Coffee Bean Image Classification With Deep Learning

Zeyu Chen
https://github.com/zeyuleochen/Deep-Learning-Project.git

Problem Statement

Introduction:

Roasted coffee beans come in various forms and flavors, each with its unique characteristics. The ability to accurately classify and identify different types of coffee beans plays a crucial role in quality control. Traditional methods of visual inspection and classification are time-consuming, subjective. Therefore, there is a need to leverage the power of deep learning techniques to develop an automated image classification system for roasted coffee beans.

Problem Description:

The coffee bean classification problem aims to to develop a Deep learning model that can accurately classify different levels of roasted coffee beans based on visual characteristics. The goal is to address the challenge of distinguishing between closely related classes and achieve high classification performance, while considering the limitations of dataset size and potential misclassification of similar objects.

Data Source & Contents

- Roasted coffee beans have four roasting levels: green, lightly roasted, medium roasted, and dark roasted.
- The coffee bean photos are captured using an iPhone 12 Mini with a 12-megapixel back camera, including Ultra-wide and WideCamera settings.
- The camera is positioned with a plane parallel to the object's path during image capture.
- Images of roasted coffee beans are taken in various settings to validate a wide range of inputs.
- Both LED light from a lightbox and natural light sources are used to shoot the dataset.
- Image noise is enhanced by placing each variety of coffee beans in a container.
- The images are saved in PNG format and have a resolution of 3024x3032 pixels and resized to 224x224 pixels.

Source: https://www.kaggle.com/datasets/gpiosenka/coffee-bean-dataset-resized-224-x-224

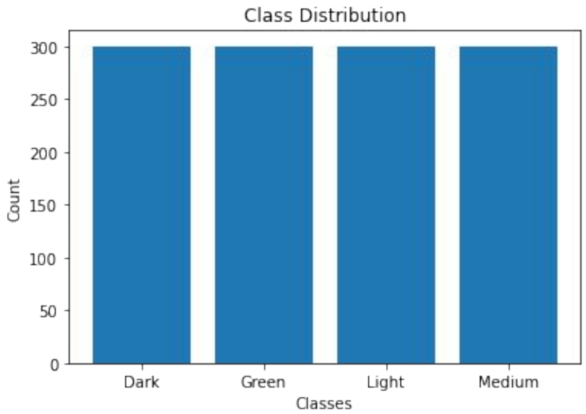
Data Assumptions

1. **Consistency of Image Format**: All input images follow the PNG format, as mentioned previously. It is crucial to ensure that the dataset contains images in a standardized format for proper processing and analysis.
2. **Consistent Image Resolution**: The assumption is that all input images have a consistent resolution of 224x224 pixels.
3. **Image Preprocessing**: It is assumed that necessary preprocessing steps, such as resizing, normalization, and noise enhancement using containers, have been applied to the images.
4. **Limited Varieties**: The dataset assumes a limited range of coffee bean varieties. The classification model trained on this dataset may not generalize well to other coffee bean varieties or blends.
5. **Limited Image Augmentation Techniques**: The Dataset does not specify if other augmentation methods, such as rotation, cropping, or color transformations, were applied.
6. **Labeling Accuracy**: The accuracy of the classification heavily depends on the accuracy of the labels assigned to the coffee bean images.

Exploratory Data Analysis: Distribution

The dataset has a balanced distribution, with each category containing an equal number of images.

This ensures model training, enhanced generalization, and fair performance evaluation. However, drawbacks include limited representation of real-world distributions and potential limitations in handling complex characteristics.



Dark

Green

Light

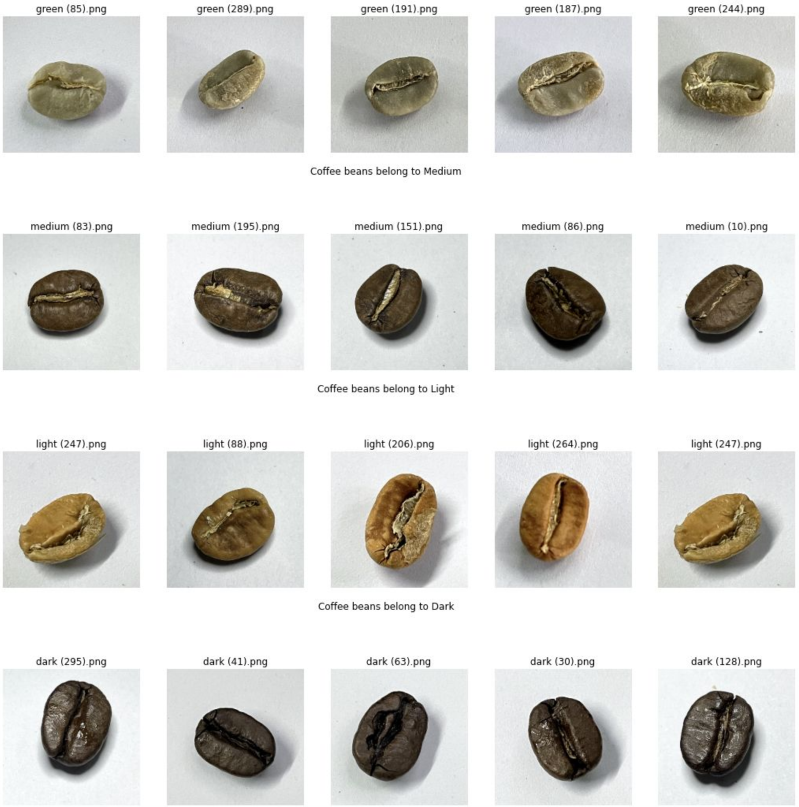
Medium

	class index	filepaths	labels	data set
0	0	train/Dark/dark (1).png	Dark	train
1	0	train/Dark/dark (10).png	Dark	train
2	0	train/Dark/dark (100).png	Dark	train
3	0	train/Dark/dark (101).png	Dark	train
4	0	train/Dark/dark (102).png	Dark	train

Exploratory Data Analysis: Visualization

The images displayed were randomly selected from each category in the training dataset. This random selection ensures a representative sample from each category, providing a balanced view of the dataset.

1. **Consistent Image Quality**: The EDA suggests that the image quality remains constant across the dataset..
2. **No Image Issues**: The EDA does not indicate any issues with the images. This implies that there are no apparent anomalies, distortions that hinder the model's ability to accurately classify the coffee beans.
3. **Potential Distinction Challenges**: The EDA highlights a potential challenge in distinguishing between some classes..



Feature Engineering and Transformations

- **Rescale**: The pixel values of the images are rescaled to a range between 0 and 1 using the 1/255. factor, effectively normalizing the pixel values.
- **Rotation**: Random rotation of the images within a range of -20 to +20 degrees introduces variations in the orientation of the coffee bean images.
- **Shift**: Random horizontal and vertical shifting of the images by a fraction of their width and height (0.1) creates translations in the position of the coffee bean within the image.
- **Shear**: Random shearing transformations on the images (0.1) distort the shape of the coffee bean images, introducing additional variations.
- **Zoom**: Random zooming in and out of the images (0.1) changes the scale of the coffee bean within the image.
- **Flipping**: Random horizontal and vertical flipping of the images create mirror reflections and upside-down versions of the coffee bean, respectively.
- **Validation Split**: 20% of the dataset is designated for validation during training.
- **Fill Mode**: Any newly created pixels during transformations are filled with the nearest existing pixel value using the 'nearest' fill mode.

Feature Engineering and Transformations

After the transformation process, the following changes and improvements can be observed in the randomly selected coffee bean images:

1. **More Portion of the Bean at the Edge**: Moving the beans towards the edges of the image provides a larger portion of the bean for the model to analyze.
2. **Color Distinction**: The applied transformations result in less distinct color differences between the roasting classes. This reduced color distinction could pose challenges.
3. **Lighting Changes**: The transformations can introduce changes in lighting conditions, affecting the shadows and highlights on the beans.
4. **Different Angles of Beans**: The transformations allow the model to learn from various angles and orientations of the coffee beans.

Proposed Model Solution: Convolutional Neural Networks (CNN)

1. **Image-specific Features**: CNNs excel at capturing image-specific features, such as texture, shape, and patterns. This is crucial for accurately distinguishing between different types of coffee beans based on their unique visual characteristics.
2. **Spatial Relationship Preservation**: CNNs preserve the spatial relationships between pixels in an image, allowing them to capture local patterns and structures specific to coffee beans.
3. **Robustness to Variations**: Coffee beans can exhibit variations in size, orientation, and position within the image. CNNs are robust to such variations due to their ability to detect and classify objects regardless of their specific location or orientation.
4. **Scale and Translation Invariance**: CNNs are capable of handling different scales and translations of coffee beans within the image.

Trade-offs:

- **Limited Generalization**: CNNs are prone to struggle in generalizing unseen or uncommon variations in coffee beans.
- **Computational Resources**: CNNs can be computationally intensive, requiring substantial computational resources for training and inference

CNN Model Building: Hyperparameters

- **Input image size**: The width and height of the input images for the CNN model have been set to 128 pixels. This choice strikes a balance between computational efficiency and the preservation of important image details for accurate coffee bean classification.
- **Number of epochs**: The training process consists of 20 iterations over the complete training dataset, enabling the model to capture intricate patterns and variations in coffee bean images. This approach strikes a balance between learning complexity and considering the computational constraints of the machine being used.
- **Batch size**: Each training iteration processes a batch of 32 samples. This choice optimizes the utilization of computational resources while introducing sufficient variability in the training process, leading to effective model training.
- **Input image shape**: The input images are expected to have a shape of (128, 128). This specification ensures that the images are correctly processed by the model's architecture and layers, maintaining compatibility throughout the model.
- **Compile learning rate of Adam() compiler**: The Model use the default learning rate used by the Adam optimizer, which is 0.001.

CNN Model Architecture Summary

Input Layer:

- This layer processes the input images by applying a convolution operation, capturing important features using 32 filters.

Convolutional Layers:

- These convolutional layers apply filters to capture and extract different features and patterns from the input images at increasing levels of complexity.
- Max pooling reduces spatial dimensions, extracting important features while reducing computational complexity.
- Dropout layers help prevent overfitting by randomly dropping out a fraction of the connections during training.

Fully-Connected Layers:

- The flatten layer converts the 2D feature maps into a 1D vector to connect to the fully-connected layers.
- Dense layers perform high-level feature extraction and classification.
- Dropout layers help prevent overfitting in the fully-connected layers.

Layer (type)	Output Shape	Param #
conv2d_32 (Conv2D)	(None, 128, 128, 32)	896
conv2d_33 (Conv2D)	(None, 126, 126, 32)	9248
max_pooling2d_16 (MaxPooling2D)	(None, 63, 63, 32)	0
dropout_19 (Dropout)	(None, 63, 63, 32)	0
conv2d_34 (Conv2D)	(None, 63, 63, 64)	18496
conv2d_35 (Conv2D)	(None, 61, 61, 64)	36928
max_pooling2d_17 (MaxPooling2D)	(None, 30, 30, 64)	0
dropout_20 (Dropout)	(None, 30, 30, 64)	0
conv2d_36 (Conv2D)	(None, 30, 30, 128)	204928
conv2d_37 (Conv2D)	(None, 26, 26, 128)	409728
max_pooling2d_18 (MaxPooling2D)	(None, 6, 6, 128)	0
dropout_21 (Dropout)	(None, 6, 6, 128)	0
flatten_4 (Flatten)	(None, 4608)	0
dense_8 (Dense)	(None, 256)	1179904
dropout_22 (Dropout)	(None, 256)	0
dense_9 (Dense)	(None, 4)	1028

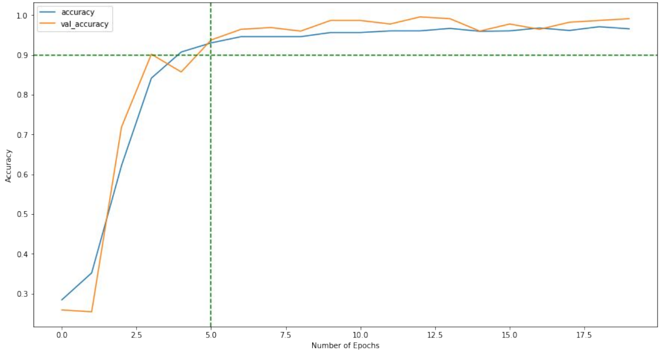
Total params: 1,861,156
Trainable params: 1,861,156
Non-trainable params: 0

Model Regularization Methods

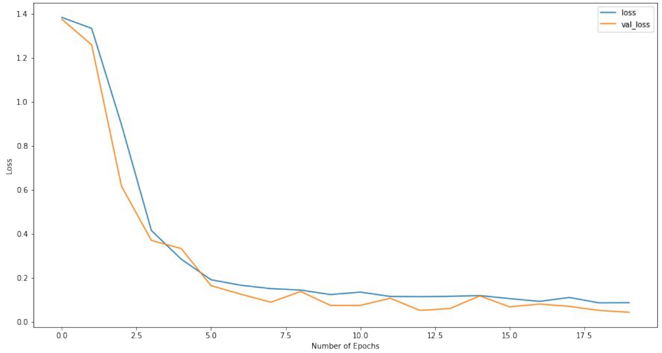
- **Data Augmentation**:
  - Augmentation techniques, including rotation, shifting, shearing, zooming, and flipping, are applied to the training images during Data Processing to introduce uncertainty.
  - Data augmentation increases the diversity of the training dataset, exposing the model to a wider range of variations and reducing overfitting to specific instances.
  - This improves the model's generalization capability, making it more robust to variations in coffee bean images.
- **Dropout Layers**:
  - Dropout is applied after each convolutional layer to regularize the model and prevent overfitting. It helps to randomly deactivate neurons, forcing the network to learn more robust and generalizable features from the input data.

Model Evaluations: Overfitting/Underfitting

For our balanced dataset, accuracy was chosen as the evaluation metric since it provides a clear measure of overall classification performance. However, in cases of imbalanced data, using additional metrics such as AUC is recommended to account for variations in true positive and false positive rates.



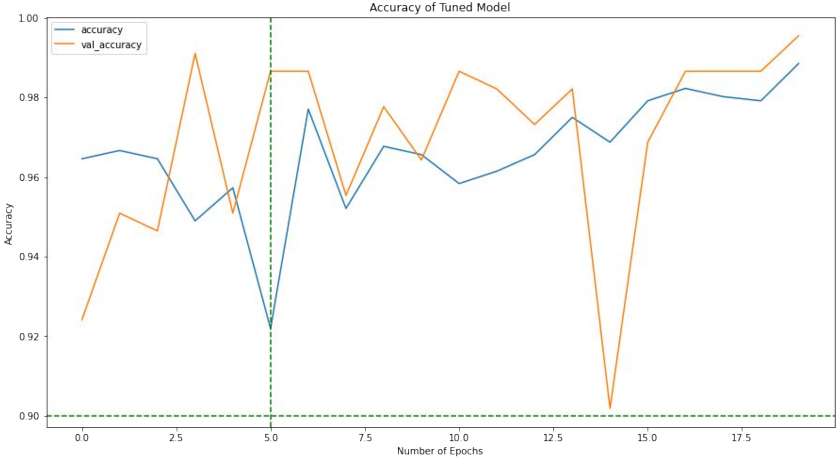
Based on the validation accuracy, accuracy, val_loss, and loss metrics, the model does not display significant signs of overfitting as both accuracy and val_accuracy improve while val_loss and loss decrease. This indicates that the model generalizes well to unseen data.

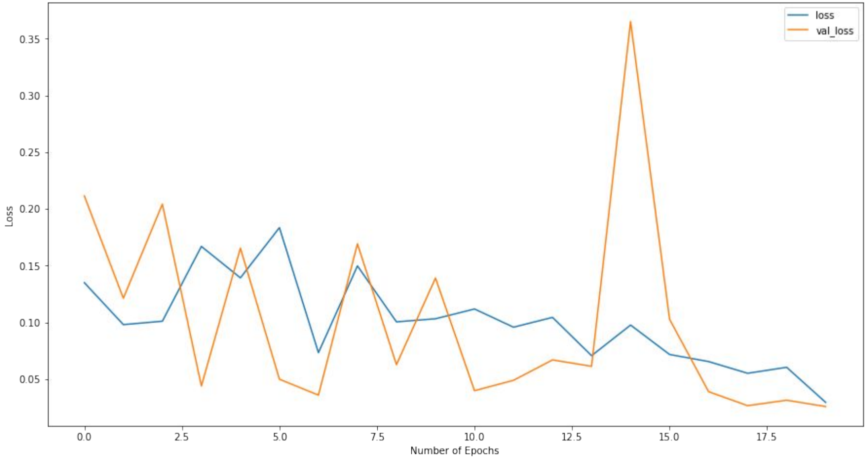


Model Tuning

- The model was tuned using Keras Tuner, and the best hyperparameters to yield the best results of validation accuracy
    - Number of filters in the convolutional layers: 128
    - Dropout rate: 0.33
    - Number of units in the dense layer: 640
    - Adam Compile Learning rate: 0.0008

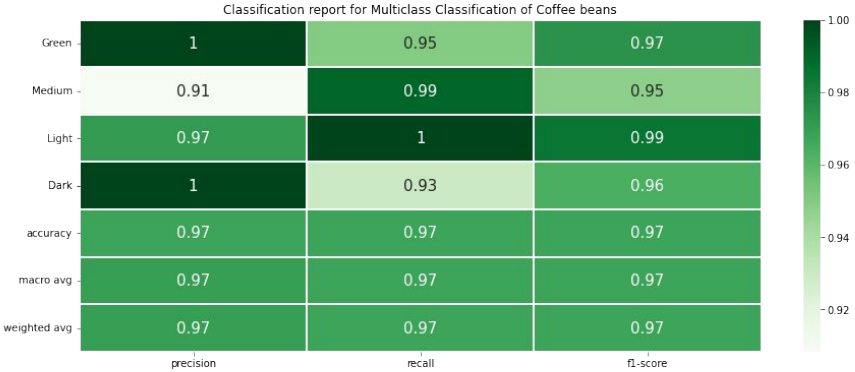
Despite achieving an impressive validation accuracy of 98.75%, the tuned model exhibited significant fluctuations in accuracy and the presence of overfitting in certain epochs. This inconsistency may be attributed to the model's increasing complexity and the limited size of the training dataset. Consequently, the original model, prior to tuning, may be considered a more favorable choice.

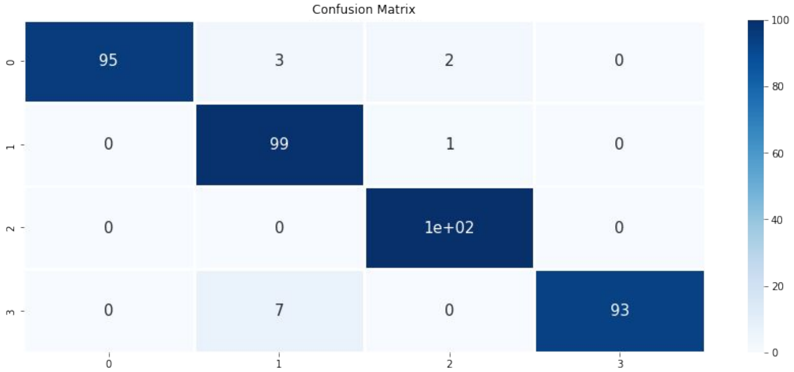




Model Results & Accuracy

- The deployed model achieved an overall accuracy of 96.75% on the test set. It displayed high precision, recall, and F1-scores for each class, indicating its effectiveness in classifying coffee beans based on their color.
- The model exhibited excellent performance in accurately classifying the "Green" and "Dark" coffee bean classes.
- These results align with our initial exploratory data analysis, which indicated that distinguishing between the "Light" and "Medium" classes could be more challenging.

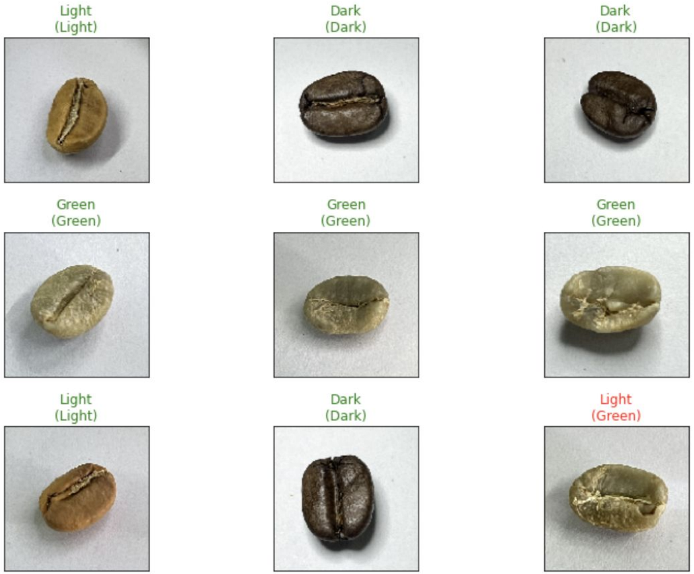




Model Misclassifications

The challenges of differentiating between the "Light" and "Medium" classes might be more difficult due to their similar color characteristics that we assumed initially.

Furthermore, classifying the image of a red bean as "Medium" roast provides evidence that the model is specifically trained to distinguish coffee beans, emphasizing its limited ability to classify objects beyond the scope of coffee beans.



Prediction: Medium

Limitations & Future Works

Limitations:

1. **Limited dataset size and Epochs**: The data size is less than desirable. With a smaller dataset, the model may struggle to learn complex patterns and generalize well to unseen data and it is easy to overfit the data with CNN model. Additionally, due to computation limitations, the model can only be trained with 20 epochs, switch to cloud platforms to increase the computation power and increase the model training.
2. **Lack of versatility**: The model's focus on coffee bean classification limits its ability to accurately classify other types of beans or objects resembling coffee beans. It may misclassify similar objects, such as red beans, as coffee beans.
3. **Risks of labeling biases**: The accuracy of the model heavily relies on the quality and accuracy of the labeling process for the training data. Labeling in the case of this data set can be highly subjective and influenced.

Future Works:

1. **Transfer learning with pre-trained models**: Experimenting with different CNN architectures or more complex models, such as deeper networks or advanced architectures like ResNet or DenseNet.
2. **Dataset expansion:** Collecting a larger and more diverse dataset by acquiring additional images from different sources or employing data augmentation techniques can enhance the model's performance.