18th IMEKO TC10 Conference
"Measurement for Diagnostics, Optimisation and Control to Support Sustainability and Resilience"
Warsaw, Poland, September 26–27, 2022

Classification of coffee bean varieties based on a deep learning approach

D. Buonocore, M. Carratù, M. Lamberti

University of Salerno, Via Giovanni Paolo II, 132 Fisciano (SA)
{dbuonocore,mcarratu}@unisa.it, m.lamberti57@studenti.unisa.it

***Abstract* – In this article, a coffee beans fraud detection based on a deep learning approach is proposed, which has been achieved after classifying the two coffee varieties to distinguish them in a real-time industrial scenario. The coffee bean quality is typically defined by visual inspection, which is subjective, needing significant effort and time, and susceptible to fault detection. For these reasons, a different method is required to be objective and precise. Therefore, object detection techniques were employed to automatically classify the coffee bean samples according to their specie using an own dataset consisting of over 2500 coffee beans. Furthermore, a convolutional neural network (CNN) based on the YOLO algorithm was employed to categorize the coffee beans automatically. The result of this study has revealed that the object detection technique could be used as an effective method to classify coffee bean species and discover food fraud.**

Keywords – Deep Learning, Food Quality, Coffee Beans, Fraud Detection, Classification, YOLOv5, Object Detection, Industry Innovation and Infrastructure, SDG 9.

I. INTRODUCTION

Coffee is a beverage obtained by roasting and grinding the seeds of certain species of tropical trees, *Coffea*, belonging to the botanical family of *Rubiaceae*.

In the world, more than hundreds of species of Coffea exist, however, there are about ten commercially useful ones, which differ in some characteristics, such as plant height, flower fragrance, caffeine content, and size and color of the seeds. Since the most widespread and cultivated species are '*Coffea Arabica*' and '*Coffea Canephora*', better known as *'Robusta'*, the following study focused on the classification and the extraction of the major features of the mentioned species.

As the second most popular beverage behind water, the industry of coffee has generated a significant economy of its own, and, as often happens when the market is relatively large, the coffee industry is not free from misdeeds towards consumers who use blending the two species. In fact, Arabica is harder to grow and is generally considered higher quality; Robusta, on the other hand, is lower quality, though it is easier to grow. The economic benefits of selling Arabica versus Robusta have favoured coffee fraud. In recent years, European Union and other regions of the world are the stage for serious cases of fraud concerning Arabica coffee [1]. For these reasons, nowadays coffee identification and classification have become important issues since its variety is a key concerned factor in coffee trading and consumption. According to that, Figure 1 shows one of the possible use cases of the proposed classification method: starting from a mixture of Robusta and Arabica varieties, each coffee bean is detected and recognised with the respective anchor box (red box for Arabica coffee and pink box for Robusta coffee).

Figure 1. Mixture of Robusta and Arabica coffee beans.

The wording "100% Arabica" is not a coffee quality assurance, it is simply stating that the contents are Arabica beans as opposed to Robusta beans. In this regard, in 2018 the Quadram Institute conducted a study [2] whose results confirm that a consistent percentage (5-20%) of pure Arabica coffee contains substantial traces of Robusta. However, identifying Arabica versus Robusta coffee is no easy task. Arabica beans are oval, have a pronounced center crease, and are generally larger, while Robusta ones are paler, center crease is less pronounced, more circular, and generally smaller, but, once the beans have been roasted and ground up, human visual inspections and computer vision are often useless worsening the detection probability.

Several analytical techniques can be used to discern the differences between Arabica and Robusta, including, machine learning approaches [3], molecular genetics approaches [4], [5], nuclear magnetic resonance spectroscopy [6]-[8], liquid- and gas chromatography-mass spectrometry [9], [10] among others. Also, laser-induced breakdown spectroscopy (LIBS) [11], Raman spectroscopy [12], and vibrational techniques like mid-infrared (MIRS) [13] and near-infrared spectroscopies (NIRS) [14] can provide direct, non-destructive, and rapid measurements. Although the approaches mentioned above are effective and mature for the coffee quality assessment, they are expensive or complex due to the expensive equipment and complex measurement process. In addition, even if all these procedures can offer high sensitivity and thus be appropriate for a confirmatory examination of suspect fraudulent beans, the application of these methods would not be reasonable for the quick screening and checking of coffee authenticity.

Moreover, coffee roasters and coffee distributors often blend different varieties of coffee beans to create a specific and consistent flavor profile. These blends are formulated from different specific varieties of beans, and suppliers compete to create new and distinct blends. As a result, the batches of beans come from different suppliers, and their identity must be confirmed before they can be inventoried. Including the wrong grain-type can change the final product's flavor profile and lead to consumer dissatisfaction. Each incoming batch or lot for blending must also be checked for physical contaminants such as plant material, pebbles, and other materials.

In this article, a coffee beans fraud detection based on a deep learning approach is proposed, which has been achieved after classifying the two coffee species to distinguish them in a real-time industrial scenario.

II. DESCRIPTION OF THE METHOD

To describe the methodologies and the setup, it is first necessary to summarize the process:

- Investigating on morphological characteristics of varieties such as shape, size, and color of the coffee beans
- Extracting features from the image, using a neural network that suits to a classification of different varieties
- Measure the performance of the neural network created.

To acquire the pictures, the "DFK 23G445 - GigE" camera was used, with 1.2 MP, equipped with Sony ICX445AQA, a diagonal 6.0mm (Type 1/3) interline CCD sensor. The camera has been mounted on a fixed stand that ensures stable support, preventing movement in any direction. The camera was placed at a distance of 340 mm from the table to get images clearly.

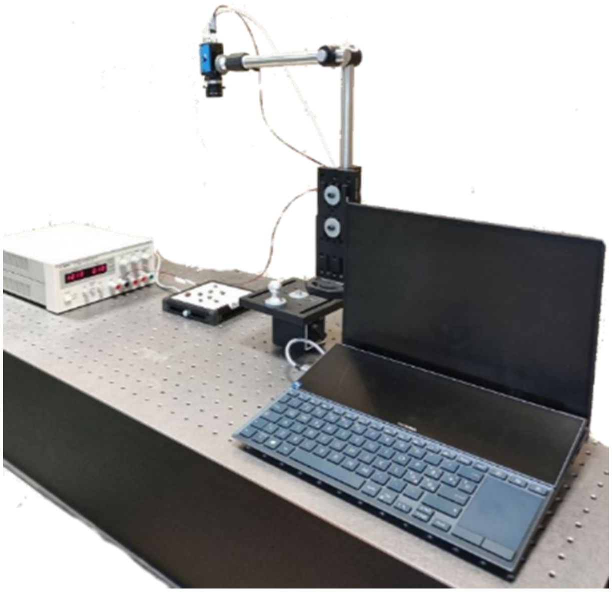

Figure 2. Measurement setup.

The coffee beans were positioned on a whiteboard to remove the background during the analysis easily. The coffee beans were tidily arranged on the plane so they did not touch each other to make preliminary segmentation easier. To obtain uniform lightning, it was used a 100 W incandescent lamp. The images were taken at 1280x960 pixels; the camera was mastered using a MATLAB script running on a laptop, that also provides to separate and crop the single coffee beans from the complete pictures. The measurement setup is shown in the Figure 2.

To create the dataset, four different 100 grams of coffee beans were used below listed:

- Ethiopia Sidamo, Arabica (Heirloom), cultivated at an altitude of 1900/2000 m
- Guatemala Shb, Arabica (Caturra), cultivated at an altitude of over 1300 m
- Tanzania Superior, Robusta, cultivated at an altitude of 1200/1500 m
- Indonesia Flores, Robusta, cultivated at an altitude of 800/1200 m.

Once created, the dataset contains 144 complete pictures and more than 2500 coffee beans. The 80% of this dataset was used for training and the residual 20% for validation. Image classification, used to extract features from beans, goes through incremental levels of complexity. The objectives are to assign each object to its own category, draw bounding boxes and calculate the probability of error. YOLO (You Only Look Once) and SSD (Single Shot Multibox Detector) family algorithms are the two best-known examples of object detection methods. For the purpose, a YOLOv5 [15] algorithm was used, employing convolutional neural networks (CNN) to detect objects in real-time. YOLO was chosen for its speed, high accuracy and learning capabilities [16], [17]. The algorithm splits each image (size 960 x 960) into cells using a 30 x 30 grid, where each cell is responsible for the predicting of 3 bounding boxes, as in Figure 3. In order to recognize objects faster and to ensure real-time recognition, the accuracy is reduced.

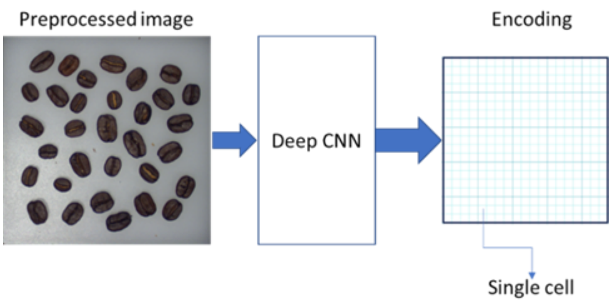

Figure 3. Image size reduction.

Each of the grid cells in the image is matched with the class with the highest probability. Then, the next step is non-max suppression, which allows the algorithm to discard unnecessary anchor boxes. Once selection the highest-class probability, YOLO determines the Intersection over Union, IoU, for all the bounding boxes, given in Eq. 1:

$$IoU = \frac{B1 \cap B2}{B1 \cup B2} \qquad \textit{Eq. 1}$$

Class confidence scores for each bounding box respond to:

$$\Pr(Class_i) * IoU_{pred}^{truth} \qquad \textit{Eq. 2}$$

Eq. 2 is given by multiplying the conditional class probabilities by the confidence predictions of the individual boxes [18].

YOLOv5 is a single-stage object detector, the main part of which consists of:

- model Backbone
- model Neck
- model Head

To extract important features from the pictures, in YOLOv5 Cross Stage Partial Networks (CSP) was used as a model Backbone [19]. The model Neck is used to generate feature pyramids to help the network recognize the same object with different sizes and scales; YOLOv5 uses PANet for this purpose [20]. In YOLO, Feature Pyramid Network (FPN) is a feature extractor projected for accuracy and speed, replacing the feature extractor of detectors, for example Faster R-CNN, generating a multu-scale feature maps with an improvement in quality of information than the normal feature pyramid.

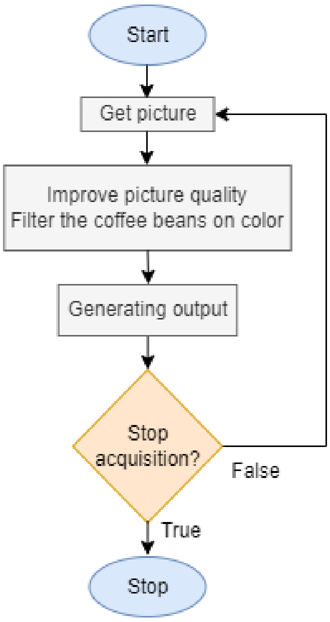

Figure 4. MATLAB algorithm

The last detection part uses the model Head, which generates final vectors with class probabilities, confidence scores and bounding boxes to show where the objects are located, YOLOv5 uses the same model Head as the previous YOLO version.
YOLOv5 authors [15] decided to use the Leaky ReLU and Sigmoid activation function, which is the most crucial choice of deep neural networks.
Two kinds of optimization functions could be used in YOLOv5:

- SGD [21]
- Adam [22]

By default, the optimization function for training is SGD. The SGD is an optimizer which is a variant of gradient descent, it does not compute on the whole dataset but only on a subset, producing a good performance as a gradient descent. Adam, instead, is an algorithm for gradient-based optimization of stochastic objective functions, which combines two SGD extensions (RMSProp and AdaGrad). It was decided to use SGD despite the popularity of Adam because it sometimes fails to converge on an optimal solution.

The training of YOLO model was made using Nvidia RTX 8000 with a memory size of 48 GB and Compute

Unified Device Architecture (CUDA) graphic features.
The procedure consists of the following steps:

- Acquiring pictures using MATLAB script
- Elaborating the picture to split and crop the coffee beans
- Building the dataset suitable to YOLO algorithm
- Training neural networks using Python
- Using Python to validate and to measure the network performances.

Figure 5. The outcome of YOLO detection

III. RESULTS AND DISCUSSIONS

Images were acquired in accordance with the approach described in the previous section; using appropriate processing techniques, grains with a correct detection rate of 98% were extracted.
To build the dataset it was used MATLAB; the flow chart of the algorithm was reported in Figure 4.
Then, a convolutional neural network was trained with the YOLO algorithm using Python programming language. The algorithm consists of three steps: training, validation and testing.

The training was made using a PC with a GPU using CUDA, as described above, on 250 epochs and the results are described below.
The metrics to be taken into account are:

$$Precision = \frac{TP}{(TP + FP)} \quad \text{Eq. 3}$$

$$Recall = \frac{TP}{(TP + FN)} \quad \text{Eq. 4}$$

where TP stands for *true positives* (good prediction), FN means *false negatives* (failure to predict an object) and FP is related to *false positives* (incorrect positive prediction). Precision indicates what proportion of positive identifications is actually correct. Recall, also known as sensitivity, instead, indicates what proportion of actual positives is identified correctly. The area under the precision-recall curve is the general definition for the mean Average Precision (mAP). The best coefficients in terms of precision, recall and mAP, obtained from the validation test, are reported in *Table 1*. In particular, for the mean Average Precision two values are shown: mAP@0.5 means the mAP with IoU threshold equals to 0.50, whereas mAP@0.5-0.95 corresponds to the average mAP for IoU from 0.5 to 0.95 with a step size of 0.05 [23].

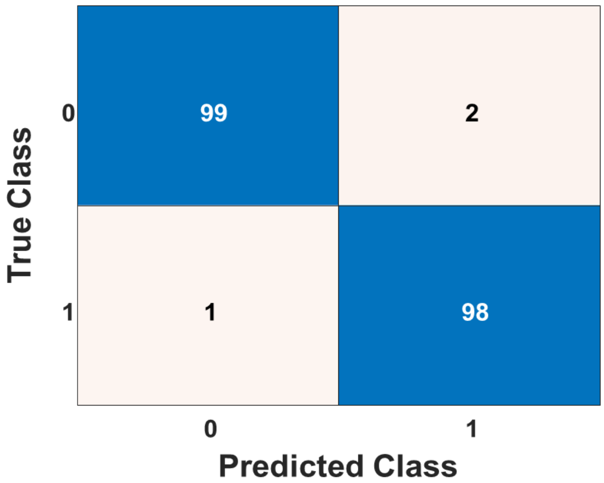

Figure 6. Confusion matrix

The testing step was evaluated on photos not in the dataset, whose variety arrangement is known (odd rows Robusta, even rows Arabica). The image returned from the network contains information regarding the individual bean (label and confidence interval). Label '0' indicates Arabica quality, whereas label '1' indicates Robusta. Figure 6 shows an example of network results in which grain detection and class assignment with high confidence intervals are appreciated.
The results of the trained network are summarized in the confusion matrix shown in Figure 5.As can be seen from the matrix, the trained network correctly identifies the two coffee varieties minimizing the false positive and true negative, according to the optimal coefficients from the validation in *Table 1*.

Table 1. Optimal coefficients from the validation.

Precision	Recall	mAP@0.5	mAP@0.5-0.95
0.99	0.99	0.99	0.94

Table 2 contains the results, in terms of error percentage, of tests performed on 100 grains for each variety described above.

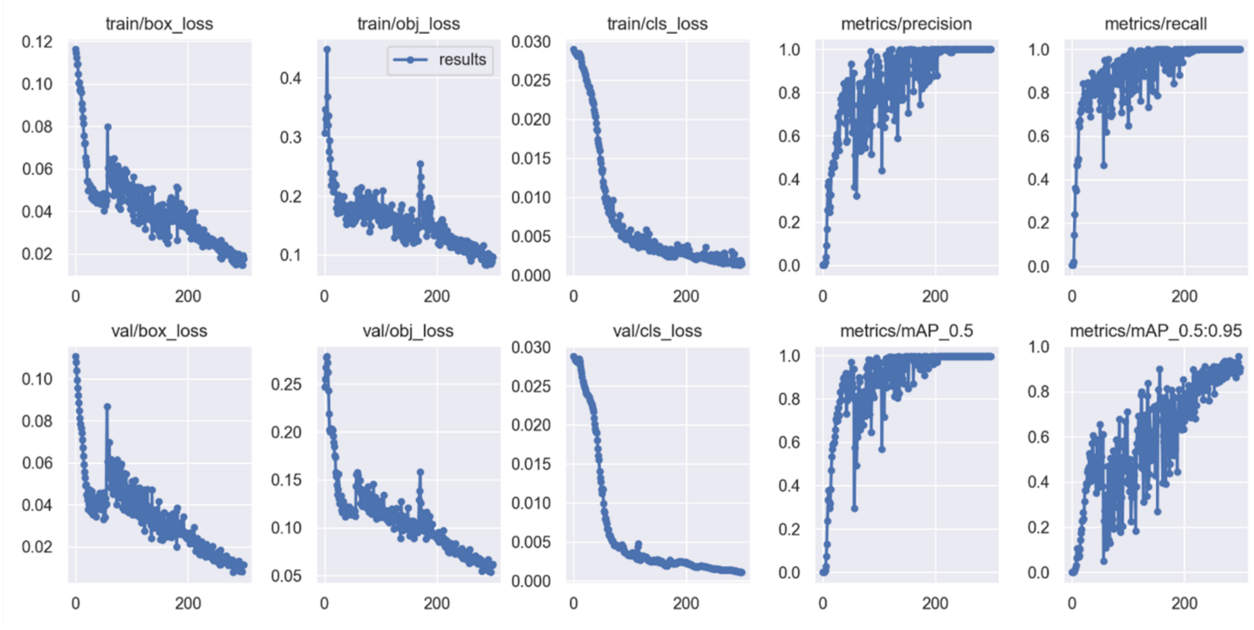

Figure 7: The progress in the performance of the proposed model.

Table 2. Analysis of different varieties.

		Grains	Error [%]	Mean Score [%]
Arabica	Ethiopia Sidamo	100	5	93
	Guatemala Shb	100	4	92
Robusta	Tanzania Superior	100	8	91
	Indonesia Flores	100	3	94

Figure 7 shows the graphs of performance characteristics during training and validation in terms of losses and metrics. YOLO loss function is composed of three parts: box_loss (bounding box regression loss), obj_loss (objectness loss) and cls_loss (classification loss). In this figure, the images in the top row refer to the performance of the model for the training set, while the images in the bottom row refer to the performance of the model using the validation set. From these images, it can be seen that the loss in the detection of coffee beans in the training set reached the minimum value after approximately 250 epochs. The loss in the detection of coffee beans in the validation set had about the same behaviour. The number of epochs was chosen as a compromise between training time and the best data results before overfitting.

Moreover, a 5-fold cross validation is used to measure the performance of the machine learning model, in order to estimate how the model is expected to perform on data not used during the training of the model. This approach involves randomly dividing the set of observations into 5 groups, or folds, of approximately equal size. Since the 80% of images are included in the training set and the remaining part in the validation one, this means that 1 fold is treated as a validation set whereas the method is trained on the remaining 4 folds. This procedure is consists of considering the following 5 configurations given as:

- Round 1: trained on folds 1-2-3-4, validated on fold 5
- Round 2: trained on folds 1-2-3-5, validated on fold 4
- Round 3: trained on folds 1-2-4-5, validated on fold 3
- Round 4: trained on folds 1-3-4-5, validated on fold 2
- Round 5: trained on folds 2-3-4-5, validated on fold 1

Table 3. Maximum values of mAP@0.5-0.95 for each round of the 5-fold cross validation method.

Round 1	Round 2	Round 3	Round 4	Round 5
0.929	0.943	0.947	0.932	0.950

The maximum values of mAP@0.5-0.95 obtained from each round of the 5-fold cross validation method are summarized in Table 3. Starting from these values, statistic parameters, such as mean ($\mu = 0.940$) and variance ($\sigma^2 = 8.570 \times 10^{-5}$) , are calculated. Based on these results, it is possible to assert that the model performances are promising for the recognition of coffee beans varieties.

IV. CONCLUSIONS

This paper introduces a convolutional neural network (CNN) approach for coffee bean detection and classification. The algorithm can automatically classify the species of coffee via the bean images. The experimental result indicates that the YOLO algorithm can correctly distinguish coffee species.

The proposed method can be improved by expanding the coffee dataset, both in terms of the number of bean photos for each variety and by introducing other varieties, so that the network has more sources from which to extract characteristics and features to refine the predictive capability. As mentioned in the introduction, once the network has been trained with a high degree of confidence, further perspectives could be engineering, not only in industrial applications, but also in the private sector, in order to detect fraud and guarantee the quality of coffee to consumers.

In conclusion, the results of this study have revealed that the object detection technique could be used as an effective method to classify coffee bean species and fight food fraud.

REFERENCES

[1] Europol - INTERPOL, 2019 https://www.europol.europa.eu/media-press/newsroom/news/over-%E2%82%AC100-million-worth-of-fake-food-and-drinks-seized-in-latest-europol-interpol-operation

[2] https://quadram.ac.uk/coffee-adulteration-uncovered-using-new-method/

[3] **Capriglione, D., Carratù, M., Sommella, P., & Pietrosanto, A.** (2017). ANN-based IFD in motorcycle rear suspension. Paper presented at the 15th IMEKO TC10 Workshop on Technical Diagnostics 2017 - "Technical Diagnostics in Cyber-Physical Era", 22-27.

[4] **Combes, M. C., Joët, T., & Lashermes, P.** (2018). Development of a rapid and efficient DNA-based method to detect and quantify adulterations in coffee (Arabica versus Robusta). Food Control, 88, 198–206. https://doi.org/10.1016/j.foodcont.2018.01.014

[5] **Spaniolas, S., May, S. T., Bennett, M. J., & Tucker, G. A.** (2006). Authentication of coffee by means of PCR-RFLP analysis and lab-on-a-chip capillary electrophoresis. Journal of Agricultural and Food Chemistry, 54(20), 7466–7470. https://doi.org/10.1021/jf061164n

[6] **Cagliani, L. R., Pellegrino, G., Giugno, G., & Consonni, R.** (2013). Quantification of Coffea arabica and Coffea canephora var. robusta in roasted and ground coffee blends. Talanta, 106, 169–173. https://doi.org/10.1016/j.talanta.2012.12.003

[7] **Defernez, M., Wren, E., Watson, A. D., Gunning, Y., Colquhoun, I. J., Le Gall, G., Kemsley, E. K.** (2017). Low-field 1H NMR spectroscopy for distinguishing between arabica and robusta ground roast coffees. Food Chemistry, 216, 106–113. https://doi.org/10.1016/j.foodchem.2016.08.028

[8] **Monakhova, Y. B., Ruge, W., Kuballa, T., Ilse, M., Winkelmann, O., Diehl, B., Lachenmeier, D. W.** (2015). Rapid approach to identify the presence of Arabica and Robusta species in coffee using 1H NMR spectroscopy. Food Chemistry, 182, 178–184. https://doi.org/10.1016/j.foodchem.2015.02.132

[9] **Garrett, R., Vaz, B. G., Hovell, A. M. C., Eberlin, M. N., & Rezende, C. M.** (2012). Arabica and robusta coffees: Identification of major polar compounds and quantification of blends by direct-infusion electrospray ionization–mass spectrometry. Journal of Agricultural and Food Chemistry, 60(17), 4253–4258. https://doi.org/10.1021/jf300388m

[10] **Procida, G., Lagazio, C., Cateni, F., Zacchigna, M., & Cichelli, A.** (2020). Characterization of Arabica and Robusta volatile coffees composition by reverse carrier gas headspace gas chromatography–mass spectrometry based on a statistical approach. Food Science and Biotechnology, 29(10), 1319–1330. https://doi.org/10.1007/s10068-020-00779-7

[11] **Liu, F.; Ye, L.; Peng, J.; Song, K.; Shen, T.; Zhang, C.; He, Y.**: *Fast Detection of Copper Content in Rice by Laser-Induced Breakdown Spectroscopy with Uni- and Multivariate Analysis*. Sensors 2018, 18, 705. https://doi.org/10.3390/s18030705

[12] **A.S. Luna, A.P. da Silva, C.S. da Silva, I.C.A. Lima, J.S. de Gois**, *Chemometric methods for classification of clonal varieties of green coffee using Raman spectroscopy and direct sample analysis*, Journal of Food Composition and Analysis, 76 (2019), pp. 44-50, 10.1016/j.jfca.2018.12.001

[13] **Wang, J., Jun, S., Bittenbender, H., Gautz, L. and Li, Q.X.** (2009*), Fourier Transform Infrared Spectroscopy for Kona Coffee Authentication*. Journal of Food Science, 74: C385-C391. https://doi.org/10.1111/j.1750-3841.2009.01173.x

[14] **A. Giraudo, S. Grassi, F. Savorani, G. Gavoci, E. Casiraghi, F. Geobaldo**, *Determination of the geographical origin of green coffee beans using NIR spectroscopy and multivariate data analysis,* Food Control, Volume 99, 2019, Pages 137-145, ISSN 0956-7135, https://doi.org/10.1016/j.foodcont.2018.12.033.

[15] **Ultralytics,** *YOLOv5***,** https://github.com/ultralytics/yolov5

[16] **Carratu, M., Gallo, V., Liguori, C., Pietrosanto, A., O'Nils, M., & Lundgren, J.** (2021). A CNN-based approach to measure wood quality in timber bundle images. Paper presented at the Conference Record - IEEE Instrumentation and Measurement Technology Conference, , 2021-May doi:10.1109/I2MTC50364.2021.9459906

[17] **Carratu, M., Gallo, V., Liguori, C., & Paciello, V.** (2022). Development of a new speed measurement technique based on deep learning. Paper presented at the Conference Record - IEEE Instrumentation and Measurement Technology Conference, doi:10.1109/I2MTC48687.2022.9806625

[18] **J. Redmon, S. Divvala, R. Girshick and A. Farhadi,** "You Only Look Once: Unified, Real-Time Object Detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779-788, doi: 10.1109/CVPR.2016.91.

[19] **Wong Kin Yiu,** *Cross Stage Partial Networks,* see https://github.com/WongKinYiu/CrossStagePartialNetworks

[20] **Liu, Shu and Qi, Lu and Qin, Haifang and Shi, Jianping and Jia, Jiaya,** *Path Aggregation Network for Instance Segmentation,* https://doi.org/10.48550/arxiv.1803.01534

[21] **Sebastian Ruder,** *An overview of gradient descent optimization algorithms,* https://arxiv.org/pdf/1609.04747.pdf

[22] **Kingma, Diederik P. and Ba, Jimmy,** *Adam: A Method for Stochastic Optimization,* https://doi.org/10.48550/arxiv.1412.6980

[23] https://cocodataset.org/#detection-eval, " COCO, Common Objects in Context", Available online in July 2022