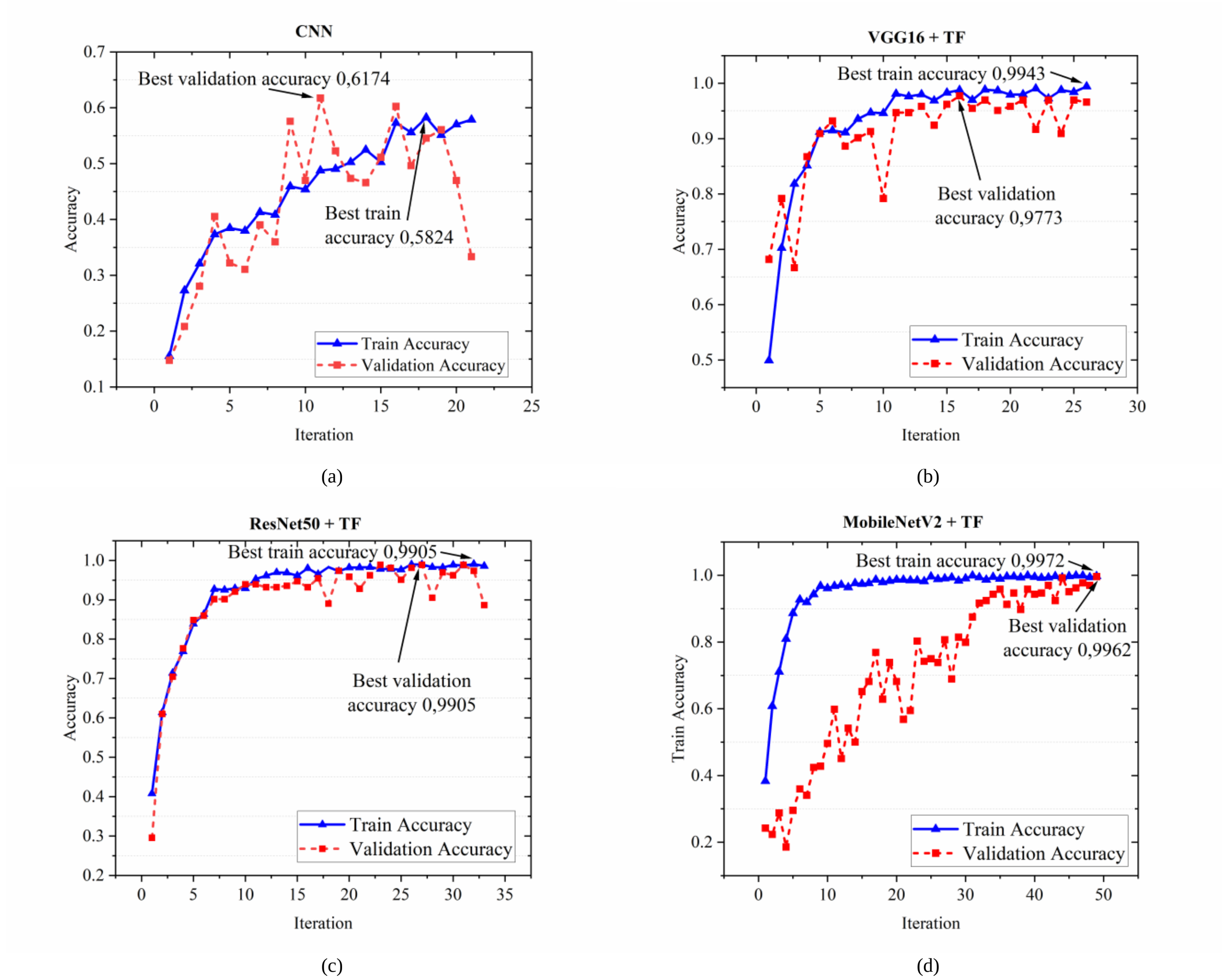


Fig. 7. Performance accuracy of CNN (a), VGG16+TF (b), ResNet50+TF models (c), and MobileNetV2+TF (d)

Fig. 7 indicated that all accuracy models were fluctuating, and a number of techniques were underfitting, including the CNN without transfer learning. Also, VGG16 and ResNet50 appeared to be gradually increasing the ideal model. Specifically for VGG16, accuracy were consistent after 35 epochs, whereas ResNet50 swings after 20 epochs, showing that the model had converged.

CNN and tranfer learning pre-trained comparison analysis. The foreign object rice is directly fed into the CNN layer and Pre-Trained Layer for comparison. The accuracy of the CNN model and the VGG-TF model can be seen in Figure 8(a) and (b) accordingly. (a) The foreign object detection accuracy of the CNN layer is 0.47, whereas the VGG+TF layer achieves an accuracy of 0.97. The comparison between the ResNet50+TF model and the MobileNetV2+TF model was conducted to assess their accuracy. The results are presented in Figure 8(c) and (d) for each model, respectively. Model CNN and model CWT-CNN were selected for comparison. The accuracy of the ResNet50+TF model is 0.93, as shown in Table 4. Similarly, the accuracy of the MobileNetV2+TF model is 0.90. In summary, the VGG+TF model should be selected.

Based on the findings presented in Figure 8, it is evident that Random Forest (RF) with 10 K-Fold cross-validation emerged as the most accurate machine learning model, achieving an accuracy rate of 94.88%. In comparison, the accuracy obtained with 5 K-Fold cross-validation was slightly lower at 94.60%, resulting in a marginal difference of 0.30%. Furthermore, when cross-validation was not applied, the disparity increased slightly to 0.59%. Meanwhile, Decision Tree (DT) with cross-validation exhibited a significant difference of 24.86%, marking the most substantial gap when compared to the accuracy obtained with 10 K-Fold validation.

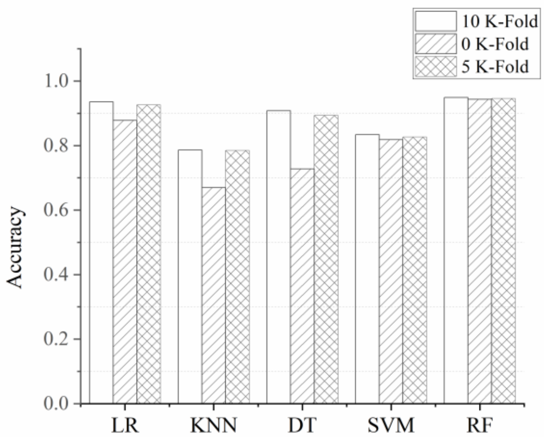


Fig. 8. Comparative accuracy analysis models machine learning

Fig. 9 showed that VGG16, when used with transfer learning, exhibited superior performance compared to others, achieving an accuracy rate of 97%. In close proximity, ResNet50 achieved a commendable accuracy rate of 87%. The disparity between VGG16 and CNN was found to be 50%, while the dissimilarity between ResNet50 and CNN was 40%. These findings underscore the efficacy of transfer learning, particularly with VGG16, in enhancing the accuracy of the model compared to traditional convolutional neural networks.

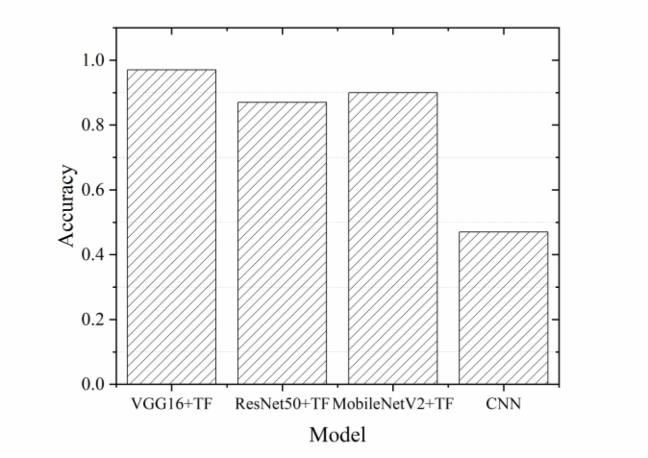


Fig. 9. Comparative accuracy deep learning model

Fig. 10 depicts the test results of a model designed to recognize foreign objects within a collection of images of mixed rice. The test findings indicate that the model can identify red and black (rb) rice objects, damaged yellow (yb) rice objects, and fragments (fr).

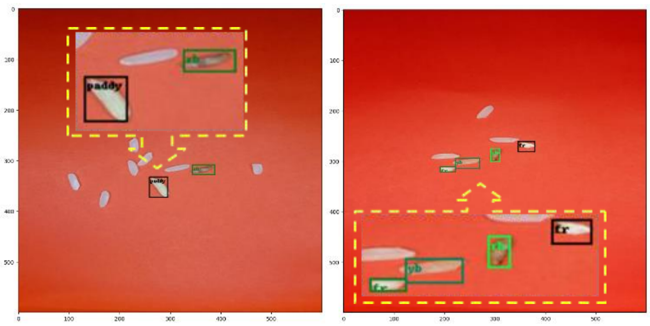


Fig 10. Deep learning model for detecting foreign objects

Table 8 compares accuracy results from several related research using both machine and deep learning models. The proposed model showed a 1-2% improvement compared to the comparative model used by other methods. With the inclusion of the VGG16+TF architectural model for further development, the proposed method is significant for future research and may be compared to other object detection algorithms.

TABLE VIII
SOME RESEARCH RESULTS AS A COMPARISON

Classifier	Methodology/Features	Accuracy
Fuzzy Logic [36]	The morphology features of rice grains, including parameters like area, perimeter, equivalent diameter, and roundness, used a fuzzy logic algorithm.	83,33 %
VGG16 [37]	Adaptive Mean Filter (AMF), VGG16, and cross-validation to avoid bias.	88,62 %
CNN ShuffleNet +SVM [38]	Use of Adaptive Anisotropic Diffusion Filter and Adaptive Mean Adjustment for preprocessing, segmentation with Fuzzy C-Means and Adaptive Otsu Thresholding, GLCM, and feature optimization with PCA.	89,37 %
VGG19 [39]	Classified using SVM and CNN architectures such as AlexNet, VGG16, and VGG19.	94,74 %
CNN U-NET [13]	CNN U-NET with data augmentation techniques, such as color jitter, are used to increase the model robustness to lighting changes.	95 %
ResNet50+ TF [40]	Convolutional Neural Networks (CNN) with transfer learning, such as DenseNet201, InceptionResnetV2, InceptionV3, ResNet50, ResNet152V2, and Xception.	95,24 %
SVM [18]	HSV color and GLCM texture were extracted and trained in the classifiers.	96,83 %
The proposed model (CNN VGG16+TF)	RGB images to classification with CNN+VGG+Transfer Learning.	97 %

IV. Conclusion

The use of DCNN-based models, with transfer learning and fine-tuning techniques, was suggested as a means to identify foreign objects present in rice. The proposed design used transfer learning by leveraging pre-trained techniques, incorporating many convolutional layers with carefully tuned hyperparameters. These convolutional layers were used to extract relevant visual features, which were fed into the fully connected convolutional network for the purpose of classifying images into six distinct categories. As an integral component of the model validation and evaluation procedure, the multi-class classification solution was incorporated with a conventional model that integrates digital image processing and a machine learning classification model. Accuracy, precision, recall, and F1-scores were often used metrics for evaluating the performance of a model. The results of the comparative analysis showed that the proposed strategy exhibited a high level of accuracy compared to the traditional model. This strategy incorporated transfer learning, thereby surpassing alternative strategies in terms of precision, even during the cross-validation phase of the traditional model.

Consequently, the proposed DCNN-based computer model can aid farmers and the agricultural industry significantly in the field of quality control using deep learning techniques. This research could be expanded by including additional samples, distinguishing more diversified foreign objects, and applying comparative object detection strategies. Future research should concentrate on using VGG16 and transfer learning for object detection with YOLO, SSD, or Faster RCNN.

References

[1] Y. Yan, “Food safety and social risk in contemporary China,” J. Asian Stud., vol. 71, no. 3, 2012, doi: 10.1017/S0021911812000678.

[2] FAO and WHO, “Food and Agriculture Organization Assuring Food Safety and Quality: Guidelines for strengthening national food control systems,” *Food Nutr. Pap.*, vol. 76, 2003.

[3] N. N. Suh, G. F. Njimanted, and N. Thalut, “Effect of farmers’ management practices on safety and quality standards of cocoa production: A structural equation modeling approach.,” *Cogent Food Agric.*, vol. 6, no. 1, 2020, doi: 10.1080/23311932.2020.1844848.

[4] A. Bhargava and A. Bansal, “Fruits and vegetables quality evaluation using computer vision: A review,” *Journal of King Saud University - Computer and Information Sciences*, vol. 33, no. 3. 2021. doi: 10.1016/j.jksuci.2018.06.002.

[5] L. Zhu, P. Spachos, E. Pensini, and K. N. Plataniotis, “Deep learning and machine vision for food processing: A survey,” *Curr. Res. Food Sci.*, vol. 4, pp. 233–249, 2021, doi: https://doi.org/10.1016/j.crfs.2021.03.009.

[6] S. Mittal, M. K. Dutta, and A. Issac, "Non-destructive image processing based system for assessment of rice quality and defects for classification according to inferred commercial value," *Measurement*, vol. 148, p. 106969, Dec. 2019, doi: https://doi.org/10.1016/j.measurement.2019.106969.

[7] K. Adi, S. Pujiyanto, O. Dwi Nurhayati, and A. Pamungkas, "Beef Quality Identification Using Thresholding Method and Decision Tree Classification Based on Android Smartphone," *J. Food Qual.*, vol. 2017, 2017, doi: 10.1155/2017/1674718.

[8] J. M. Soon, A. K. M. Brazier, and C. A. Wallace, "Determining common contributory factors in food safety incidents – A review of global outbreaks and recalls 2008–2018," *Trends Food Sci. Technol.*, vol. 97, pp. 76–87, Mar. 2020, doi: https://doi.org/10.1016/j.tifs.2019.12.030.

[9] M. T. Mohd Khairi, S. Ibrahim, M. A. Md Yunus, and M. Faramarzi, "Noninvasive techniques for detection of foreign bodies in food: A review," *Journal of Food Process Engineering*, vol. 41, no. 6. 2018. doi: 10.1111/jfpe.12808.

[10] D. Rong, L. Xie, and Y. Ying, "Computer vision detection of foreign objects in walnuts using deep learning," *Comput. Electron. Agric.*, vol. 162, 2019, doi: 10.1016/j.compag.2019.05.019.

[11] V. Bhole and A. Kumar, "Mango Quality Grading using Deep Learning Technique: Perspectives from Agriculture and Food Industry," in SIGITE 2020 - Proceedings of the 21st Annual Conference on Information Technology Education, 2020. doi: 10.1145/3368308.3415370.

[12] A. O. Adedoja, P. A. Owolawi, T. Mapayi, and C. Tu, "Intelligent Mobile Plant Disease Diagnostic System Using NASNet-Mobile Deep Learning," *IAENG International Journal of Computer Science*, vol. 49, no. 1, pp.216-231, 2022.

[13] G. J. Son, D. H. Kwak, M. K. Park, Y. D. Kim, and H. C. Jung, "U-Net-based foreign object detection method using effective image acquisition system: A case of almond and green onion flake food process," *Sustain.*, vol. 13, no. 24, 2021, doi: 10.3390/su132413834.

[14] K. Adi, C. E. Widodo, A. P. Widodo, and U. S. Margiati, "Detection of foreign object debris (FOD) using convolutional neural network (CNN)," *J. Theor. Appl. Inf. Technol.*, vol. 100, no. 1, 2022.

[15] K. Zhang, W. Wang, Z. Lv, Y. Fan, and Y. Song, "Computer vision detection of foreign objects in coal processing using attention CNN," *Eng. Appl. Artif. Intell.*, vol. 102, 2021, doi: 10.1016/j.engappai.2021.104242.

[16] N. Mamat, M. F. Othman, R. Abdulghafor, A. A. Alwan, and Y. Gulzar, "Enhancing Image Annotation Technique of Fruit Classification Using a Deep Learning Approach," *Sustain.*, vol. 15, no. 2, 2023, doi: 10.3390/su15020901.

[17] B. SravyaPranati, D. Suma, C. ManjuLatha, and S. Putheti, "Large-Scale Video Classification with Convolutional Neural Networks," in *Smart Innovation, Systems and Technologies*, 2021, vol. 196. doi: 10.1007/978-981-15-7062-9_69.

[18] A. Setiawan, K. Adi, and C. E. Widodo, "Rice Foreign Object Classification Based on Integrated Color and Textural Feature Using Machine Learning," *Math. Model. Eng. Probl.*, vol. 10, no. 2, pp. 572–580, Apr. 2023, doi: 10.18280/mmep.100226.

[19] N. E. Khalifa, M. Loey, and S. Mirjalili, "A comprehensive survey of recent trends in deep learning for digital images augmentation," *Artif. Intell. Rev*., vol. 55, no. 3, 2022, doi: 10.1007/s10462-021-10066-4.

[20] T. Yun *et al.*, "Individual tree crown segmentation from airborne LiDAR data using a novel Gaussian filter and energy function minimization-based approach," *Remote Sens. Environ.*, vol. 256, 2021, doi: 10.1016/j.rse.2021.112307.

[21] A. Ibrahim, N. El-Bialee, M. Saad, and E. Romano, "Non-Destructive Quality Inspection of Potato Tubers Using Automated Vision System," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 10, no. 6, p. 2419, Dec. 2020, doi: 10.18517/ijaseit.10.6.13079.

[22] P. K. Sethy, N. K. Barpanda, A. K. Rath, and S. K. Behera, "Deep feature based rice leaf disease identification using support vector machine," *Comput. Electron. Agric*., vol. 175, 2020, doi: 10.1016/j.compag.2020.105527.

[23] K. K. Leong and L. L. Tze, "Plant Leaf Diseases Identification using Convolutional Neural Network with Treatment Handling System," in 2020 IEEE International Conference on Automatic Control and Intelligent Systems, I2CACIS 2020 - Proceedings, 2020. doi: 10.1109/I2CACIS49202.2020.9140103.

[24] C. Jing, Z. Haopeng, C. K. On, E. G. Moung, and P. Anthony, "Face Recognition Based on Deep Convolutional Support Vector Machine with Bottleneck Attention," *IAENG International Journal Computer Science*, vol. 49, no. 4, pp. 1284-1296, 2022.

[25] U. Iruansi, J. R. Tapamo, and I. E. Davidson, "Classification of power-line insulator condition using local binary patterns with support vector machines," *IAENG International Journal Computer Science*, vol. 46, no. 2, pp. 300-310, 2019.

[26] Md. Asfi-Ar-Raihan Asif, Mirza Muntasir Nishat, Fahim Faisal, Rezuanur Rahman Dip, Mahmudul Hasan Udoy, Md. Fahim Shikder, and Ragib Ahsan, "Performance Evaluation and Comparative Analysis of Different Machine Learning Algorithms in Predicting Cardiovascular Disease," *Engineering Letters*, vol. 29, no.2, pp731-741, 2021.

[27] S. M. Mostafa, S. A. Salem, and S. M. Habashyis, "Predictive Model for Accident Severity," *IAENG International Journal Computer Science*, vol. 49, no. 1, pp. 110-124, 2022.

[28] A. Nugroho, H. L. H. S. Warnars, F. L. Gaol, and T. Matsuo, "Trend of Stunting Weight for Infants and Toddlers Using Decision Tree," *IAENG International Journal of Applied Mathematics*, vol. 52, no. 1, pp. 144-148, 2022.

[29] Z. Xiong, Y. Cui, Z. Liu, Y. Zhao, M. Hu, and J. Hu, "Evaluating explorative prediction power of machine learning algorithms for materials discovery using k-fold forward cross-validation," *Comput. Mater. Sci.*, vol. 171, 2020, doi: 10.1016/j.commatsci.2019.109203.

[30] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings, 2015.

[31] O. K. Oyedotun, A. E. R. Shabayek, D. Aouada, and B. Ottersten, "Going Deeper with Neural Networks Without Skip Connections," in Proceedings - International Conference on Image Processing, ICIP, 2020, vol. 2020-October. doi: 10.1109/ICIP40778.2020.9191356.

[32] L. Alzubaidi *et al.*, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *J. Big Data*, vol. 8, no. 1, 2021, doi: 10.1186/s40537-021-00444-8.

[33] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2018. doi: 10.1109/CVPR.2018.00474.

[34] G. S. Nandini, A. P. S. Kumar, and C. K, "Dropout technique for image classification based on extreme learning machine," *Glob. Transitions Proc*., vol. 2, no. 1, pp. 111–116, Jun. 2021, doi: 10.1016/j.gltp.2021.01.015.

[35] G. Moreira, S. A. Magalhães, T. Pinho, F. N. dos Santos, and M. Cunha, "Benchmark of Deep Learning and a Proposed HSV Colour Space Models for the Detection and Classification of Greenhouse Tomato," *Agronomy*, vol. 12, no. 2, p. 356, Jan. 2022, doi: https://doi.org/10.3390/agronomy12020356.

[36] C. L. Macalalad, E. R. Arboleda, A. A. Andilab, and R. M. Dellosa, "Morphological based grain comparison of three rice grain variety," *Int. J. Sci. Technol. Res*., vol. 8, no. 8, 2019.

[37] Daniel S. Wangsaputra, Choirul Anam, Kusworo Adi, and Ariij Naufal, "Impact of Adaptive Mean Filter as the Preprocessing Stage of Histopathological Image Classification of Breast Tumor Using Transfer Learning VGG16 for Various Magnifications," *Int. J. Sci. Res. Sci. Technol.*, pp. 274–280, Mar. 2023, doi: 10.32628/IJSRST52310239.

[38] M. Yogeshwari and G. Thailambal, "Automatic feature extraction and detection of plant leaf disease using GLCM features and convolutional neural networks," *Mater. Today Proc.*, 2021, doi: 10.1016/j.matpr.2021.03.700.

[39] S. Agustin, H. Tjandrasa, and R. V. H. Ginardi, "Deep Learning-based Method for Multi-Class Classification of Oil Palm Planted Area on Plant Ages Using Ikonos Panchromatic Imagery," *Int. J. Adv. Sci. Eng. Inf. Technol*., vol. 10, no. 6, 2020, doi: 10.18517/ijaseit.10.6.12030.

[40] A. Majumder, A. Rajbongshi, M. M. Rahman, and A. A. Biswas, "Local Freshwater Fish Recognition Using Different CNN Architectures with Transfer Learning," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 11, no. 3, 2021, doi: 10.18517/ijaseit.11.3.14134.

Aji Setiawan is a Doctoral student in the information systems department at Diponegoro University, Indonesia. His research interests include Computer Vision, Agriculture Technology, and Deep Learning.

Kusworo Adi is a full Professor of Computer Science at Diponegoro University, Indonesia. His main research interests include Pattern Recognition, Image Processing, Data Mining, and Artificial Intelligence.

Catur Edi Widodo is a full Lecturer of the Department of Physics, Faculty of Science and Mathematics, Diponegoro University, Indonesia. His main research interests include Pattern Recognition, Image Processing, Data Mining, and Artificial Intelligence.