SINET: Ethiop. J. Sci., 45(3):309-321, 2022 ISSN: 0379–2897 (PRINT)
 eISSN: 2520–7997

Date received: July 03, 2022; Date revised: December 20, 2022; Date accepted: December 25, 2022
DOI: https://dx.doi.org/10.4314/sinet.v45i3.6

Multiclass classification of Ethiopian coffee bean using deep learning

Getabalew Amtate[1] and Dereje Teferi[2*]

[1] Faculty of Informatics, Hawassa University. E-mail: getabalew@hu.edu.et
[2] School of Information Science, Addis Ababa University. E-mail: dereje.teferi@aau.edu.et

ABSTRACT: Ethiopia is the homeland of Coffee Arabica. Coffee is the major export commodity and a high-income source of foreign currency for the country. In addition to this, coffee has a great role in social interaction between people and is also a source of income for the coffee-producing farmers. Several types of coffee beans grow in Ethiopia. These beans are distinct from each other in terms of quality, color, shape etc. based on their geographical origins. Classification of these coffee beans are based on growing origin, altitude, bean shape and color, preparation method and others. However, the quality of the coffee beans is determined by visual inspection, which is subjective, laborious, and prone to error. This creates the necessity for the development of an automatic method that is precise, non-destructive and objective. Thus, this research aims to develop a model that classifies coffee beans of six different origins of Ethiopia (Jimma, Limmu, Nekemte, Yirgacheffe, Bebeka, and Sidama) in to nine classes. The dataset for this research is collected from the Ethiopian Coffee Quality Inspection and Auction Center (ECQIAC). This research followed design science research (DSR) to investigate the problem. Image processing and the state-of-the-art deep-learning techniques were employed to automatically classify coffee bean images into nine different classes grown in six different regions of Ethiopia. A total of 8646 coffee bean images were collected and 1190 images were added using augmentation to make the total dataset 9836. The model is trained and tested by tuning the hyper-parameters of the CNN algorithm. When 80% of the dataset is used for training, 10% for validation, and the remaining 10% for testing, the proposed model achieved a 99.89% overall classification accuracy with 0.92% generalization log-loss. In conclusion, the result of this research shows that deep learning is an effective technique for classification of Ethiopian coffee beans and can be implemented in the coffee industry.

Keywords/ Phrases: Ethiopian coffee, Coffee bean classification, Deep learning, CNN

INTRODUCTION

Ethiopia is the birthplace of coffee Arabica species (Danakil, 2016), where the coffee plant originated. Nowadays, coffee Arabica is grown in different parts of the world. Different varieties of coffee grow in different regions of Ethiopia. Ethiopian coffee beans can be categorized as Yirgacheffe, Sidamo, Harrar, Bebeka, Teppi, Limmu, Jimma, Illubabor, Lekempti, Wellega, and Gimbi based on their growing areas. Ethiopia has three major coffee brandings, namely: Sidamo, Yirgacheffe, and Harar (Espresso and Coffee Guide, 2020). Coffee beans from these three regions have their own unique characteristics, brand, and label (Nomad Coffee Club, 2020).

Green coffee is classified for export to define quality criteria for a fair system of pricing. However, there is no universally accepted classification system and hence many countries develop their own classification criteria to set the minimum standards. Some of the criteria are altitude, growing region/origin, preparation method (dry and wet), bean size, shape and color, number of defects, and cup quality (International Coffee Organization, 2018). In Ethiopia, green coffee is mainly classified based on coffee class/preparation method (washed and unwashed), and type based on growing region (Chemonics International Inc, 2010). The identification and classification process are done manually by a group of people through traditional eye inspection and the experts previous experiences. This approach is subjective, time taking, and prone to error (Habtamu Minassie, 2008). As a result, objective identification and classification of coffee varieties has become a pressing issue.

The purpose of this research is therefore to design and develop a coffee bean classification method using deep learning. Nine coffee bean classes that are commercially grown in six different regions in Ethiopia are considered for

*Author to whom correspondence should be addressed.

this research. These coffee bean classes are: ***washed Limmu, unwashed Limmu, washed Sidamo, unwashed Sidamo, washed Yirgacheffe, unwashed Yirgacheffe, unwashed Jimma, unwashed Nekemte, and washed Bebeka.***

Related works

Several researches are conducted using image processing along with different learning-based classifier algorithms to classify coffee beans mainly for marketing.

(Habtamu Minassie, 2008) used digital image processing with neural network for classification of Ethiopian coffee beans and achieved a 77.4% accuracy using 309 image dataset.

(Binyam Girma, et al., 2013) attempted to classify Hararghe coffee bean images genotype based on their morphological, color and texture features. They Artificial Neural Network (ANN) along with morphological features of the images and scored 99.4% accuracy on using 390 image data sets.

(de Oliveira et al., 2016) used ANN and Bayes classifiers to classify green coffee bean samples into their market grades. The authors selected 120 (50 g) samples (30 per color) of four-color groups: whitish, green, cane green, and bluish green and achieved 100% accuracy of classifying them into four groups based on their color. These color classes are often used to classify coffee beans commercially.

(Belay Birhanu and Tesfa Tegegn, 2015; Nasution and Andayani, 2017) used ANN with and without back propagation to classify roasted coffee beans into 16 different categories of testes and achieved a 98.2% and 97.5% accuracy respectively.

(Arboleda et al. 2018) used ANN and K-nearest neighbor to classify coffee beans from different towns of Cavite, the Philippines according to their species (Robusta, Excelsa, and Liberica) using 190 coffee bean images for training and 55 coffee bean images for testing. The authors found that ANN classifier performed better than k-nearest neighbor achieving a 96.66% classification accuracy.

Research Methodology

Design Science Research method (DSR) was used in this research. The DSR process model proposed by (Peffers et al. 2007; Peffers et al. 2016) consists six activities (see Figure 1).

These six activities of DSR are adopted for the development of the coffee bean detection and classification model as follows:

Step 1. Problem identification and motivation: to find a research gap i.e. to find the problem, different works of literature on the same domain are reviewed;

Step 2. The objective of a solution: to design and build a model that detects and classifies coffee beans found in different parts of Ethiopia;

Step 3. Design and development: to design and develop the model, relevant data is collected and pre-processed. Deep learning is used to develop the model and conduct rigorous experiments;

Step 4. Demonstration: The model was presented to academicians and experts of ECQIAC to collect comments for improvement;

Step 5. Evaluation: the model was evaluated using precision, recall, f-score and the result was compared with similar researches;

Step 6. Communication: This is done through publication on conferences and journals to reach a larger research community.

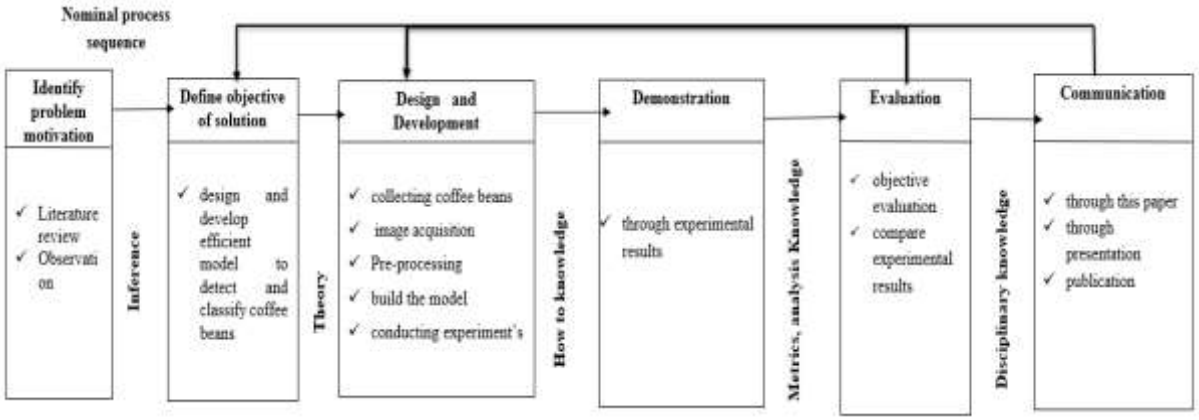


Figure 1: DSR process for coffee bean classification (adopted from (Binyam Girma *et al.*, 2013).

Data Collection

Data collection is one of the challenging tasks in deep learning. This is mainly because deep learning requires a large amount of data to train the algorithm. The coffee beans for this research are collected from the Ethiopian Coffee Quality Inspection and Auction Center (ECQIAC) because all coffee beans samples at the office are certified by domain experts of the laboratory. The coffee bean samples are collected from six different coffee-growing regions (Jimma, Limmu, Sidamo, Yirgacheffe, Bebeka, and Nekemte) in July 2019 G.C. Coffee from these regions is the most widely planted and popular brands grown in Ethiopia and hence are selected for this research. However, coffee beans that is: washed Jimma, unwashed Bebeka, and washed Nekemte are not included in this research because their samples were not available in ECQIAC.

Image Acquisition

The input for the deep learning algorithm is digital images of the coffee beans. The coffee beans collected from ECQIAC for each coffee growing region were captured using a Samsung Galaxy A20 smart phone camera having 13 Mega Pixels resolution. The snapshots of the coffee bean samples were taken by placing the coffee bean on a white background. The camera was set at 3:4 ratio, x3.0 zoom, held in a position normal to the plane of the coffee bean roughly 10 inches above the surface where the coffee bean is placed. Both the front and backside of the coffee beans pictures are taken and each is considered as a different image when the model is deployed to perform the classification. A total of 8,646 images were obtained from 4323 coffee beans (front and back) and each of these images are labelled with the actual class. The images are stored in Joint Photographic Experts Group (.jpg) format with a 4128 X 3096 resolution.

Table 1 shows the number of coffee bean samples, the number of images, and the size of all images datasets in each category (coffee origin and processing type).

Table 1: Coffee bean image data captured from six growing origins.

Coffee origin	Processing type	Number of coffee beans	Coffee bean images
Jimma	Unwashed	527	1054
Limmu	Unwashed	450	900
	Washed	591	1182
Sidamo	Unwashed	279	558
	Washed	569	1138
Yirgacheffe	Unwashed	447	894
	Washed	560	1120
Bebeka	Washed	528	1056
Nekemte	Unwashed	372	744
Total		**4323**	**8646**

Image Pre-processing

The main goal of image pre-processing is to improve image data by suppressing unnecessary distortion to enhance some image features which are important for further processing. The pre-processing stages consists the following steps:

a) Appropriate naming of coffee bean images in each class;
b) Removing the background of the image;
c) Cropping and resizing the images to the same size of (224 x 224) as shown in Figure 2.
d) *Figure* 2Smoothing the images by removing the noise without losing the principal feature of the image;
e) Balancing coffee images across each class using augmentation. A total of 1190 images were created for the under-sampled datasets using different augmentation techniques such as rotation and flipping. The augmented images are only used for training the model; and
f) Data transformation from categorical to numerical representation using encoding techniques.

Finally, the coffee bean images were split into training, validation and testing dataset. Two types of dataset splitting are used in the experiments. The first split is (70, 15, 15) where 70% of the dataset is used for training, 15% is used for validation, and the remaining 15% is used for testing. The second split is (80, 10, 10) where 80% of the dataset is used for training, 10% is used for validation, and the remaining 10% is used for testing. The training data is used to train the model, the validation data is used to measure the model generalization and to halt training when generalization stops improving, and the testing data is used measure the performance of the model.

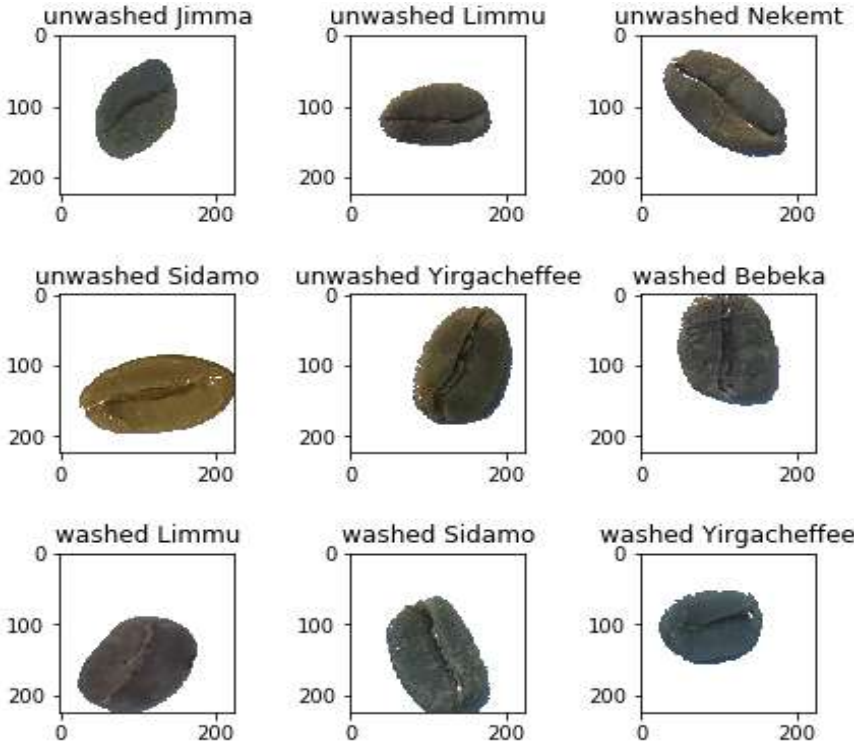


Figure 2: Coffee bean images of each class after pre-processing.

The Learning Model

The learning model is at the center of any machine learning algorithm. Deep learning, the learning algorithm used in this research, is a subfield of artificial intelligence that uses multiple stacked layers of processing units to learn high-level representation from a huge volume of data (Foster, 2019). Each layer learns a complex intrinsic features and multi-level abstract representation of data by itself (Alom, et al., 2019; Sudharsan, 2019).

Convolutional Neural Networks (CNN) are one type of deep neural network commonly used in computer vision applications (Reddy et al., 2018). CNN has been used in several previous researches for face detection (Kwolek, 2005; Deffo et al., 2018), object detection (Galvez et al. 2018), smile detection (Sang et al. 2017; Qu, et al., 2018). Human activity recognition (Gadzicki et al., 2018) etc. A CNN consists of an input layer, an output layer as well as two or more hidden layers. The hidden layers consist of a series of convolutional layers**.**

In this research, CNN with back-propagation algorithm is used. CNN takes an image as input, processes it, and classifies it under certain categories. To train and test a CNN model in deep learning, each image should pass through a series of convolutional layers with kernels, pooling layer, fully connected layer, and finally an activation function to classify an object based on its probability.

The proposed CNN architecture has one input layer, four convolutional layers, two fully connected layers, and one output layer as shown in Figure 3. The first layer is the input layer which provides normalized coffee bean images to the first convolutional layer. The convolutional layer has the role of extracting the features from the images automatically using multiple spatial filters. The extracted features include: edges, colors, textures, corners etc. The four convolutional layers have 32, 64, 128, and 128 kernel respectively.

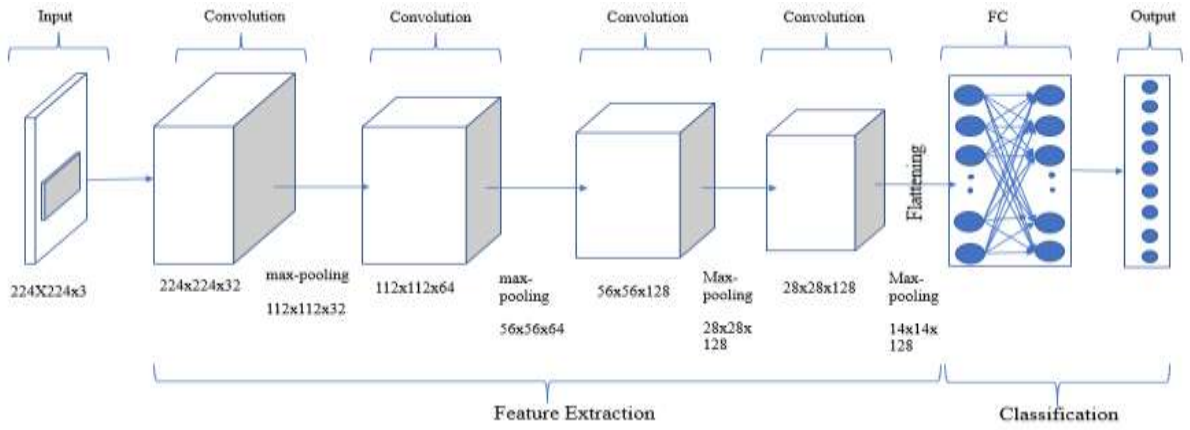


Figure 3: The proposed CNN architecture.

Let's assume that the input coffee bean images (224X224X3) are $a^{[0]}$ and the spatial filters (3X3X3) are the weights $W^{[1]}$ of layer one. The feature maps generated by the convolutional layer is calculated by Equation 1.

$$Z^{[l]} = \sum_{nc=0}^{nc-1} a_{nc}^{[l-1]} . w_{nc}^{[l]} + b^{l} \quad (1)$$

Where *nc* is the number of channels of the coffee bean images (3 for color images), $Z^{[l]}$ is the output of the convolution, *b* is the bias, and *l* is the layer of the convolution. The convolution operation by the spatial filters is performed on each channel separately and summed up together with the bias.

The activation function has the role to activate the important features taken and keeps the non-linearity of the features. The activation function Equation 2 is applied on $Z^{[l]}$.

$$a^{[l]} = g(Z^{[l]}) \quad (2)$$

where $a^{[l]}$ is the feature map of the l^{th} layer.

The convolutional layer is followed by batch normalization and pooling layer. The pooling layer is used to extract dominant features and reduce the large dimension of the feature map. Max pooling with a kernel size of 2x2 and a stride of 2 is applied here. Max pooling takes the largest element from the rectified feature map within the given window. The pooling layer operates on each channel and feature map independently. The size of the feature map is calculated after applying max-pooling given input size ($N_h X N_w$), pooling kernel size(K), and stride (S) of the max-pooling as shown in Equation 3.

$$\frac{N_h - K}{S} + 1 \times \frac{N_w - K}{S} + 1 \quad (3)$$

The fully connected layer takes the features to predict the class by applying the weights. The fully connected layer has 256 and 512 kernel sizes respectively whereas the output layer has a one-dimensional array size of nine which is equivalent to the number of coffee bean classes. The output of the fully-connected layer is given by Equations 4 and 5.

$$u_j = \sum_{i=1}^{I} w_{ji} x_i + b_j \quad (4)$$

$$z_j = f(u_j) \quad (5)$$

where f(.) is the activation function. Rectified Linear Unit (ReLU) is used as the activation function for the convolutional as well as the fully-connected layers. The ReLU activation function is shown in equation 6.

$$f(x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (6)$$

For the final layer, softmax activation function is used. The softmax activation function generally assigns the probability of each coffee bean image belonging to each of the nine classes using Equation 7.

$$F(x) = \frac{e^{x}}{\sum_{k=1}^{k} e^{x}} \quad (7)$$

In addtion to those layers discused, drop regularazation techniqques and batch normalization layers are used as shown in Figure 4.

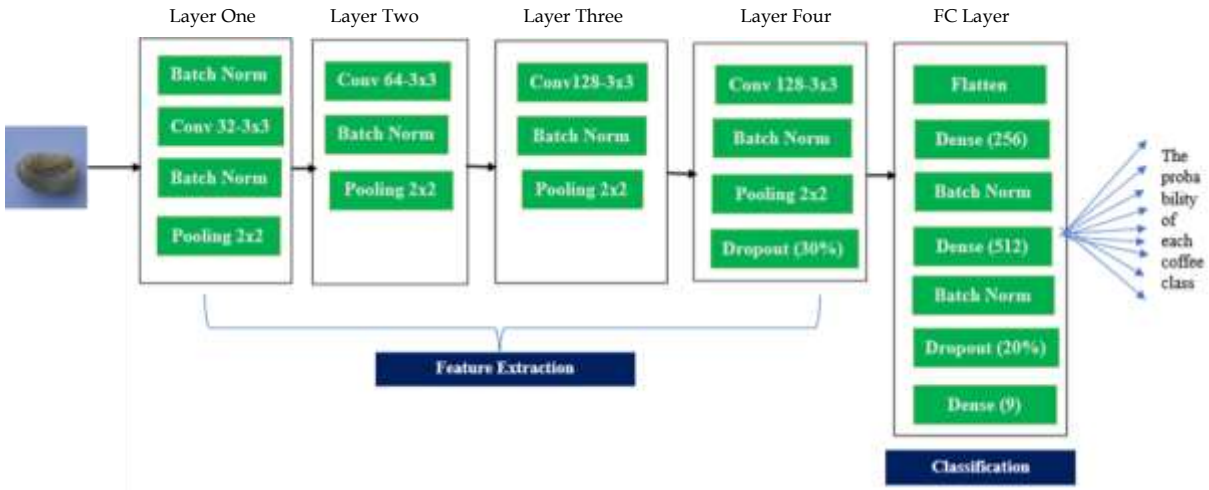


Figure 4: Components of proposed CNN model.

System Architecture

The proposed architecture is a combined effect of different processes. The system accepts coffee bean images that are captured through an image acquisition component. The coffee bean images are transferred to the pre-processing component of the system. In the pre-processing stage, different techniques are applied to get an enhanced coffee bean image. Those enhanced coffee bean images are split into training and testing datasets. The proposed CNN classification model is trained on the training dataset and validated with the validation dataset. Finally, the trained model is evaluated using the testing dataset. The proposed system architecture is shown in Figure 5.

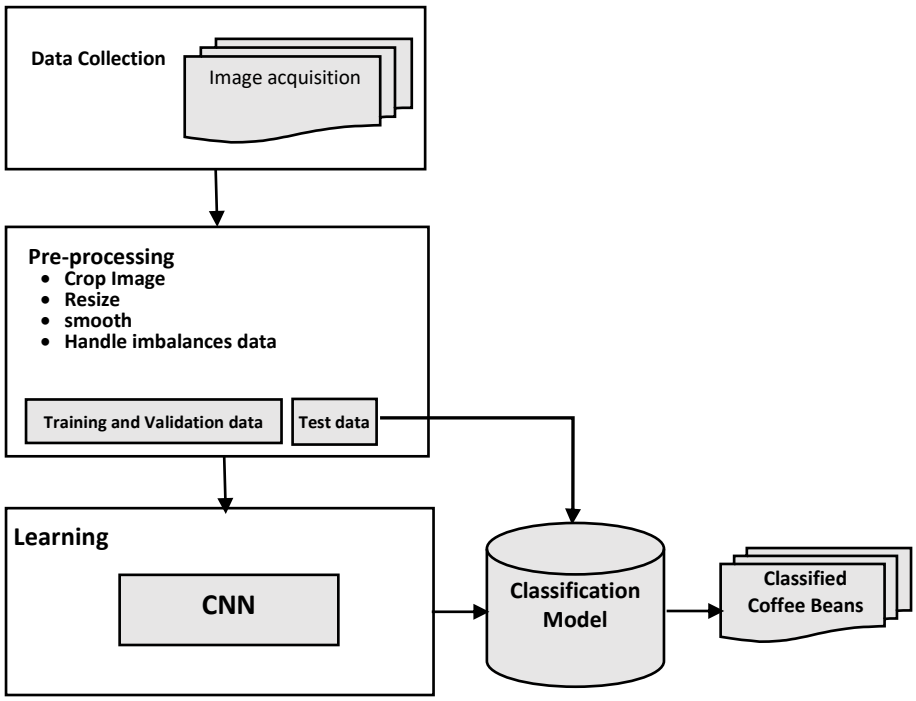


Figure 5: Architecture of the proposed system.

The system architecture in Figure 5 contains four major components, which are explained in detail in the previous sections. These are:

a) The data collection: here the images are acquired using a digital camera and saved in .jpeg format;
b) The pre-processing: here the collected images are processed to get them ready for the

learning model. The processing includes cropping, resizing, balancing, smoothing, and finally splitting into training, validation, and test datasets;

c) The learning: here the different hyper parameters are set and experiments are conducted to find the best fitting classification model for the research in question; and
d) The classification; this is where the classification model is evaluated using metrics like precision, recall, accuracy, and F-score using unseen test data.

Experiment and Discussion of Results

Three experiments were conducted to select the best model. In all the experiments, the following parameters and activities were applied. Each model was trained for 30 epochs, *Log-loss* (also called categorical_crossentropy) is used as a loss function, *Adamax* is used as an optimizer, and *Reduce LROn Plateaus* callback is used to reduce the learning rate. In each experiment a batch gradient decent with a batch size of 64 is used for back propagation and weights are initialized as proposed in (Chellot, 2017). Table 2 shows the number of coffee beans images used for training, validation and testing the model in each of the experiments.

Table 2: Number of images used for training, validation and testing the model from each coffee bean class.

Coffee bean class	Splitting of the coffee bean image dataset					
	Dataset 70-15-15			Dataset 80-10-10		
	Training	Validation	Testing	Training	Validation	Testing
Unwashed Jimma	737	158	159	843	105	106
Unwashed Limmu	712	152	154	814	101	103
Unwashed Nekemte	722	154	156	825	103	104
Unwashed Sidamo	781	167	168	892	111	113
Unwashed Yirgacheffe	784	168	168	896	112	112
Washed Bebeka	739	158	159	844	105	107
Washed Limmu	828	177	178	946	118	119
Washed Sidamo	795	170	172	909	113	115
Washed Yirgacheffe	784	168	168	896	112	112
Total	**6882**	**1472**	**1482**	**7865**	**980**	**991**
	9836			**9836**		

Note that the dataset is split as 70-15-15 means that 70% is used for training, 15% for validation, and 15% for testing the model. Similarly, the dataset is split 80-10-10 implies 80% is used for training, 10% for validation, and 10% for testing the model.

Experiment without Applying Image Augmentation

Two experiments, were conducted in this research using the 70-15-15 and 80-10-10 dataset split. The dataset is split for training, validation, and testing respectively.

Model Training: -The model is trained in both dataset splits for 30 epochs using the *Adamax* optimizer with the learning rate of 0.002, beta_1=0.9, and beta_2=0.999. A batch gradient descent with a batch size of 128 was used that defines the amount of data the model samples at a single optimization step. Since a multiclass classification (having 9 classes of coffee beans) is sought, the loss function *categorical_crossentropy* is used.

After training the model, a training accuracy of 99.7% and a validation accuracy of 95.1% is achieved for the 70-15-15 dataset split whereas a 99.8% training accuracy and a 95.2% validation accuracy is registered for the 80-10-10 split.

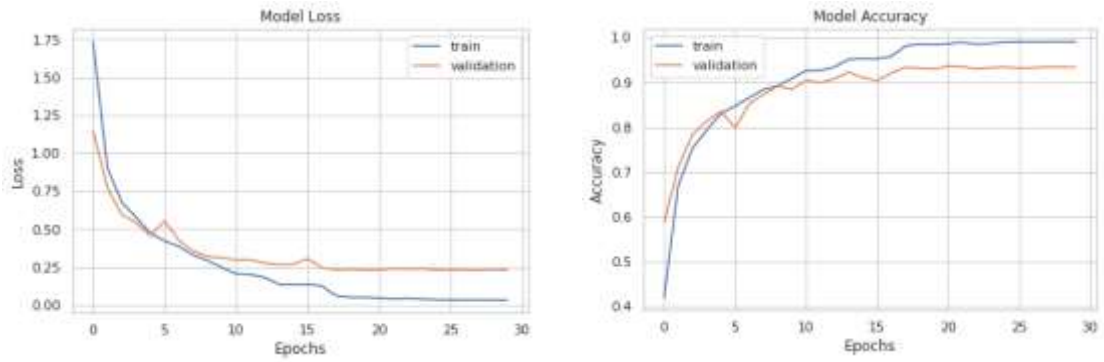


Figure 6: Training and validation curves of RGB dataset with 70-15-15 split.

The model shows some overfitting on the training data as can be seen from the training and validation curves shown in Figure 6 and Figure 7.

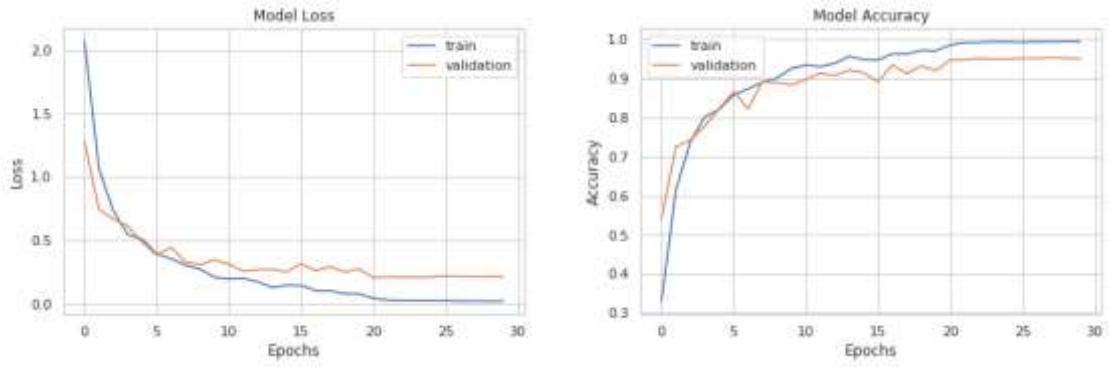
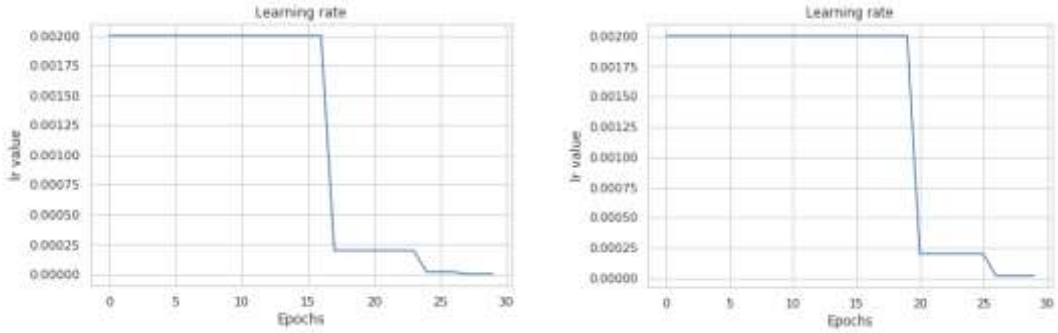


Figure 7: Training and validation accuracy of RGB dataset with 80-10-10 split.

The model accuracy curves in Figure 8 show that the accuracy of the model on the validation dataset saturated after 20 epochs even if a learning rate reducer is used.

Figure 8 shows how the default learning rate is decreased to get better performance of the model on the validation data.



(A) (B)

Figure 8: Learning rate curves of the model (A) split 70-15-15 (B) split 80-10-10.

From the learning rate curve, the validation accuracy doesn't improve after 17th epoch. After waiting for three epochs, the learning rate is reduced to 0.00025 in the first case. Whereas, the validation accuracy doesn't improve after the 20th epoch in the second case. Therefore, the default learning rate is reduced to 0.00025 and the training is continued.

Model Evaluation: The performance of the model is evaluated on the test dataset split using both the 70:15:15 and 80:10:10 cases.

Case 1: Model is trained on 70:15:15 dataset split. The model evaluation registered a 94.8% accuracy and 18.54% log-loss on the 15% test dataset.

(A)

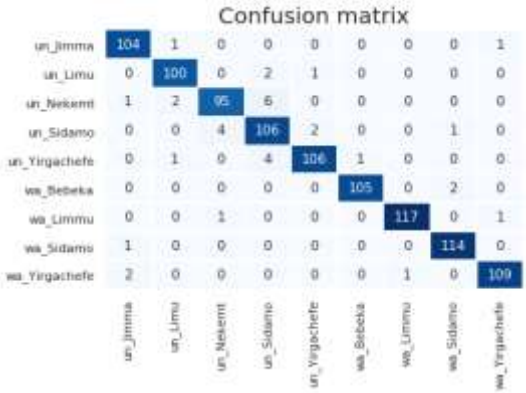
	precision	recall	f1-score	support
unwashed jimma	0.96	0.95	0.95	159
unwashed limmu	0.94	0.96	0.95	154
unwashed nekemt	0.93	0.85	0.89	156
unwashed sidamo	0.87	0.93	0.90	168
unwashed yirgachefe	0.93	0.92	0.93	168
washed bebeka	0.96	0.97	0.97	159
washed limmu	0.99	0.99	0.99	178
washed sidamo	0.98	0.97	0.97	172
washed yirgachefe	0.98	0.98	0.98	168
accuracy			0.95	1482
macro avg	0.95	0.95	0.95	1482
weighted avg	0.95	0.95	0.95	1482

(B)

Figure 9: Evaluation Result of the model on dataset split as 70-15-15 (A) confusion matrix (B) Classification report.

The confusion matrix on Figure 9 (A) shows that the model is confused to classify the coffee beans especially for *unwashed Nekemt* and *unwashed Sidamo*. The model misclassified 23 *unwashed Nekemte* coffee beans and 12 *unwashed Sidamo* coffee beans to another coffee bean class. This is observed in the low recall of those coffee bean classes (see Figure 9 (B)). On the other hand, the precision of *washed Bebeka* and *washed limmu* coffee bean classes are high. Only 2 coffee beans in each class are misclassified for another coffee bean class.

Case 2: - Model is trained on 80:10:10 dataset split. The model evaluation registered a 96.4% accuracy and 12.79% log-loss on the 10% test dataset.



(A)

	precision	recall	f1-score	support
unwashed jimma	0.96	0.98	0.97	106
unwashed limmu	0.96	0.97	0.97	103
unwashed nekemt	0.95	0.91	0.93	104
unwashed sidamo	0.90	0.94	0.92	113
unwashed yirgachefe	0.97	0.95	0.96	112
washed bebeka	0.99	0.98	0.99	107
washed limmu	0.99	0.98	0.99	119
washed sidamo	0.97	0.99	0.98	115
washed yirgachefe	0.98	0.97	0.98	112
accuracy			0.96	991
macro avg	0.96	0.96	0.96	991
weighted avg	0.97	0.96	0.96	991

(B)

Figure 10: Evaluation Result of the model on dataset 80-10-10 (A) confusion matrix (B) Classification report.

The confusion matrix on Figure 10 (A) shows that the model is confused especially on the coffee beans of *unwashed Nekemt* and *unwashed Sidamo*. The model misclassified 9 coffee beans of *unwashed Nekemt* and 6 coffee beans of *unwashed Sidamo* into another coffee bean class. For this reason, as compared to the accuracy (both recall and precision) registered for other classes, the recall of *unwashed Nekemt and* precision of *unwashed Sidamo* classes registered a lower accuracy (see Figure 10 (B)).

Experiment with Image Augmentation

In the following two experiments, image augmentation was introduced to prevent the over fitting. Image augmentation is a useful technique in CNN used to increase the size of the training set without acquiring new images. In this research, augmentation is applied to alter the coffee bean images with variations or rearrangement of the pixels so that the proposed CNN model can learn from more examples. However, image augmentation will be counterproductive if it produces images very dissimilar to what the model will be tested on.

The augmentation procedure is applied on coffee bean images on the fly during training of the model. The following four transformations are applied in the augmentation process:

a) Zoom coffee bean image randomly by 0.2.

b)Randomly shift the coffee bean images horizontally by a fraction of total width (horizontal_shift_range=0.2).

c) Randomly shift the coffee bean images vertically by a fraction of total height (vertical_shift_range=0.2).
d) Flip the coffee bean images horizontally.
The image augmentation is applied only on the training image dataset and not on the validation and test datasets.

Model Training: The model is trained for 30 epochs with Adam optimizer having a learning rate of 0.001, beta_1=0.9 and beta_2=0.999, and *categorical_crossentropy* loss function. A learning rate reducer and weight initialization mechanism proposed in the work of (Sudharsan, 2019) is also used.

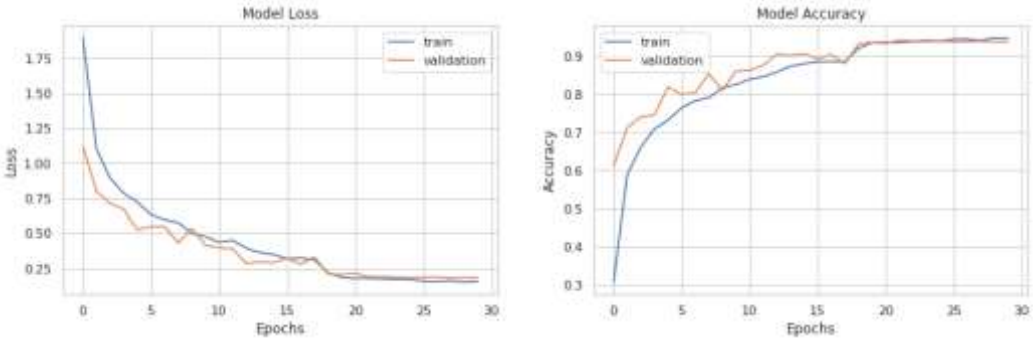
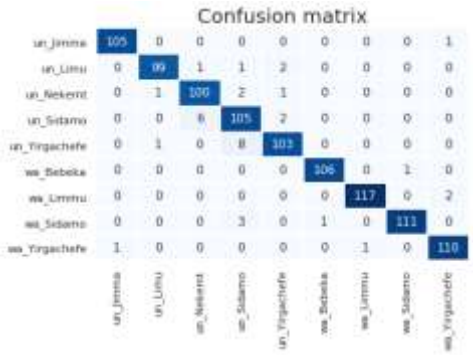


Figure 11: Training and validation curves on dataset_80-10-10 with augmentation.

The training and validation curves shown in Figure 11 does not have overfitting problem on the training dataset. The model does not show much improvement after the 20th epoch. The training accuracy and the validation accuracy are 94.56% and 93.56% respectively.

Model Evaluation: The performance of the model on the 10% of the original test dataset from 80:10:10 split is evaluated. An accuracy of 96.4% and 11.16% log-loss is registered.



(A)

	precision	recall	f1-score	support
unwashed jimma	0.99	0.99	0.99	106
unwashed limmu	0.98	0.96	0.97	103
unwashed nekemt	0.93	0.96	0.95	104
unwashed sidamo	0.88	0.93	0.91	113
unwashed yirgachefe	0.95	0.92	0.94	112
washed bebeka	0.99	0.99	0.99	107
washed limmu	0.99	0.98	0.99	119
washed sidamo	0.99	0.97	0.98	115
washed yirgachefe	0.97	0.98	0.98	112
accuracy			0.96	991
macro avg	0.97	0.96	0.96	991
weighted avg	0.97	0.96	0.96	991

(B)

Figure 12: Evaluation Result of the model on dataset 80-10-10 (A) confusion matrix (B) Classification report.

This model showed more confusion on *unwashed Yirgacheffe* and *unwashed Sidamo* coffee beans. It misclassified 9 *unwashed Yirgacheffe* coffee beans and 8 *unwashed Sidamo* coffee beans into another coffee bean class. Due to this, the recall of those classes is lower. On the other hand, the model performs well on *unwashed Jimma* coffee beans. The model only misclassified one coffee bean of *unwashed Jimma* as *unwashed Yirgacheffe*.

Experiment with Batch Normalization

Batch Normalization (BN) is introduced here to the proposed CNN model. The batch normalization layer is added after the convolutional layer followed by the ReLU activation function. Batch normalization is used to standardize the inputs to the layer in a deep neural network. The batch normalization layer transforms the inputs so that they are standardized. That is, they will have a mean of zero and a standard deviation of one.

CNN Architecture: A batch normalization layer is added to the CNN architecture. The ReLU

activation function is used for the convolution layer and the first two fully connected layers, SoftMax is for the output layer, and max-pooling is used for the pooling layer as shown in Figure 4.

***Model Training*:** The model is trained for 30 epochs on the training dataset split as 80:10:10. A batch gradient with a batch size of 64 is used for backpropagation. That means, in a single epoch the model is trained for 122 steps and validates for 16 steps (total image size of the training or validation divided by the batch size). The loss function *categorical_crossentropy* with Adam optimizer in used. The model registered a 100% training accuracy and a 99.18% validation accuracy.

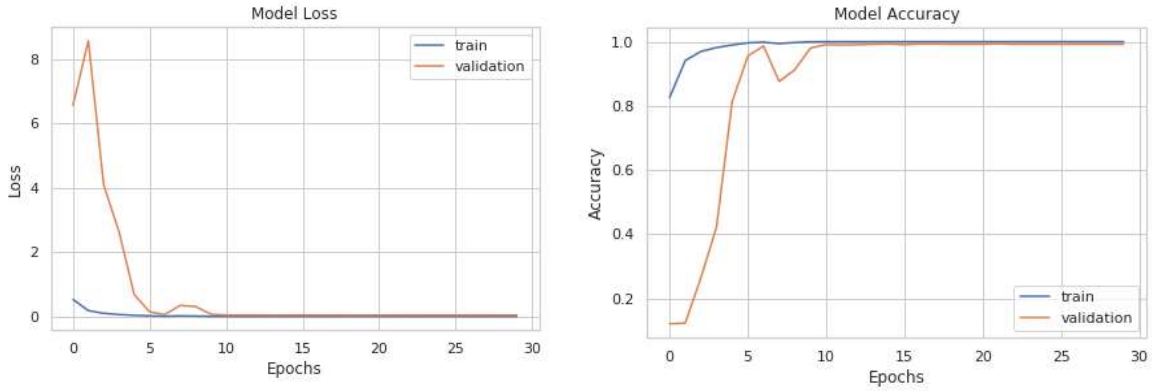


Figure 13: Training and validation curves of a CNN with BN.

As see from the model loss curves in Figure 13, the validation loss gets high at the beginning and tends to gets closer to the training loss after the 10thepoch. The model accuracy on the validation dataset was also low at the beginning of the training and get closer to the training accuracy after the 10th epoch. Even though the model is trained for 30 epochs, it did not show any improvement after the 13th epoch.

***Model Evaluation*:** The model performance is evaluated to see how well it classifies the coffee beans into their origin using the test dataset on both cases.

When the model trained on the coffee bean color images are evaluated using the coffee bean color image dataset, the model scored 99.89% accuracy and 0.92% log-loss.

(A)

	precision	recall	f1-score	support
unwashed jimma	1.00	1.00	1.00	106
unwashed limmu	1.00	1.00	1.00	103
unwashed nekemt	1.00	1.00	1.00	104
unwashed sidamo	1.00	1.00	1.00	113
unwashed yirgachefe	1.00	1.00	1.00	112
washed bebeka	1.00	0.99	1.00	107
washed limmu	1.00	1.00	1.00	119
washed sidamo	0.99	1.00	1.00	115
washed yirgachefe	1.00	1.00	1.00	112
accuracy			1.00	991
macro avg	1.00	1.00	1.00	991
weighted avg	1.00	1.00	1.00	991

(B)

Figure 14: evaluation result of CNN model with BN on rgb images (A) confusion matrix (B) classification report.

From the confusion matrix of this model (Figure 14 (A)), it can be seen that only one coffee bean from *washed Bebeka* is misclassified as *washed Sidama*. The classification report (Figure 14 (B)) shows that the precision of *washed Sidamo* coffee class (99%) is because of the one false-positive misclassification. The recall of the *washed*

Bebeka coffee bean class (99%) is also low because the one false-negative misclassification.

CONCLUSION

Coffee is an economically important crop, which plays a major role in earning foreign currency among other export commodities of Ethiopia. The quality of the coffee beans is evaluated by Ethiopian Coffee Quality and Inspection center (ECQIC) to give a standard certificate for export. The evaluation is currently done using traditional manual methods. In this research, an attempt is made to show the possibility of automating coffee quality inspection. To this end, a CNN model is developed to automatically classify 9 coffee bean classes namely: *unwashed Limmu, washed Limmu, washed Sidamo, unwashed Sidamo, washed Yirgacheffe, unwashed Yirgacheffe, unwashed Jimma, washed Bebeka, and unwashed Nekemte*. Three experiments were conducted on coffee bean image dataset split as 70:15:15 and 80:10:10 (for training, validation, and testing)by varying the hyper-parameters such as learning rate, momentum, batch size, number of epochs, and dropout rate.

The best model is found to be the one conducted using CNN with batch normalization and augmentation when the dataset is split as 80% for training, 10% for validation, and 10% for testing. The classification model registered a 99.89% classification accuracy with a log-loss of 0.92%. This confirms that CNN works well for Ethiopian coffee bean classification. A better result could be achieved if the model is trained on a large dataset which is recommended here as a way forward to come up with an applicable model for the industry.

REFERENCES

1. Alom, Z., Taha, T. M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M., Asari, a. V. (2019). A State-of-the-Art Survey on Deep Learning: Theory and Architectures. *Electronics, **8*(3),** 292.

2. Arboleda, E. R., Fajardo, A. C., & Medina, R. C. (2018). Recognition of Roasted Coffee Bean Levels using Image Processing and Neural Network and K-Nearest Neighbors. *IEEE Intl Conference Innov Res Dev.* IEEE.

3. Belay Birhanu & Tesfa Tegegn. (2015). Ethiopian Roated Cofee Classification Using Imaging Techniques. ICAST. Bahir Dar, Ethiopia.

4. Binyam Girma, Getachew Abebe, & Girma Goro. (2013). Characterization and Classification of Hararghe Coffee (Coffee Arabica L.) Beans Using Morphological Attributes Based on Classical Images. *East Africa Journal of Science, **7*(1),** 11-16.

5. Chellot, F. (2017). *Deep Learning with Python.* O'Reilly Media Inc.

6. Chemonics International Inc. (2010). *Ethiopian Coffee Industry Value Chain Analysis.* USAID.

7. Danakil, D. (2016). Ethiopian Coffee: Origin and History. *EtBUNA, 103780*, 1-4.

8. de Oliveira, E. M., Leme, D. S., G. Barbosa, B. H., Rodarte, M. P., Fonseca, R. G., & Pereira, A. (2016). A Computer Vision System for Coffee Beans Classification Based on Computational Intelligence Techniques. *Journal of Food Eng, **171***, 22-27.

9. Deffo, L. S., Tegne, E., & Tonye, E. (2018). CNNSFR: A Convolutional Neural Network System for Face Detection and Recognition. *Int Journal of Advanced Computer Science Applications, **9*(12**).

10. Espresso and Coffee Guide. (2020, March 30). *Ethiopian Coffee*. Retrieved from Ethiopian Coffee Beans: https://espressocoffeeguide.com/gourmet-coffee/arabian-and-african-coffees/ethiopian-coffee/

11. Foster, D. (2019). *Generative Deep Learning: Teaching Machines to Paint, Write, Compose, and Play.* O'Reilly Media Inc.

12. Gadzicki, K., Ashari, R. K., & Zetzsche, C. (2018). Multimodal Convolutional Neural Networks for Human Activity Recognition. *RSJ International Conference on Intelligent Robots and Systems.* Madrid, Spain.

13. Galvez, R. L., Bandala, A. A., Dadios, E. P., Vicerra, R. R., & Maningo, J. M. (2018). Object Detection Using Convolutional Neural Networks. *TENCON- IEEE Region 10 Conference,*, (pp. 2023–2027).

14. International Coffee Organization. (2018). *Grading and Classification of Coffee Grade.* FAO.

15. Kwolek, B. (2005). Face Detection Using Convolutional Neural Networks and Gabor filters. In W. Duch, J. Kacprzyk, E. Oja, & S. Zadrozny (Ed.), *Springer,.3696.* Berlin, Heidelberg: ICANN.

16. Habtamu Minassie. (2008). Image Analysis for Ethiopian Coffee Classification. *MSc Thesis*. Addis Ababa University: Addis Ababa.

17. Nasution, T., & Andayani, U. (2017). Recognition of Roasted Coffee Bean Levels using Image Processing and Neural Network. *IOP Conference Series Material Science and Engineering*.

18. Nomad Coffee Club. (2020, March 30). *Ethiopia's Coffee Farm Region – Nomad Coffee Club*. Retrieved from Ethiopian Coffee Farm Region-Nomad coffee Club: https://www.nomadcoffeeclub.com/pages/ethiopia

19. Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Manag Info System,* **24(3**), 45-77.

20. Peffers, K., Tuunannen, T., Gengler, C. E., Rossi, M., Hui, W., & Bragge, J. (2016). The Design Science Research Process: A Model for Producing and Presenting Information. *1st International Conference on Design Science Research in Information Systems and Technology*, (pp. 8-16). Claremont, California, USA.

21. Qu, D., Huang, Z., Gao, Z., Zhao, Y., Zhao, X., & Song, G. (2018). An Automatic System for Smile Recognition Based on CNN and Face Detection. *IEEE Int. Conf. Robot. Biomimetics*, (pp. 243-247).

22. Reddy, S. V., Reddy, K. T., & Kumari, V. V. (2018). Optimization of deep learning using various optimizers, loss functions and dropout. *Int. Journal of Innov. Technol. Explor. Eng., 8***(2),** 272-279.

23. Sang, D. V., Cuong, L. T., & Thuan, D. P. (2017). Facial smile detection using convolutional neural networks. *9th Int. Conf. Knowledge System Eng.*, (pp. 136-141).

24. Sudharsan, R. (2019). *Hands-On Deep Learning Algorithms with Python.* Packet Publishing.