Feature Matching and Determining Similarities Algorithms: A Comparison

Seh Hong Low
Faculty of Computing and Informatics
Multimedia University
Cyberjaya, Malaysia
1161104400@student.mmu.edu.my

Weng Hoong Seng
Faculty of Computing and Informatics
Multimedia University
Cyberjaya, Malaysia
1161103872@student.mmu.edu.my

Ee Kein Ivan Boo
Faculty of Computing and Informatics
Multimedia University
Cyberjaya, Malaysia
1161104032@student.mmu.edu.my

***Abstract*—Image classification is no longer a stranger in every sort of domain, especially in industries. This type of classification problem exists in various fields such as precision agriculture, control and sorting of automated crop tracking and more. The application of image classification has been used widely as it is capable of identifying products without any direct contact which can save human resources and time, which will increase the productivity of a business and lower its cost. As there are plenty of techniques in image classification, three standard methods such as SIFT, SURF and CNN are being compared in this study to understand which of the methods is more effective in completing such tasks. Background study has been done to understand the applications of image classification in different domains and the techniques used, background regarding the standard methods such as SIFT, SURF and CNN. To test out the techniques, we have introduced a fruit classification problem using the "Fruit-360" dataset and construct the models using SIFT and SURF as feature extraction functions and SVM as the classifier. CNN is used to classify the fruits without any external feature extraction methods. Then, several evaluation methods are proposed to evaluate the efficiency of the models in classifying the classes of fruits. Based on the findings, CNN exhibits the most accurate classification of the fruits compared to the other two models, which are using SIFT and SURF as feature extraction methods. This paper ends with a conclusion of the whole study and the direction of future studies that should be focused on.**

***Keywords*—Fruit Classification, Scale-Invariant Feature Transform, Speeded-Up Robust Features, Support Vector Machine, Convolutional Neural Network**

I. Introduction

In this modern period, image processing has been commonly used in all kinds of industries and various fields, such as precision agriculture, control and sorting of automated crop tracking and many more [1]. Reasons that are invoked behind these wide ranges of applications of image processing in these industries are because of its capability of identifying products without direct contact physically. With regards to image processing capabilities, identification of products based on size, colour, texture, etc., can be easily achieved. Since human error is known to be relatively large in the manual sorting process, automated sorting such as image processing in the food industry is one of the most challenging problems, imprinting a long-term impact on the process industry [2]. In this area, challenges have been addressed where various types of algorithms have been proposed for the classification of an object in an image. Still, the efficiency and accuracy of the algorithms vary. In specific industries, precision and accuracy, as well as the efficiency of classifying items in a picture, is very significant if it can affect the company as the days pass by where the lower the accuracy and efficiency would affect the cost of production. This project holds the objective of comparing variations of well-known algorithms such as SIFT (Scale-Invariant Feature Transform), SURF (Speeded-Up Robust Features), and convolutional neural network in the contrast of their accuracy and efficiency. A small sample of train and training sets will be selected from the dataset of images containing fruits and vegetables which is titled "Fruits 360 dataset" from Kaggle having different fruits and vegetables with other varieties for the evaluations of the algorithms [3].

II. Background

In this section, several related works will be reviewed, and an extensive background study will be performed on the literature materials that are related to the problem of the study.

A. Techniques in Image Classification

In the image classification field, there are a wide variety of techniques implemented, from image feature extraction to image classification. A literature study has been carried out to obtain domain knowledge on image classification implemented by current and recent research works conducted, subjectively on the models and techniques used to classify and identify images. The process of image classification includes pre-processing, object detection, image sensors, segmentation of objects, feature extraction and finally object classification. Therefore, although the classification of objects seems to be an easy task, for machines, it is a challenging task [4]. Generally, the application for image classification consists of a database containing predefined patterns that fit the object to be categorised into the appropriate category.

According to Table I image classification approach is essential for various domains such as medical, business, urban planning, transportation, and environment. Image classification can be categorised into two main categories, namely supervised classification and unsupervised classification. Figure 1

TABLE I
AUTHORS AND DOMAINS RELATED TO IMAGE CLASSIFICATION

Author	Technique	Domain				
		Medical	Business	Urban Planing	Transportation	Environment
[6]	Deep convolutional Neural Network	✓				
[7]	Hyperparameter Tuning Deep Learning	✓				
[8]	Deep Neural Network	✓				
[9]	Deep Wavelet Autoencoder-Based Deep Neural Network	✓				
[10]	Deep Convolutional Neural Network	✓				
[11]	Convolutional Neural Network	✓				
[12]	Convolutional Neural Networks		✓			
[13]	Neural Networks		✓			
[14]	Lacunarity Approaches			✓		
[15]	Gray Level Co-occurrence Matrix (GLCM)			✓		
[16]	Fatigue Classification Algorithm (KNN algorithm)				✓	
[17]	Superframe Segmentation				✓	
[18]	Learning Vector Quantization Method				✓	
[19]	Support Vector Machine and Back Propagation Neural Network					✓
[20]	Support Vector Machine (SVM), Artificial Neural Network (ANN), and Random Forest (RF)					✓

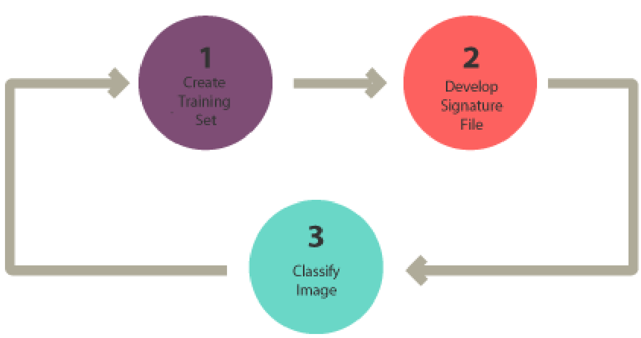

Fig. 1. The Flowchart of Supervised Classification [41]

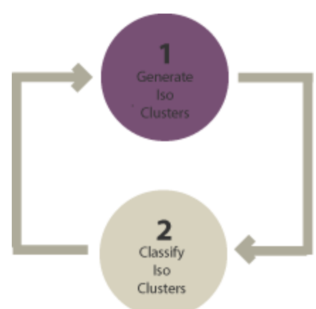

Fig. 2. The Flowchart of Unsupervised Classification [41]

and Figure 2 shows the steps of supervised classification and unsupervised classification, respectively. In supervised classification, some pixels are known to be clustered and give the classes a label; this process is known as training. After the process of training, classify other images by using the classifier trained pixels. In contract, the unsupervised classification is used when there are no trained pixels, which are dividing large numbers of pixels into several classes based on natural image groupings [5].

In image classification, the state-of-the-art techniques that are used in various domains are usually deep learning techniques such as neural networks. Image classification techniques are considered as necessary in the domain of healthcare and medic. Convolutional neural networks are used by Yada in [6] on chest X-rays dataset to classify pneumonia and it performs better than the ORB and SVM classifier. Besides, Guan has also implemented convolutional neural networks in classifying multi-label X-rays images in [11]. As seen in the study, a huge dataset is used to fit into the neural network as implementing a small dataset without parameter tuning will likely cause overfitting in the model, as mentioned by Yadav in [6]. Deep neural networks (DNN), till date, have demonstrated excellent performance in classification and segmentation tasks, and it has been used in brain MRI image classification for cancer detection [9]. Convolutional neural networks have also been used in clinical melanoma image classification tasks which have been done by Brinker in [10]. Besides in the medical field, neural networks have also been mainly applied in business such as fashion image classification [12] and retails such as solve this complex problem of planogram compliance for the retail use case to observed retail market certain policy for arranging items (planogram) so that sale of those items can be improved [13].

1) Support Vector Machine (SVM): Support Vector Machine (SVM) is a known classification method. SVM is a modern learning method introduced by [42] as a very efficient method for identifying common and essential patterns of intent. Support Vector Machine is referred to as a hyperplane classifier to decide an optimal line for the training set of the two classes to be separated [43]. The objective of the SVM Classification is to provide a model based on training data that will allow it to predict class labels of test data that go through dot-products [44]. SVM classification will have many attributes; however, there are just a few cases because of the training process. This has become one of the advantages of SVM classification. The figure 3 shows the sample of SVM classification.

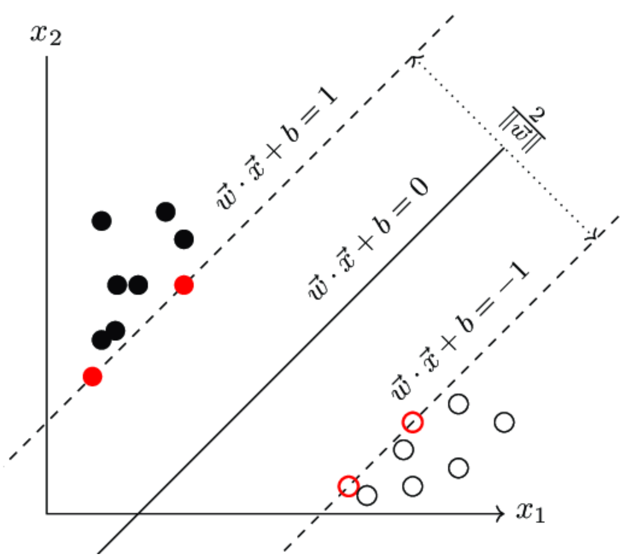

Fig. 3. SVM Classification [41]

In the study conducted by [45], they used the combination of Scale-invariant feature transform (SIFT) and SVM classification to do image classification. In their studies, they classifier 3 type of vehicles such as cars, trains, and aeroplanes. In this paper, the SIFT algorithm is used to extract feature points, and all the extracted feature points are clustered by K-means clustering. Finally, a multi-class classifier is equipped to identify images using SVM. The combination of SIFT and SVM classification obtained 93.3% correct rate. Besides, in the study conducted by [44], they classify ten persons' signatures in their paper. Each person includes 24 signatures consisting of 160 training signatures and 80 test signatures, and 240 signatures with a total signature picture. They used SIFT and Speeded-Up Robust Features (SURF), each of the feature extraction methods combined with SVM classification. In their conclusion, the combination of SIFT and SVM classification performed better than the combination of SURF and SVM classification; the accuracy was 98.75% and 96.25%, respectively.

2) Convolutional Neural Network (CNN): In the computer vision domain, convolutional neural networks (CNN), have gained great attention and become popular for their high accuracy and good performance in solving problems such as image classification. CNN is a class of deep feed-forward artificial neural networks, composed of multilayers including convolution layer, pooling layer, and full connection layer [49]–[51]. Compared to multilayer perceptron network (MLP), CNN utilizes convolutional layers and pooling, as well as options to deal with non-linearities [51]. In the convolution layer, there exists a filter of $x * y * c$, where x denotes the width, y denotes the height and c denotes the channel of image. The feature maps are then generated by computing the dot product of the region of input with the weight learning parameter. To further reduce the redundant features, a pooling layer reduces the size of images generated by the convolution layer. In general, the pooling layer performs down-sampling by computing the local average and maximum value. The convolutional layer and pooling layer can be added multiple times with different settings. Lastly, a full connection layer is adopted to perform classification. The neurons within this layer take the generated feature maps from preceding layers as inputs [50], [51].

B. Feature Extraction Algorithm

The extraction of features begins from an initial collection of calculated data. It produces derived features intended to be insightful and non-redundant, enabling the subsequent steps of learning and generalisation and, in some cases, leading to better human interpretations. Extraction of features is linked to reducing dimensionality [34].

1) Scale-invariant feature transform (SIFT): Proposed by David Lowe in 1999, is an algorithm used in the detection of features in an image. SIFT works by extracting the distinct points or unique characteristics of an image, such as the scale, location, position and so on [31]. These features can be used to identify an image [33], and most importantly, to perform comparison and matching between images [32] as well as identifying the target objects [31]. However, SIFT is computationally intensive and requires an extension in arithmetic and memory access [30]. SIFT is not sensitive to image size and rotation [31].

To explain the main concept of the SIFT algorithm, the steps can be separated into two main stages: keypoint detection and feature descriptor construction. In general, SIFT uses Difference-Of-Gaussian (DOG) scale to detect the key points and the extreme spatial points to remove the lousy feature points. Secondly, to ensure that the descriptor can be rotated

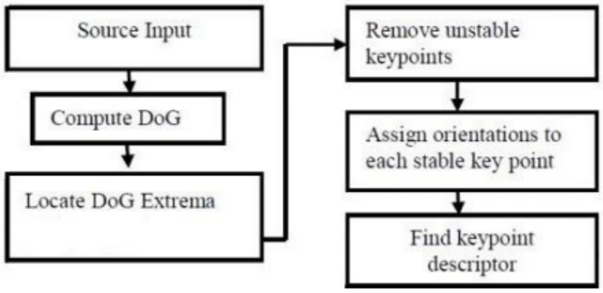

Fig. 4. Flowchart of SIFT Algorithm [40]

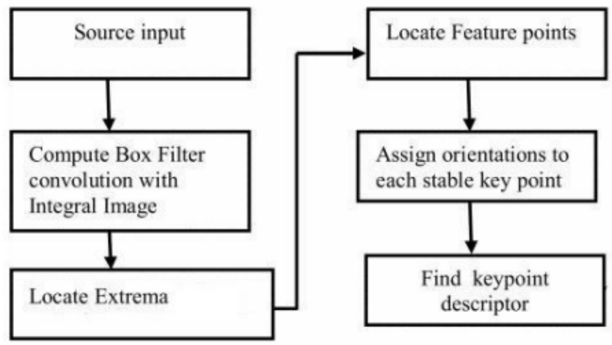

Fig. 5. Flowchart of SURF Algorithm [40]

in a non-deform manner, a directional gradient can be given to each identified vital point, by computing a gradient magnitude and orientation weighted by Gaussian Window around the location of the detected key point. A unique vector will be generated for each key point, computed using block gradient histogram of the region around each identified vital point. The vector describes the individual information of the image in this region [31]–[33].

2) Speeded-Up Robust Features (SURF): This is another image feature extraction algorithm proposed by Herbert Bay et al. in 2006 [39], which can be referred to as an improved version of SIFT [35], [38]. SURF is faster than SIFT in feature detection [35], [36], making it a more popular choice recently. Similar to SIFT, SURF is also scaling-invariant and rotation-invariant, which makes it possible to identify scaled and rotated images [37]. SURF can be implemented to perform object recognition, registration of images, image classification, 3D reconstruction, and object tracking [35], [38]. The SURF algorithm can be explained generally in three steps, which are keypoint detection, feature description and feature matching [35]. SURF detects the key points based on the approximate Hessian matrix. It computes the coordinational sum of the original image by the origin and a location, by utilizing the integral image approach to speed up the computation. Next, the SURF algorithm describes the features after detecting them, by generating a descriptor using the sum of the Haar wavelet responses. By this way, SURF can achieve high robustness with lower computational time. Lastly, SURF performs feature matching based on calculating the distance between key points [35]–[37].

C. Bag of word (BOW)

The BoW technique has become well known in the computer vision field because of its simplicity and high efficiency [46]. This approach is similar to the BoW approach used in the domain of text documents. It treats the image as a text document, and as words, the features are extracted. The image is composed of visual words describing a local segment that can be characterized by a region or by a neighbourhood reference point [47]. Using this approach, a picture can be represented by a few characteristics.

1) Bag of Visual Words (BOVW): With applications in image classification and categorization, the BoW model applied to images is known as BoVW. The bag model appears to be more discriminating and more general than a local descriptor than a global descriptor. There is a vocabulary in BoVW of the primary visual patterns of an image set of visual words, often also called a dictionary or codebook. The bag of words is generated using a pre-defined visual codebook based on the occurrences of visual words within an image [48]. The method has the same sequence of steps, regardless of the application domain: extraction, codebook, coding, and pooling; and varies mainly in the description of the vocabulary according to the intrinsic characteristics of each domain from one application to another. The Figure 6 presents this process.

D. Fruit Classification

In the business and retail domain, for instance, retail stores, cashiers and self-service checkout systems are implemented to process the transactions made. It is expected that goods to be sold generally follow the barcode scanning registration system, where the goods are scanned to aid in the purchasing process, to identify and register the transactions efficiently. However, that is not the case in vegetables and fruit retailing. The cashier often needs to identify the class of the product while weighing by finding the product class manually in the system [21]. Numerous researches have been conducted in fruit classification, namely classification on different types of fruits [21]–[24], same kind of fruit but different varieties [21], the maturity of fruit [25] as well as the grading and quality of fruit [26]–[28]. The classification on types of fruits can be beneficial towards the consumers and retail market, even the growing agriculture industries, such as helping the consumers to select fruits that are suitable for them, teaches people to identify the characteristics of fruits, implementing them into the robots responsible for fruit harvesting [23], as well as smart refrigerators [22]. Besides, fruit classification aids in the retail market by providing automatic fruit detection and pricing possibilities [21], [23].

III. Methodology

A. Introduction

In this section, the raw dataset will be described together with the methods that are being applied in order to achieve

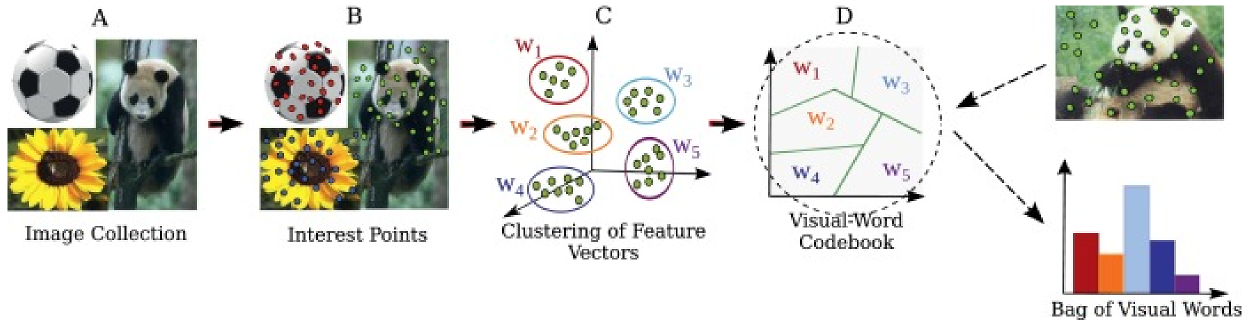

Fig. 6. The Process of Bag of Visual Words [48]

the objectives that are proposed which is to compare different image classification techniques that involves SIFT, SURF and CNN. Several evaluation methods will also be discussed on how to evaluate the efficiency of all the techniques and methods mentioned. References for the methods proposed is from Kaggle [52].

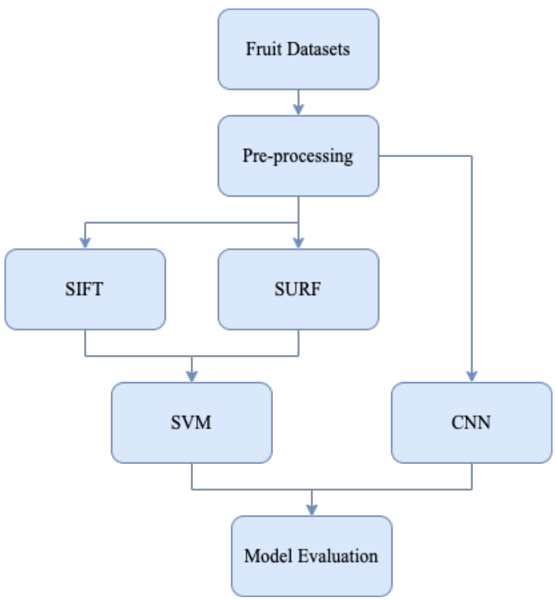

Fig. 7. Flowchart of Methodology

In summary, the methods and the sequence are demonstrated in Figure 7. After preprocessing the dataset obtained, there are basically three methods to do fruit classification. For the first method, SIFT is used for image feature extraction and supported by SVM for classifying the fruits. Secondly, another feature extraction algorithm, SURF, will be used to extract the fruit features and combined with SVM to do fruit classification. In contrast, a deep learning model, Convolutional Neural Network (CNN) will be constructed to classify the fruits, without going through image feature extraction. Lastly, the three methods will be evaluated by computing their accuracy and confusion matrix. Each of the processes will be explained in detail through this section.

B. Dataset

In this section, the dataset that is going to be used for the comparisons and evaluation of the applied techniques will be discussed. The dataset is an open-source dataset on Kaggle named "Fruits-360 data set" from [29] that are images of different types of fruits obtained by filming the fruits while a motor rotates them and then extracting frames. There are 131 different types of fruit classes in the whole dataset where each of the species has already been split into training and test dataset separately. The total number of images in the original dataset is 90483 images altogether with the training size set of 67692 images and a test set consisting of 22688 images. As the classification problem of the study's case is the multi-class classification of different types of fruits, nine varieties of fruits have been selected as our training and test dataset for modelling of fruit image classification. The table II explains the datasets that are chosen from the "Fruits-360 data set" to be used in this study. Besides, Figure 8 display the example of the images of the fruit class that has been chosen.

TABLE II
DATASETS CHOSEN FROM THE "FRUITE-360 DATASET"

Fruit Class	Training Size*	Test Size*
Apple Golden 1	480	160
Avocado	427	143
Banana	490	166
Cherry 1	492	164
Cocos	490	166
Kiwi	466	156
Lemon	492	164
Mango	490	166
Orange	479	160

*Number of Images.

C. Data Preprocessing

After specifying the dataset that is going to be included in this study from the "Fruits-360 data set", several data preprocessing steps have been done for the following model construction steps to be performed smoothly. As the dataset has already been split into training and test sets in the first place, it has made the selection of the datasets easy without needing to perform a train test split. Images that have been read will be saved into two separate arrays based on train and test set accordingly. As the name of the type of fruits is the same

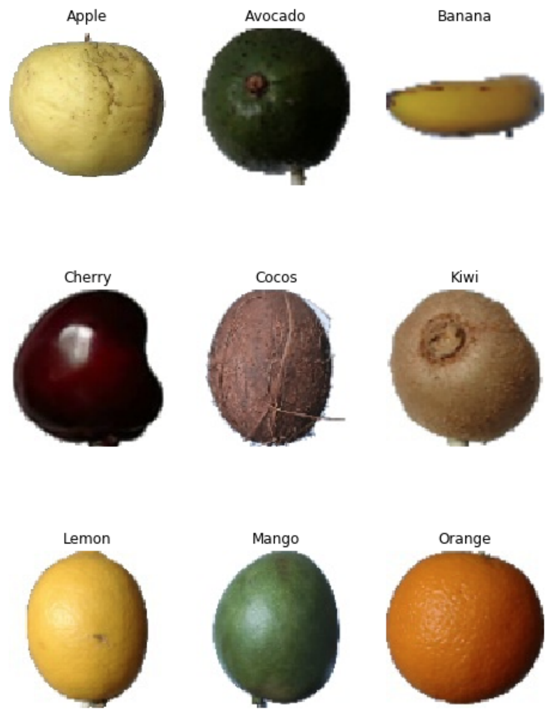

Fig. 8. Sample Image of Fruit Class

as the filename of a particular fruit, a label will also be created and read together with the images from a particular file. The images that are read will be converted from Blue-Green-Red format to Red-Green-Blue format before the construction of the model.

D. Model Construction

There are three methods involved in model construction, as listed below:

1) SIFT + SVM
2) SURF + SVM
3) CNN

The steps in constructing the models are relatively different, mainly the feature extraction related models and the deep learning model. For SIFT + SVM and SURF + SVM, the steps involved are as follows:

1) Load the preprocessed training and test dataset.
2) Use different extraction methods (SIFT and SURF) to extract the color, shape, and texture features of the fruit.
3) Build the visual vocabulary of the fruits using Mini Batch KMeans clustering.
4) Compute the Bag-of-words (BOW) representation for all the fruits.
5) Fit the BOW of all fruits and their classes to SVM classifier.

For CNN, there are no feature extraction techniques required. Hence, instead of feature extraction, several data transformation techniques are applied before fitting the dataset into the deep learning model. The steps are demonstrated as below:

1) Load the preprocessed training and test dataset.
2) Normalize and reshape the fruit images.
3) Convert the labels to categorical data type.
4) Split the training dataset to training and validation dataset with 80%-20% split.
5) Construct and tune the CNN.
6) Fit the training and validation dataset to the model.

For the Convolutional Neural Network, there are 9 layers within the model. The feature extraction within the model is done by passing through 6 layers, basically three iterations of 2D convolution layer with different filter sizes of 8, 16, 32 and kernel size of 3,3 followed by max pooling operation for 2D spatial data with 2,2 pool size. The remaining two layers are two dense layers, with matrix size of 512 and activation function of "relu". For the output dense layer, the matrix size will be 9, as there are 9 classes of the fruits. The activation function used is "softmax". To optimize the model, Adam optimizer is used. The model is made to pass through 8 epoch processes, to make sure that the best possible generalization can be achieved.

E. Evaluation Metric

Model evaluation plays an essential role after model construction, which provides insight into the performance of the model. Besides, the goal of model evaluation is to estimate the abstraction accuracy of a predictive model while fitting unseen data. In multi-class classification, simple metrics such as precision, recall, accuracy, etc., can be implemented. In this study, confusion matrix and the report scores of precision, recall, f1-score and support will also be used as a way to evaluate the models. The equations of accuracy, precision, recall, f1-score and support will be shown below. All of the scores above are based on True Positive, False Positive, True Negative and False Negative. True Positives (TP) are the correctly predicted positive values which mean that the value of the actual class is yes and the value of the predicted class is also yes. True Negatives (TN) are the correctly predicted negative values which mean that the value of the actual class is no and the value of the predicted class is also no. False positives and false negatives, these values occur when your actual class contradicts with the predicted class. False Positives (FP) is when the actual class is no and predicted class is yes, whereas False Negatives (FN) is when the actual class is yes but predicted class is no.

1) Accuracy: Accuracy is the most intuitive performance measure, and it is merely a ratio of correctly predicted observation to the total observations.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

2) Precision: The ratio of correctly predicted positive observations of the total predicted positive observations.

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

3) Recall: Sometimes being mentioned as Sensitivity, is the ratio of correctly predicted positive observations to all observations in the actual class.

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

4) F1-score: The weighted average of Precision and Recall. Therefore, this Score takes both false positives and false negatives into account. Intuitively it is not as easy to understand as accuracy, but F1 is usually more useful than accuracy, especially if you have an uneven class distribution. Accuracy works best if false positives and false negatives have similar cost. If the cost of false positives and false negatives are very different, it's better to look at both Precision and Recall.

$$F1score = \frac{2TP}{2TP + FP + FN} \tag{4}$$

5) Support: The number of occurrences of each class in the ground truth when it is compared to the dependent variable in the test set.

6) Macro-Averaging: The average of the performances of each individual class are calculated.

$$PRE_{macro} = \frac{PRE_1 + ... + PRE_k}{k} \tag{5}$$

IV. Findings and Discussion

Based on the methods that have been proposed in the previous section, several models such as implementing SIFT and SURF as feature extractions and classifying the images to its distributed classes using Support Vector Machine, and using Convolutional Neural Network to perform classification on the images without an external feature extraction method. Findings and observations will be summarised in tables and figures followed by discussions.

Three methods that have been applied which are image classification using SIFT with SVM, image classification using SURF with SVM and also image classification using CNN have all displayed good results which are generally achieving accuracy score that is above 80%. Still, image classification from elicits the highest accuracy in multiclass classification among all three methods that have been used. The summarized accuracy for all of the methods is shown in the table III.

TABLE III
Accuracy of Models

Technique of Image Classification	Accuracy (%)
SIFT and SVM	84.29
SURF and SVM	85.33
CNN	100.00

SURF performs better in the sense of feature extraction in this case than SIFT while CNN performs the best in extracting the features and classifying the images. Through classification of fruits using CNN, an accuracy score of 100% has been obtained after running through a series of epochs.

A. Comparison on Heatmaps of Confusion Matrix and the Classification Report of the Models

The figures 9, 10, and 11 show the heatmaps of confusion Matrix of the models where the classes are correctly classified or misclassified.

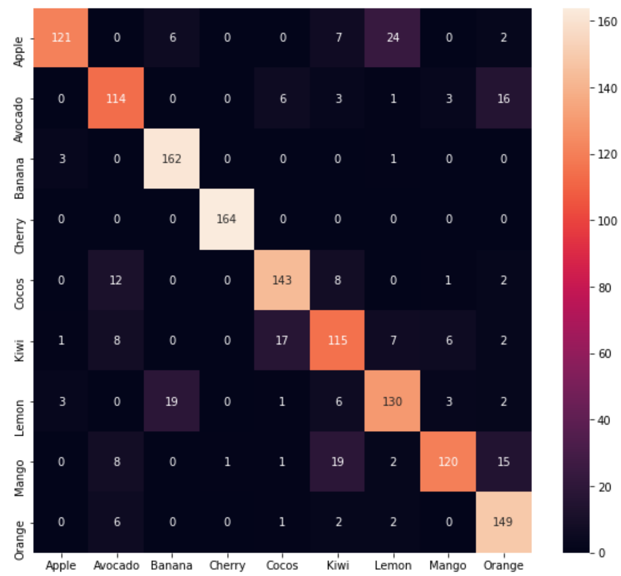

Fig. 9. Confusion Matrix Heatmap of SIFT + SVM

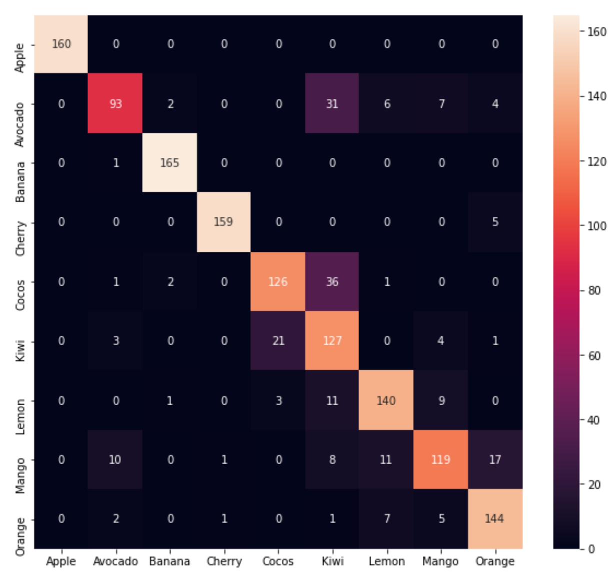

Fig. 10. Confusion Matrix Heatmap of SURF + SVM

The heatmaps of the confusion matrices of the models can explain clearly on the efficiency of the models. The matrix behaves in a two-dimensional plot where the X-axis and Y-axis resemble the classes of types of fruits used in the classification. The value where X is the same type of fruit as Y elicits the correct number of classification and the value where X

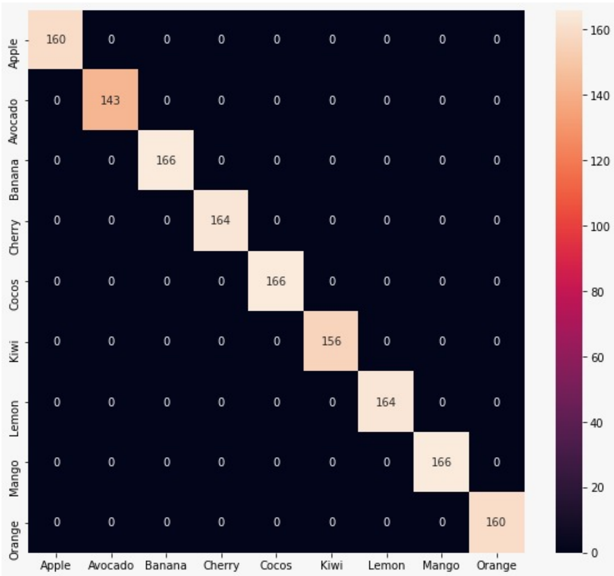

Fig. 11. Confusion Matrix Heatmap of CNN

and Y is not the same type of fruit exhibits the number of misclassification.

The possible reason behind this misclassification based on the observations above might infer that the fruit shares a lot of similar features found on its texture and colour. For example, based on the heatmap of SIFT, the types of fruits that are usually misclassified are kiwis and cocos, lemon and bananas, etc. as these fruits share similar characteristics in terms of colours and texture. The heatmap of SURF exhibits a similar scenario as that of the heatmap of SIFT. Even though the model with SURF is exhibiting the same scenario with SIFT, it is still a slightly better model because it has a lesser number of misclassification than the model that is using SIFT. The model which uses CNN does have any misclassification in classifying the fruits. Therefore the model's accuracy is at a hundred per cent. CNN works the best in this scenario mainly because the dataset that is applied in modelling is clean and specific.

To further dive into more precise evaluations of the models, the classification report that consists of different evaluation score metrics will be shown in table IV, V, and VI.

TABLE IV
CLASSIFICATION REPORT OF SIFT + SVM

	Precision	Recall	F1-Score	Support
Apple	0.95	0.76	0.84	160
Avocado	0.77	0.80	0.78	143
Banana	0.87	0.98	0.92	166
Cherry	0.99	1.00	1.00	164
Cocos	0.85	0.86	0.85	166
Kiwi	0.72	0.74	0.73	156
Lemon	0.78	0.79	0.79	164
Mango	0.90	0.72	0.80	166
Orange	0.79	0.93	0.86	160
Macro Average	0.85	0.84	0.84	1445

TABLE V
CLASSIFICATION REPORT OF SURF + SVM

	Precision	Recall	F1-Score	Support
Apple	1.00	1.00	1.00	160
Avocado	0.85	0.65	0.74	143
Banana	0.97	0.99	0.98	166
Cherry	0.99	0.97	0.98	164
Cocos	0.84	0.76	0.80	166
Kiwi	0.59	0.81	0.69	156
Lemon	0.85	0.85	0.85	164
Mango	0.83	0.72	0.77	166
Orange	0.84	0.90	0.87	160
Macro Average	0.86	0.85	0.85	1445

TABLE VI
CLASSIFICATION REPORT OF CNN

	Precision	Recall	F1-Score	Support
Apple	1.00	1.00	1.00	160
Avocado	1.00	1.00	1.00	143
Banana	1.00	1.00	1.00	166
Cherry	1.00	1.00	1.00	164
Cocos	1.00	1.00	1.00	166
Kiwi	1.00	1.00	1.00	156
Lemon	1.00	1.00	1.00	164
Mango	1.00	1.00	1.00	166
Orange	1.00	1.00	1.00	160
Macro Average	1.00	1.00	1.00	1445

The precision score of the classes on the model that is using SIFT shown in the table, there are a few classes that are exhibiting the score around 0.7 as the result of having much more misclassification, whereas Cherry is the class that has the most correctly classified results based on other classes. In the model using SURF, the classes generally have a good score in terms of F1 scores, where it is the weighted average of both Precision and Recall except for Kiwi as it is greatly misclassified than the other classes of fruits. Apple, Banana and Cherry is classified almost without fault, which exhibits a high score for Precision, Recall and F1 score. As the accuracy of the model is based on the average of the F1-score, it is misleading if the efficiency of the model is solely based on the observation of the accuracy. In this case, even though the classification for some fruits are almost without any fault, there exist one or two greatly misclassified fruits such as Kiwi in the model using SURF. Even though the model with SIFT is having a slightly lower accuracy than the model using SURF, there is no misclassification of the classes of the fruits that is higher than the misclassification of Kiwi in the model using SURF, but instead having distributed misclassifications across the classes of fruits. In this particular case in this particular study, both of these models might not seem to be better models than CNN for image classification problems in industries such as product identification as both the models that use SIFT and SURF might exhibit fault in identification of the product. In this particular case, CNN is having an optimum score and accuracy in predicting the classes of the fruits in the testing set. From the comparison between the scores of these models, CNN yields the best performance against the other models.

This can particularly explain that CNN is the state-of-the-art method in image classification problems as it is accurate and accuracy means saving cost and increasing profit in business and industries.

V. Conclusion and Future Studies

In this project, SIFT and SURF are used to extract features of all fruit images, such as the shape, color, and texture. The extracted feature points are clustered using Mini Batch K-Means clustering to construct BOW for every image. To classify the images, an SVM classifier is used to train a multi-class classifier. Besides, CNN is also used as a comparison to classify the images, with 9 layers including feature extraction and classification inside the model. The performances and results of the three experiments are evaluated by computing the confusion matrix, accuracy, precision, recall and F1 score. The experiment proves that in overall, CNN was able to achieve the best performance and classification results out of the three models. Besides, SURF achieved a slightly better feature extraction result compared to SIFT not to mention SURF obviously required less computational time, compared to SIFT. Thus, it can be said that when dealing with real-world data, the feature extraction methods need to be chosen and evaluated carefully, as many factors such as the computational time and the quality of extracted features have to be taken into consideration with proper weight distribution and concern. The other concern from the experiment taken out in this project is the dataset used. In real-world collected data, there are extra factors that will influence the accuracy of classification, such as noises and technical problems. It is possible to reduce the effect of these factors by carrying out suitable preprocessing techniques to minimize the noise and technical issues. In addition, classification in a single complex image can be studied and experimented further.

References

[1] M. K. Monir Rabby, B. Chowdhury and J. H. Kim, "A Modified Canny Edge Detection Algorithm for Fruit Detection Classification," 2018 10th International Conference on Electrical and Computer Engineering (ICECE), Dhaka, Bangladesh, 2018, pp. 237-240, doi: 10.1109/ICECE.2018.8636811.

[2] M. K. M. Rabby, M. M. Islam and S. M. Imon, "A Review of IoT Application in a Smart Traffic Management System," 2019 5th International Conference on Advances in Electrical Engineering (ICAEE), Dhaka, Bangladesh, 2019, pp. 280-285, doi: 10.1109/ICAEE48663.2019.8975582.

[3] Mureșan, Horea Oltean, Mihai. (2018). "Fruit recognition from images using deep learning," Acta Universitatis Sapientiae, Informatica. 10. 26-42.

[4] Dhaware, C., Wanjale, K. H. (2016). Survey on image classification methods in image processing. Int. J. Comput. Sci. Trends Technol.(IJCST), 4(3), 246-248. Jog 2016 - S. Jog and M. Dixit, "Supervised classification of satellite images," 2016 Conference on Advances in Signal Processing (CASP), Pune, 2016, pp. 93-98, doi: 10.1109/CASP.2016.7746144.

[5] S. Jog and M. Dixit, "Supervised classification of satellite images," 2016 Conference on Advances in Signal Processing (CASP), Pune, 2016, pp. 93-98, doi: 10.1109/CASP.2016.7746144.

[6] Yadav, S.S., Jadhav, S.M. Deep convolutional neural network based medical image classification for disease diagnosis. J Big Data 6, 113 (2019). https://doi.org/10.1186/s40537-019-0276-

[7] K. Shankar, Y. Zhang, Y. Liu, L. Wu and C. Chen, "Hyperparameter Tuning Deep Learning for Diabetic Retinopathy Fundus Image Classification," in IEEE Access, vol. 8, pp. 118164-118173, 2020, doi: 10.1109/ACCESS.2020.3005152.

[8] Yan, R., Ren, F., Wang, Z., Wang, L., Zhang, T., Liu, Y., ... Zhang, F. (2020). Breast cancer histopathological image classification using a hybrid deep neural network. Methods, 173, 52-60.

[9] P. Kumar Mallick, S. H. Ryu, S. K. Satapathy, S. Mishra, G. N. Nguyen and P. Tiwari, "Brain MRI Image Classification for Cancer Detection Using Deep Wavelet Autoencoder-Based Deep Neural Network," in IEEE Access, vol. 7, pp. 46278-46287, 2019, doi: 10.1109/ACCESS.2019.2902252.

[10] Brinker, T. J., Hekler, A., Enk, A. H., Klode, J., Hauschild, A., Berking, C., ... Utikal, J. S. (2019). A convolutional neural network trained with dermoscopic images performed on par with 145 dermatologists in a clinical melanoma image classification task. European Journal of Cancer, 111, 148-154.

[11] Guan, Q., Huang, Y. (2020). Multi-label chest x-ray image classification via category-wise residual attention learning. Pattern Recognition Letters, 130, 259-266.

[12] Seo, Y., Shin, K. S. (2019). Hierarchical convolutional neural networks for fashion image classification. Expert Systems with Applications, 116, 328-339.

[13] P. Wajire, S. Angadi and L. Nagar, "Image Classification for Retail," 2020 International Conference on Industry 4.0 Technology (I4Tech), Pune, India, 2020, pp. 23-28, doi: 10.1109/I4Tech48345.2020.9102699.

[14] Myint, S. W., Lam, N. (2005). A study of lacunarity-based texture analysis approaches to improve urban image classification. Computers, environment and urban systems, 29(5), 501-523.

[15] M. De Martinao, F. Causa and S. B. Serpico, "Classification of optical high resolution images in urban environment using spectral and textural information," IGARSS 2003. 2003 IEEE International Geoscience and Remote Sensing Symposium. Proceedings (IEEE Cat. No.03CH37477), Toulouse, 2003, pp. 467-469 vol.1, doi: 10.1109/IGARSS.2003.1293811.

[16] Wang, J., Yu, X., Liu, Q. et al. Research on key technologies of intelligent transportation based on image recognition and anti-fatigue driving. J Image Video Proc. 2019, 33 (2019). https://doi.org/10.1186/s13640-018-0403-6

[17] S. Wan, X. Xu, T. Wang and Z. Gu, "An Intelligent Video Analysis Method for Abnormal Event Detection in Intelligent Transportation Systems," in IEEE Transactions on Intelligent Transportation Systems, doi: 10.1109/TITS.2020.3017505.

[18] Nurhasanah, Y. I., Kartika, A. D., Nur'aini, N. (2020). Public Transportation Identification System in Prohibited Areas based on Traffic Signs using Image Processing. Electrotehnica, Electronica, Automatica, 62(2).

[19] A. Ding and X. Zhou, "Land-Use Classification with Remote Sensing Image Based on Stacked Autoencoder," 2016 International Conference on Industrial Informatics - Computing Technology, Intelligent Technology, Industrial Information Integration (ICIICII), Wuhan, 2016, pp. 145-149, doi: 10.1109/ICIICII.2016.0044.

[20] Li, W., Fu, H., Yu, L., Gong, P., Feng, D., Li, C., Clinton, N. (2016). Stacked Autoencoder-based deep learning for remote-sensing image classification: a case study of African land-cover mapping. International journal of remote sensing, 37(23), 5632-5646.

[21] Rojas-Aranda J.L., Nunez-Varela J.I., Cuevas-Tello J.C., Rangel-Ramirez G. (2020) Fruit Classification for Retail Stores Using Deep Learning. In: Figueroa Mora K., Anzurez Marín J., Cerda J., Carrasco-Ochoa J., Martínez-Trinidad J., Olvera-López J. (eds) Pattern Recognition. MCPR 2020. Lecture Notes in Computer Science, vol 12088. Springer, Cham. https://doi.org/10.1007/978-3-030-49076-8_1

[22] Khan, R., Debnath, R. (2019). Multi class fruit classification using efficient object detection and recognition techniques. International Journal of Image, Graphics and Signal Processing, 11(8), 1.

[23] Pl, Chithra Henila, M. (2019). Fruits Classification Using Image Processing Techniques. 7. 131-135. 10.26438/ijcse/v7si5.131135.

[24] Khaing, Z. M., Naung, Y., Htut, P. H. (2018). Development of control system for fruit classification based on convolutional neural network. 2018 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus). doi:10.1109/eiconrus.2018.8317456

[25] Hambali, Hamirul'Aini Abdullah, Sharifah Jamil, Nursuriati Harun, Hazaruddin. (2016). Fruit Classification using Neural Network Model.

[26] M. A. Shahin, E. W. Tollner, R. W. McClendon, H. R. Arabnia. (2002). APPLE CLASSIFICATION BASED ON SURFACE BRUISES USING IMAGE PROCESSING AND NEURAL NETWORKS. Transactions of the ASAE, 45(5). doi:10.13031/2013.11047

[27] Gupta, A., Kushwaha, M. R., Patidar, M. K. (2019). A Comparative Study of Fruit Recognition System Using Machine Learning Approaches. International Journal of Engineering Technology and Applied Science, 5(7).

[28] J. M. Ponce, A. Aquino and J. M. Andújar, "Olive-Fruit Variety Classification by Means of Image Processing and Convolutional Neural Networks," in IEEE Access, vol. 7, pp. 147629-147641, 2019, doi: 10.1109/ACCESS.2019.2947160.

[29] Horea Muresan, Mihai Oltean, Fruit recognition from images using deep learning, Acta Univ. Sapientiae, Informatica Vol. 10, Issue 1, pp. 26-42, 2018.

[30] Moren, K., Göhringer, D. A framework for accelerating local feature extraction with OpenCL on multi-core CPUs and co-processors. J Real-Time Image Proc 16, 901–918 (2019).

[31] Shuqi Cui, Hong Jiang, Zheng Wang, Chaomin Shen. (2017). Application of neural network based on SIFT local feature extraction in medical image classification. 2017 2nd International Conference on Image, Vision and Computing (ICIVC). doi:10.1109/icivc.2017.7984525

[32] Zeglazi, O., Amine, A., Rziza, M. (2016). Sift Descriptors Modeling and Application in Texture Image Classification. 2016 13th International Conference on Computer Graphics, Imaging and Visualization (CGiV). doi:10.1109/cgiv.2016.58

[33] Li, Q., Wang, X. (2018). Image Classification Based on SIFT and SVM. 2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS). doi:10.1109/icis.2018.8466432

[34] Sarangi, S., Sahidullah, M., Saha, G. (2020). Optimization of data-driven filterbank for automatic speaker verification. Digital Signal Processing, 102795.

[35] Naik, S., Patel, B. (2017). Machine vision based fruit classification and grading-a review. International Journal of Computer Applications, 170(9), 22-34.

[36] Yuk, E., Park, S., Park, C.-S., Baek, J.-G. (2018). Feature-Learning-Based Printed Circuit Board Inspection via Speeded-Up Robust Features and Random Forest. Applied Sciences, 8(6), 932. doi:10.3390/app8060932

[37] Cao, P., Shu, H., Yang, B., Dong, J., Fang, Y., Yu, T. (2018). Speeded-up robust features based single-ended travelling wave fault location: a practical case study in Yunnan power grid of China . IET Generation, Transmission Distribution, 12(4), 886–894. doi:10.1049/iet-gtd.2017.0735

[38] Muhathir, M., Hidayah, W., Ifantiska, D. (2020). Utilization of Support Vector Machine and Speeded up Robust Features Extraction in Classifying Fruit Imagery. Computer Engineering and Applications Journal, 9(3), 183-193.

[39] Wang, R., Shi, Y., Cao, W. (2019). GA-SURF: A new speeded-up robust feature extraction algorithm for multispectral images based on geometric algebra. Pattern Recognition Letters, 127, 11-17.

[40] M. Manzoor, Y. Morgan, "Support Vector Machine based Vehicle Make and Model Recognition System", Advances in Science, Technology and Engineering Systems Journal, vol. 2, no. 3, pp. 1080-1085 (2017).

[41] GIS Geography. (2020, July 30). Supervised and Unsupervised Classification in Remote Sensing. https://gisgeography.com/supervised-unsupervised-classification-arcgis/

[42] Drucker, H., Burges, C. J., Kaufman, L., Smola, A. J., Vapnik, V. (1997). Support vector regression machines. In Advances in neural information processing systems (pp. 155-161).

[43] Olgun, M., Onarcan, A. O., Özkan, K., Işik, Ş., Sezer, O., Özgişi, K., ... Koyuncu, O. (2016). Wheat grain classification by using dense SIFT features with SVM classifier. Computers and Electronics in Agriculture, 122, 185-190.

[44] A. T. Nasser and N. Dogru, "Signature recognition by using SIFT and SURF with SVM basic on RBF for voting online," 2017 International Conference on Engineering and Technology (ICET), Antalya, 2017, pp. 1-5, doi: 10.1109/ICEngTechnol.2017.8308208.

[45] Q. Li and X. Wang, "Image Classification Based on SIFT and SVM," 2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS), Singapore, 2018, pp. 762-765, doi: 10.1109/ICIS.2018.8466432.

[46] Ayadi, W., Elhamzi, W., Charfi, I., Atri, M. (2019). A hybrid feature extraction approach for brain MRI classification based on Bag-of-words. Biomedical Signal Processing and Control, 48, 144-152.

[47] Cheng, J., Huang, W., Cao, S., Yang, R., Yang, W., Yun, Z., ... Feng, Q. (2015). Enhanced performance of brain tumor classification via tumor region augmentation and partition. PloS one, 10(10), e0140381.

[48] Silva, F. B., Werneck, R. D. O., Goldenstein, S., Tabbone, S., Torres, R. D. S. (2018). Graph-based bag-of-words for classification. Pattern Recognition, 74, 266-285.

[49] Zhang, M., Li, W., Du, Q. (2018). Diverse Region-Based CNN for Hyperspectral Image Classification. IEEE Transactions on Image Processing, 27(6), 2623–2634. doi:10.1109/tip.2018.2809606

[50] Yang, X., Ye, Y., Li, X., Lau, R. Y. K., Zhang, X., Huang, X. (2018). Hyperspectral Image Classification With Deep Learning Models. IEEE Transactions on Geoscience and Remote Sensing, 1–16. doi:10.1109/tgrs.2018.2815613

[51] Agarap, A. F. (2017). An architecture combining convolutional neural network (CNN) and support vector machine (SVM) for image classification. arXiv preprint arXiv:1712.03541.

[52] Kaggle: Your Machine Learning and Data Science Community. (n.d.). Kaggle. Retrieved October 22, 2020, from https://www.kaggle.com/