

# Coffee Grading with Convolutional Neural Networks using Small Datasets with High Variance

Serawork Walleign<sup>1,2</sup>, Mihai Polceanu<sup>1,\*</sup>, Towfik Jemal<sup>2</sup>, Cédric Buche<sup>1</sup>

<sup>1</sup>LAB\_STICC, ENIB, CNRS, France

<sup>2</sup>JiT, Jimma University, Ethiopia

[walleign, polceanu, buche]@enib.fr, towfikjemal@yahoo.com

## ABSTRACT

Convolutional Neural Networks (CNNs) have been established as a powerful class of models for image recognition problems. Despite their success in other areas, CNNs have been applied only for very limited agricultural applications due to the need for large datasets. The aim of this research is to design a robust CNN model that classifies raw coffee beans into their 12 quality grades using small datasets which have high data variability. The dataset contains images of raw coffee beans acquired in two sets using different acquisition technique under varying illuminations which poses a complex challenge to designing a robust model. To design the model, preprocessing techniques were applied to the input in order to reduce task irrelevant features. But adding the preprocessing techniques did not improve the performance of the CNN model for our dataset. We have also used ensemble methods to solve the high variance that exists in networks when working with small datasets. Finally, we were able to design a model that classifies the beans into their quality grades with an accuracy of **89.01%** on the test dataset.

## Keywords

CNN, deep learning, ensemble methods, small dataset.

## 1 INTRODUCTION

Deep learning is making major advances in solving problems that have resisted the best attempts of the artificial intelligence community for many years. It has shown outstanding performance and is applied in many areas like business, science and government [LBH15]. As deep learning has been successfully applied in various domains, it has recently also entered the domain of agriculture. In particular, Convolutional Neural Network (CNN) has been applied for recognition and classification tasks in agricultural applications [FPF11] [RCC15]. However, a considerable barrier in the use of CNN is the need for large datasets, which would serve as the input during the training procedure. In spite of data augmentation techniques, which can enhance datasets with label-preserving transformations, in reality a dataset containing several

hundreds of images may often prove to be insufficient, depending on the complexity of the problem under study [KPB18]. In the domain of agriculture, only few publicly available datasets (eg. PlantVillage[HS<sup>+</sup>15], LifeCLEF[JMG<sup>+</sup>14], MalayaKew[LCWR15] and UC Merced[YN10]) exist for researchers to work with, and in many cases, researchers need to develop their own sets of images manually. The images are usually collected at multiple sites using different acquisition techniques under varying illuminations which poses a complex challenge to designing a robust model.

The fact that CNNs can learn important features from the raw data minimizes the need to use image processing techniques especially when enough data is available. The most commonly used preprocessing techniques are mean normalization, i.e, subtracting the mean of the dataset from each image, and scaling (limiting the dataset to be in the same range) the input before feeding to the network. This is important to speedup learning as well as increase the generalization ability of the network when the dataset contains enough images to represent different acquisition variations. However, this may not be enough when working with small datasets which have high data variability. We will investigate the effect of applying additional preprocessing on the performance of a CNN model when working with small dataset with high data variability.

Coffee is one of the most important globally traded agricultural commodities, with consumption occurring mostly in developed countries and production in devel-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

\* Current affiliation: University of Greenwich, UK

oping ones. It is one of the most important agriculture activities to earn foreign exchange in many tropical and subtropical countries [Ayi14] [FPF11]. The quality of produced coffee depends on the way it is cultivated and processed, which in turn determines its price on the market. Therefore, all produced coffee has to go through the grading process before it is introduced into the market.

Grading is the process of categorizing coffee beans on the basis of various criteria such as size of the bean, where and at what altitude it was grown, how it was prepared and picked, and how good it tastes [KJM<sup>+</sup>15]. There is no single agreed method of coffee grading among the top producing countries of the world. For example, countries like Indonesia base grading on the amount and type of defects present in the beans [FPF11], while other countries like Ethiopia use the physical characteristics (size, color, shape and uniformity of the beans), defects, presence of foreign objects as well as cup taste.

Despite the fact that coffee is an important cash crop, grading of coffee is performed manually. This method employs visual and manual methods of inspection exposing the quality assessment to inconsistent results and subjectivity. As a result, supporting the human operator systems with the consistent, non-destructive and cost effective automated system of coffee quality classification and determination is necessary for such commercial products that generate an important amount of income.

To the best of our knowledge, only [Ayi14], [TAG13], [FPF11] and [dOLB<sup>+</sup>16] applied computer vision to detect defective coffee beans, classify beans according to their geographic origin and grade the raw beans based defect systems respectively. Although research in computer vision has advanced in recent years, only image processing methods and traditional artificial neural networks were used. This limits the performance of the model to depend heavily on the underlining image processing and feature extracted methods used. However, deep learning techniques, which can automate the feature design step, are outperforming traditional, hand-crafted features based methods. In this study, modern methods to train CNNs are used to design a system that grades coffee using images of raw coffee beans. The images are collected in two sets where they differ in resolution, scale and imaging protocol used. We will also investigate how these variations affect the training of CNN. Our contribution is to automate the coffee grading process by designing a robust CNN model that classifies coffee beans into their quality grades from images of raw coffee beans. We were able to design a strong classifier from several weak models by applying ensemble methods. Though color correction pre-processing and augmentation methods were applied to

decrease task irrelevant features, they did not improve the performance of the model.

The rest the paper is organized as follows: section 2 discusses the manual grading of raw coffee beans; Section 3 deals with data collection and preparation; Section 4 describes the methodology used to design the model. The experimental results obtained are presented in section 5 and we conclude by recommending methods for future improvement.

## 2 MANUAL GRADING OF RAW COFFEE BEANS

The current practice of coffee grading is conducted manually based on a classification into four basic groups by type and market: washed beans for the domestic and international markets, and unwashed beans for each of these markets (Figure 1). Different criteria are used for each group to grade the beans. For example, the presence of beans covered by husk (refer to Figure 2c) is a criterion for unwashed categories but not for the washed categories. The experts (cuppers) assign scores for the raw coffee beans through one or a combination of two methods: Raw Quality and Cup Value (Cup Test).

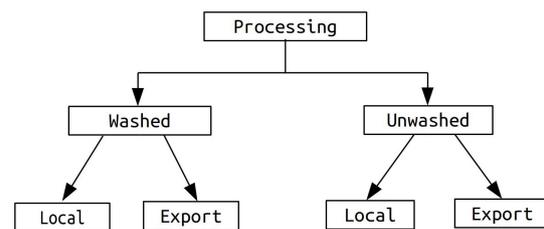


Figure 1: Grading categories based on way of processing and market. Experts use different criteria for each of the four categories



Figure 2: Some defects that affect the grade of coffee a) broken beans b) beans damaged by insects c) beans covered by husk d) unmaturing beans e) foreign objects f) unmaturing and broken

The raw quality is performed by examining the physical properties of the raw coffee beans. It entails characteristics like defects, shape and make, color and odor. The

cup value is determined by roasting and tasting sample coffee for acidity, cleanness and flavour. The final score is the sum of the raw value and cup value. The grade is then decided based on discretizing the score using intervals for each grade.

For each coffee shipment, a 3kg representative sample is taken from a 10-15 ton delivery. The sample is then coded and subjected to size screening and moisture tests. If it passes these tests, out of the 3kg, 300g is used for raw evaluation. The remaining sample is divided for roasting, future reference and client's display. The sample for raw evaluation is sorted manually by hand picking to separate each defect. A score is then given based on the ratio of defects to the entire sample or using the number of defective beans present. Some of the defects are shown in Figure 2. This accounts for 20% of the total score. Then each cupper works individually on each sample and records their opinion for every criteria. After every expert finished examining the sample, the final value of each attribute is decided by discussion among the experts. If there is a difference on the points, they will convince one another and elaborate their reasons to adjust their scores in order to reach a consensus. The final score will be given by an overall agreement of all the experts. The score is converted to the grade value based on in which range it lies. For example, coffee beans whose score is in the range 85 and above will be classified in grade-1 while beans with a score in the range 75 and 84 will be classified as Grade-2 coffee.

### 3 DATA COLLECTION AND PREPARATION

The images of coffee bean samples were collected at Jimma grading center in Ethiopia. The beans from the 300 gram sample used for raw evaluation were used to prepare the dataset. A small portion is taken from the sample, dispersed evenly in an A4 white paper then a picture is captured after which the beans are placed in a separate container. The process is repeated until the picture of all the beans in the sample have been taken. Capturing the images this way prevents the beans from overlapping and results in a sample image that is representative of the whole container.

The images were collected in two rounds nine months apart using different capturing devices. The first set of images was collected during the harvest season resulting in a dataset dominated by samples from the washed category. Whereas, the second set was collected when the beans from the previous year's harvest were graded which results in more samples from the unwashed category.

In the first round 1266 images of coffee beans from 12 grades with resolution of 2268x4032 pixels were captured using Samsung s7 edge camera (Figure 3). This

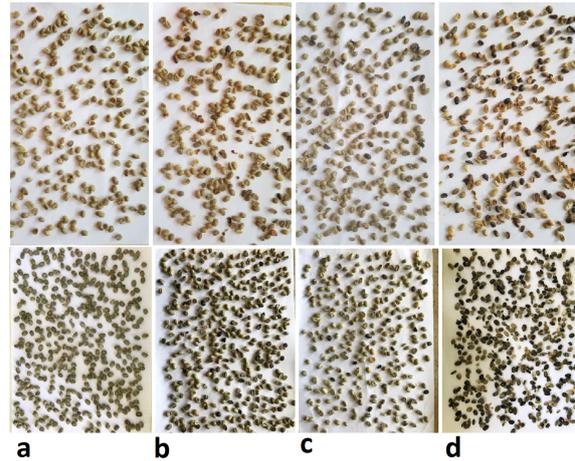


Figure 3: Sample images from the dataset. Top row: images collected in the first round. Bottom row: images collected in the second round. a) Grade-2 b) Grade-7 c) Local-1 d) Local-5A

set is mainly populated by samples from washed category (G1, G2 and G3). In the second round, 843 images with resolution of 3024x4032 pixels were captured by iPhone 7 plus camera dominated by samples from unwashed category (G6, G7). When trying to fit all the beans on the A4 paper, we were indirectly controlling the distance from the device to the beans. This created a scale difference between the two sets since the cameras have different resolution. There was no control on lighting while capturing the images. Table 1 shows the number of samples per grade collected for the two datasets. Since the collected images have high resolution, they are cropped into two along the longer side to increase the number of samples without losing important information.

Grade	Dataset1	Dataset2	Total
Grade-1	115	0	115
Grade-2	614	75	689
Grade-3	84	29	113
Grade-6	56	318	374
Grade-7	116	323	439
Grade-8	43	13	56
Local-1	14	16	30
Local-2	156	19	175
Local-3	30	11	41
Local-4	15	14	29
Local-5A	16	10	26
Local-5C	7	15	22
<b>Total</b>	<b>1266</b>	<b>843</b>	<b>2109</b>

Table 1: The number of samples per class in each dataset. Almost half of Dataset1 is from the washed category while about 76% Dataset2 is from unwashed category.

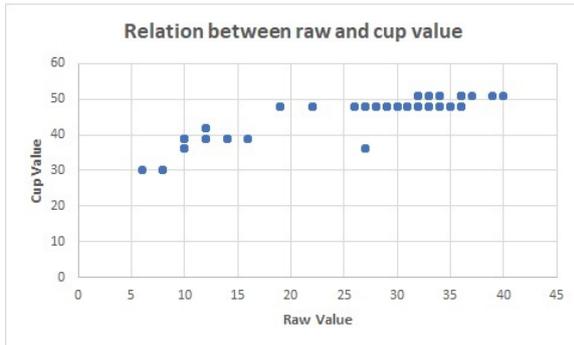


Figure 4: Correlation between the raw value and cup value

During the on-site survey at the grading organization, the experts stated that the physical appearance of the beans has an effect on the cup value since the presence of defects and foreign bodies affect the taste. To test the hypothesis, the correlation between the raw value and cup value of the sample data is calculated. The plot of the raw value versus the cup value is shown in Figure 4.

The correlation coefficient between the raw value and the cup value is 0.87 which indicates the existence of a strong relationship between the two values. Therefore, only considering the physical properties of the beans should be sufficient to obtain a model of the coffee grading process.

## 4 METHODOLOGY

This section discusses the steps followed and methods applied during the model design.

### 4.1 CNN architecture

Since we have a small dataset, the main focus during the design of the CNN architecture was to obtain a simple model with few trainable parameters. To achieve this goal, we followed two methods. First, to use transfer learning on existing trained model. It is common practice that features extracted by training a CNN in a fully supervised manner in large-scale object recognition datasets be reused for a novel generic task [DJV<sup>+</sup>14]. Therefore, we applied transfer learning on the pre-trained VGG[SZ14] model by removing the last fully connected layers. Second, to design a new architecture and compare the result with the one obtained from transfer learning.

The architecture of the new CNN model is adapted from the same model that we used for transfer learning except two modifications. Instead of using fully connected layer for the classifier, we used a 1x1 conv layer followed by global average pooling. It is stated that if the image area covered by units in the top most convolutional layer covers a portion of the image large

enough to recognize its content then fully connected layers can also be replaced by simple 1-by-1 convolutions [SDBR14]. This has a great advantage when working with small datasets by regularizing the network further since it results in a smaller number of model parameters than the fully connected layer. The second modification is, although using filters smaller than 5x5 pixels is recommended for use in the first convolutional layer, a 7x7 filter with stride=3 worked better for our dataset.

For the sub-sampling layer we used overlapping max-pooling with stride = 2 and kernel size = 3. For activation layers, using LeakyRelu (alpha = 0.001) resulted in relatively faster training and a more stable network than with the ReLU activation function. Therefore LeakyRelu is used after every convolutional layer except the last one where softmax activation is used to generate the probability of a sample being in each grade. The architecture of the designed model (referred to hereafter as CNN\_1) is summarized in Table 2.

Layer	Filters	Size, stride	Remark
Input		224x224	RGB
Conv1	96	7x7,3	LeakyReLU
Conv2	96	1x1,1	LeakyReLU
Conv3	192	3x3,1	LeakyReLU
Conv4	192	1x1,1	LeakyReLU
MaxPooling	192	3x3,2	
Conv5	384	3x3,1	LeakyReLU
Conv6	384	1x1,1	LeakyReLU
MaxPooling	384	3x3,2	0.5 dropout
Conv7	950	1x1,1	LeakyReLU
Conv8	12	1x1,1	12 outputs
GlobalAverage			
Softmax			

Table 2: Architecture of the model

### 4.2 Dataset variations

To see how the data collection method affects the model's generalization ability, we designed and trained a simple CNN model using only images from Dataset1. The model classifies the beans with an accuracy of 94.58%. But when it is tested using the images from Dataset2, the accuracy dropped to 8%. Though different image augmentation techniques were applied during the training process to cover the possible variations in the unseen data, the model fails to detect important features to classify beans correctly in the second set. The scatter plot (Figure 5) of the datasets shows that the difference in data distribution between them. This clearly shows that how the images were captured, camera differences, illumination conditions and scale introduced task irrelevant features to datasets. The data mismatch

problem can be addressed by applying preprocessing techniques to create similarity between the datasets and adding artificial synthesized data during the training to increase the model's generalization ability. In addition to another geometric augmentation techniques, random scaling is used to make the model robust to scale changes in the data. Color correction augmentation and preprocessing methods were also added to solve the mismatch because of the camera differences and lighting conditions. These techniques are discussed in the next section.

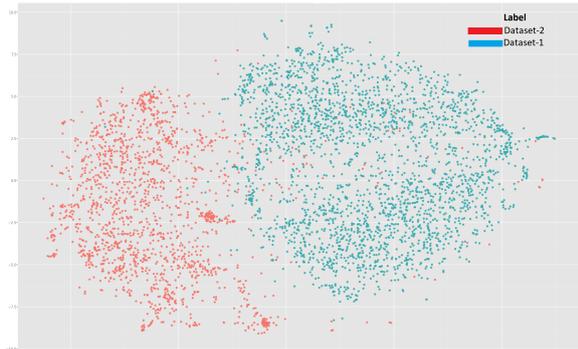


Figure 5: t-SNE plot of the distribution of the images in the datasets after feature reduction is performed using PCA. blue-Dataset1, red-Dataset2

### 4.3 Augmentation for color correction

The colors that are present in images are determined by the intrinsic properties of objects and surfaces as well as the illuminant of the scene [GGVDW11]. For a robust color-based system, these effects of the light source should be filtered out. This effect, i.e the ability to account for the color of illuminants, is known as color constancy. Many computer vision problems in both still images and videos can make use of color constancy processing as a preprocessing step to make sure that the recorded color of the objects in the scene does not change under different illumination conditions. But most often in deep learning models only intensity normalization is performed on the input images (VGG [SZ14], Inception [SLJ<sup>+</sup>15]). This may be enough when working with large datasets with millions of images but not for small datasets which lack in representation. Therefore, we applied a white balancing technique to preprocess the input images for color constancy and compare the result with intensity normalization. Gray-World[Buc80] color constancy technique is used. It is based on the assumption that the color in each sensor channel averages to gray over the entire image. We chose this algorithm because of its simplicity and fast computation time.

Gamma correction is another operation that digital cameras perform in order to match the human perception of an image. In digital cameras the received

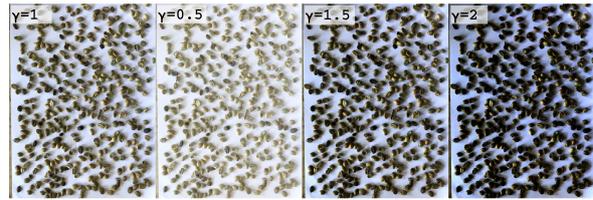


Figure 6: Sample gamma corrected images. Images appear darker as we increase the value of gamma. From left to right gamma = 1(no change), gamma = 0.5, gamma = 1.5 and gamma = 2.

signal from the sensors is linearly proportional to the light source that hits the sensor. To account for the non linearity of this process in humans, a non linear transformation function is applied. Sample gamma transformed images are shown in Figure 6. Images appear darker as the value for higher values of gamma. However this transformation is dependent on characteristics of the display and on the imaging device manufacturer. Therefore we added a gamma correction augmentation technique during the training of the model.

## 5 RESULTS

The models were implemented in Keras deep learning framework [C<sup>+</sup>15] on a single GTX 1070 GPU. The dataset is divided into three sets for training (70%), validation (10%) and testing (20%). All images were resized to 224x224 pixels, intensity values were scaled between 0 and 1. To compare the preprocessing techniques, two version of datasets were prepared. In one, let's call it DS1, images were normalized by subtracting the mean of the training set from each image. The second, DS2, was prepared using the white balancing color constancy technique. The experiment was conducted on both datasets DS1 and DS2. Each model is trained for 1000 epochs with early stopping, Adam optimization (learning rate = 0.0001) is used with categorical cross entropy loss function.

For the transfer learning, after removing the fully-connected layers of the VGG, two layers with 1024 nodes and 50% dropout were added followed by an output layer with 12 nodes equivalent to the number of grades. Then we trained only the new fully connected layers by keeping the weights in the other CNN layers the same as the one learned using the Imagenet dataset [DDS<sup>+</sup>09]. Neural networks are sensitive to initial conditions, both in terms of the initial random weights and in terms of the statistical noise in the training dataset. This stochastic nature of the learning algorithm means that each time a neural network model is trained, it may learn a slightly different version of the mapping function from inputs to outputs, that in turn will have different performance on the training and holdout datasets. Therefore, we trained the models

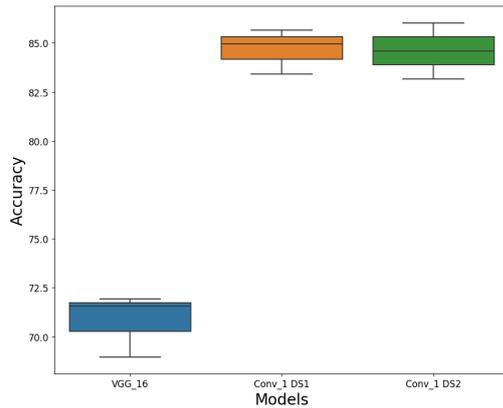


Figure 7: Performance of models for DS1 and DS2. There is high variance on test performance of the models when trained using different initial conditions

several times starting from random initial conditions. The performances of the models on DS1 and DS2 for different runs is shown in Figure 7. It can be observed from the plot that training the models starting from different initial conditions resulted on varying prediction performance making optimization and choosing the final model very challenging. We can see from the result VGG-16 has lower performance than CNN\_1 for this dataset. It was also experimented retraining the last two convolutional layers but the performance did not show much improvement compared to the CNN\_1.

A solution for high variance in networks, while also improving accuracy, is using ensemble learning [Die00]. Ensemble learning is training several models varying the initial conditions and then combining the predictions from each model to make one final prediction. We applied two types of ensemble methods, Checkpoint Smoothers (CS) and Checkpoint Ensemble (CE)[CLL17], to compare the performance between the models as well as to choose the final model. CS averages the weight of  $n$  models around the model with minimum validation (MV) loss while CE averages the predictions of the models. We picked the best performing models based on MV loss when trained with both datasets (models found at run-3 for DS1 and at run-2 for DS2) and applied CS and CE for  $n = 10$ ,  $n = 20$  and  $n = 50$ . The test accuracy of the models for different values of  $n$  is summarized in Table 3.

As it can be seen from the table, applying color correction preprocessing and augmentation did not bring any improvement on the performance of the model. Also using ensemble methods results in better prediction performance than the model selected based on minimum validation loss. Both the CE and CS models have comparable results. We chose the model designed using CS since it takes less prediction time than CE with equal

	CNN_1 (DS1) %	CNN_1 (DS2) %	Remark
run-1	84.95	84.25	MV
run-2	85.66	<b>86.02</b>	MV
run-3	<b>85.78</b>	85.07	MV
$n = 10$	88.40	87.68	CE
	88.39	87.80	CS
$n = 20$	87.80	87.20	CE
	88.51	87.68	CS
$n = 50$	88.51	87.80	CE
	<b>89.1</b>	<b>88.34</b>	CS

Table 3: Summary of the experimental results. The model's prediction performance is better when the dataset is preprocessed using only mean normalization.

number of models. Finally, the CS ensemble model, for  $n = 50$ , trained using DS1 is selected as a final model.

The Top-2 accuracy of the selected model is 98.22%. Also the confusion matrix (Figure 8) of this model on the test dataset shows most of the misclassifications occur between two neighbouring grade values. This maybe due to the fact that in the manual classification the grade value is inferred based on which range the final score lies, i.e for example if the score lies in the range 75 to 84, it will be a Grade-2 coffee and if it lies between 63 and 74 then its grade will be Grade-3. Therefore, two coffee beans with a total score of 75 and 74 will be classified in Grade-2 and Grade-3 respectively even though there is no significant difference in their appearances.

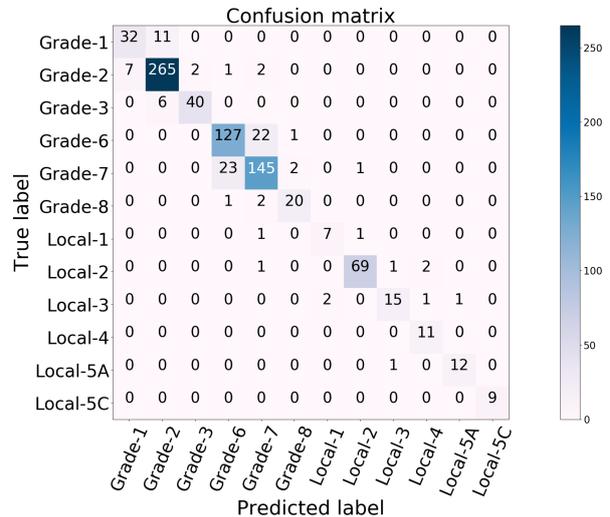


Figure 8: Confusion matrix of the model

## 6 CONCLUSIONS

In this work we designed a CNN model to classify raw coffee beans into 12 quality grades and achieved 89.1% on the test dataset. It was also shown that the CNN

model can extract features and classify the input images with minimal preprocessing. White balancing the images and adding color correction to the augmentation function did not improve the performance of the model. We have also implemented ensemble methods to design a strong model from several weak-classifiers. In the error analysis, we have seen that most misclassifications occur between two neighbour classes. In the future we will investigate the effect that treating the output as a classification problem rather than a regression problem had on the performance of the model.

During our stay in the grading organization, we were informed that the experts estimate their repeatability between 80 - 85%. Since the model was inferred based on the labels given by the experts, its prediction performance is affected by this error. We did not use any method to address wrong labels due to human errors. In the future, we will investigate using unsupervised or semi-supervised techniques to decrease the effect of noise present due to labelling.

## 7 ACKNOWLEDGMENTS

This research is funded by the French Embassy and the Ethiopian Ministry of Science and Higher Education (MoSHE) through the higher education capacity building program.

## 8 REFERENCES

- [Ayi14] Betelihem Mesfin Ayitenfsu. Method of coffee bean defect detection. *International Journal of Engineering Research & Technology*, 3:2355–57, 2014.
- [Buc80] Gershon Buchsbaum. A spatial processor model for object colour perception. *Journal of the Franklin institute*, 310(1):1–26, 1980.
- [C+15] François Chollet et al. Keras, 2015.
- [CLL17] Hugh Chen, Scott Lundberg, and Su-In Lee. Checkpoint ensembles: Ensemble methods from a single training process. *arXiv preprint arXiv:1710.03282*, 2017.
- [DDS+09] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255. IEEE, 2009.
- [Die00] Thomas G Dietterich. Ensemble methods in machine learning. In *International workshop on multiple classifier systems*, pages 1–15. Springer, 2000.
- [DJV+14] Jeff Donahue, Yangqing Jia, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng, and Trevor Darrell. De-caf: A deep convolutional activation feature for generic visual recognition. In *International conference on machine learning*, pages 647–655, 2014.
- [dOLB+16] Emanuelle Morais de Oliveira, Dimas Samid Leme, Bruno Henrique Groenner Barbosa, Mirian Pereira Rodarte, and Rosemary Gualberto Fonseca Alvarenga Pereira. A computer vision system for coffee beans classification based on computational intelligence techniques. *Journal of Food Engineering*, 171:22–27, 2016.
- [FPF11] Faridah Faridah, Gea OF Parikesit, and Ferdiansjah Ferdiansjah. Coffee bean grade determination based on image parameter. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, 9(3):547–554, 2011.
- [GGVDW11] Arjan Gijsenij, Theo Gevers, and Joost Van De Weijer. Computational color constancy: Survey and experiments. *IEEE Transactions on Image Processing*, 20(9):2475–2489, 2011.
- [HS+15] David Hughes, Marcel Salathé, et al. An open access repository of images on plant health to enable the development of mobile disease diagnostics. *arXiv preprint arXiv:1511.08060*, 2015.
- [JMG+14] Alexis Joly, Henning Müller, Hervé Goëau, Hervé Glotin, Concetto Spampinato, Andreas Rauber, Pierre Bonnet, Willem-Pier Vellinga, Robert B Fisher, and Robert Planquè. Lifeclef: Multimedia life species identification. In *EMR@ICMR*, pages 7–13, 2014.
- [KJM+15] Dae-Joong Kwon, Hee-Jeong Jeong, Hyeyoung Moon, Hak-Nam Kim, Jee-Hyun Cho, Jang-Eun Lee, Kwan Soo Hong, and Young-Shick Hong. Assessment of green coffee bean metabolites dependent on coffee quality using a 1h nmr-based metabolomics approach. *Food Research International*, 67:175–182, 2015.
- [KPB18] A Kamilaris and FX Prenafeta-Boldú. A review of the use of convolutional neural networks in agriculture. *The Journal of Agricultural Science*, 156(3):312–322, 2018.
- [LBH15] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436, 2015.

- [LCWR15] Sue Han Lee, Chee Seng Chan, Paul Wilkin, and Paolo Remagnino. Deep-plant: Plant identification with convolutional neural networks. In *Image Processing (ICIP), 2015 IEEE International Conference on*, pages 452–456. IEEE, 2015.
- [RCC15] Angie K Reyes, Juan C Caicedo, and Jorge E Camargo. Fine-tuning deep convolutional networks for plant recognition. In *CLEF (Working Notes)*, 2015.
- [SDBR14] Jost Tobias Springenberg, Alexey Dosovitskiy, Thomas Brox, and Martin Riedmiller. Striving for simplicity: The all convolutional net. *arXiv preprint arXiv:1412.6806*, 2014.
- [SLJ<sup>+</sup>15] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.
- [SZ14] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [TAG13] Birhanu Turi, Getachew Abebe, and Girma Goro. Classification of ethiopian coffee beans using imaging techniques. *East African Journal of Sciences*, 7(1):1–10, 2013.
- [YN10] Yi Yang and Shawn Newsam. Bag-of-visual-words and spatial extensions for land-use classification. In *Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems*, pages 270–279. ACM, 2010.