Jutisi: Jurnal Ilmiah Teknik Informatika dan Sistem Informasi
https://ojs.stmik-banjarbaru.ac.id/index.php/jutisi/index
Jl. Ahmad Yani, K.M. 33,5 - Kampus STMIK Banjarbaru
Loktabat – Banjarbaru (Tlp. 0511 4782881), e-mail: puslit.stmikbjb@gmail.com
e-ISSN: 2685-0893

Comparison of Coffee Bean Roasting Level Classification Using ResNet50 and VGG16

DOI: http://dx.doi.org/10.35889/jutisi.v14i2.3122

Ibnu Akil
Sistem Informasi, Universitas BSI, Jakarta, Indonesia
*e-mail *Corresponding Author:* Ibnu.ial@bsi.ac.id

Abstract
The classification of coffee bean roasting levels is an important aspect of ensuring coffee product quality. This study compares the performance of two deep learning architectures, ResNet50 and VGG16, in classifying coffee bean images into three roasting levels: light, medium, and dark. The dataset consists of 1,800 images with a resolution of 224×224 pixels, divided into training, validation, and testing sets. Both models were trained with identical configurations using transfer learning and partial fine-tuning. The evaluation results show a very small accuracy difference of only 0.01 point, with ResNet50 slightly outperforming VGG16. This indicates that both models are equally reliable for roast level classification. However, ResNet50 is more time-efficient, requiring only about 10 minutes of training compared to over 25 minutes for VGG16. This difference is suspected to be related to the complexity of VGG16's architecture. The study concludes that ResNet50 offers high efficiency with competitive accuracy. Further research is recommended to optimize VGG16's architecture to improve computational efficiency without compromising accuracy.
Keywords: *Machine learning; Restnet50; VGG16; Coffee bean roasting*

Abstrak
Klasifikasi tingkat roasting biji kopi merupakan aspek penting dalam penjaminan mutu produk kopi. Penelitian ini membandingkan performa dua arsitektur deep learning, ResNet50 dan VGG16, dalam mengklasifikasikan citra biji kopi pada tiga tingkat roasting: light, medium, dan dark. Dataset berisi 1.800 citra beresolusi 224×224 piksel, dibagi menjadi data latih, validasi, dan uji. Kedua model dilatih dengan konfigurasi identik menggunakan transfer learning dan fine-tuning parsial. Hasil evaluasi menunjukkan selisih akurasi sangat tipis, hanya 0,01 poin, dengan ResNet50 sedikit unggul. Hal ini menunjukkan kedua model sama-sama andal untuk klasifikasi tingkat roasting. Namun, ResNet50 lebih efisien secara waktu, hanya memerlukan sekitar 10 menit pelatihan dibandingkan VGG16 yang lebih dari 25 menit. Perbedaan ini diduga terkait kompleksitas arsitektur VGG16. Disimpulkan bahwa ResNet50 menawarkan efisiensi tinggi dengan akurasi kompetitif. Penelitian lanjutan disarankan mengevaluasi optimasi arsitektur VGG16 untuk meningkatkan efisiensi komputasi tanpa mengorbankan akurasi.
Kata kunci: *Machine learning; Restnet50; VGG16; Sangrai biji kopi*

1. Introduction

Coffee is one of the leading agricultural commodities, with high economic value and a widespread culture of consumption worldwide. One of the main challenges faced by specialty coffee roasters is ensuring consistency in quality control. One of the key factors that determines the flavor and aroma profile of coffee is the roasting process. The roast level—such as light, medium, and dark roast—plays a significant role in shaping the final sensory profile of the coffee [1]. Roasting is a crucial step in influencing the chemical changes and cupping quality of the roasted beans [2]. Therefore, accurate identification and classification of roast levels have become an essential need, both in large-scale industrial settings and in quality control processes within coffee-focused small and medium enterprises (SMEs). Cupping is the definitive test of roast quality, demanding a well-trained palate. However, when assessing the degree of roast, the specialty coffee market must establish clear standards or references. Moreover, there needs to be greater consensus among coffee specialists and consumers regarding the vocabulary and terminology used in roast classification.

According to Agtron and SCAA Roast Classification system, they create a numerical index and standardized roast levels to mitigate nomenclature controversies [3]. The level of roasting level is described in Table 1 below.

Table 1. The Agtron and SCAA Scale

Agtron #	SCAA Names
91 – 130	Extremely Light
81 – 90	Very Light
71 - 80	Light
61 - 70	Medium Light
51 - 60	Medium
41 - 50	Moderately Dark
31 - 40	Dark
0 - 30	Very Dark

However, in industrial coffee roasting, it is simplified into three levels: light, medium, and dark.

Nowadays, the assessment of coffee roasting degrees is generally carried out manually by expert roasters through the observation of various roasting variables such as bean color, aroma, and other sensory experiences. Several researchers have developed different models to determine the roasting process [4][5]. However, this approach is subjective and heavily dependent on human expertise. To overcome these limitations, image processing and machine learning–based approaches have emerged as increasingly researched solutions, particularly through the use of deep learning technologies.

In the field of image recognition, Artificial Neural Network (ANN) and Convolutional Neural Network (CNN) models have proven to be effective for various classification tasks. With ANN, image classification challenges arise as 2-dimensional images must be transformed into 1-dimensional vectors. This exponentially raises the count of adjustable parameters. Adding more trainable parameters requires greater storage and processing power. This is why CNN would be an ideal solution to computer vision and image classification problems [6]. Two popular CNN architectures are VGG16 and ResNet50. Both have been widely used in image classification tasks, ranging from general datasets such as ImageNet to more specific domains like plant disease recognition, medical object detection, and food material classification.

The benefit of this research is to help determine the appropriate algorithm for classifying coffee roasting profiles, so that it can be applied to modern coffee roasting machines, where modern coffee roasting machines have utilized artificial intelligence to determine coffee roasting profiles. It will be necessary to have reliable algorithms for the classification of roasting profiles automatically.

2. Literature Review

Several previous studies have demonstrated the effectiveness of CNN or ANN in image-based object classification, including:

Astuti et al. applied an Artificial Neural Network (ANN) method to distinguish roasting levels of Robusta coffee using an E-nose and TGS sensor, and achieved effective results in determining roast levels [7].

Ihsani and Ichwan attempted to combine the VGG16 architecture with DenseNet121 to classify coffee bean quality based on roasting levels. The results of this CNN architecture combination showed that using 15 epochs and a learning rate of 0.0001 produced an average precision of 98.5%, recall of 98%, F1-score of 98.5%, accuracy of 98%, and a loss value of 27.4% [8].

Tama et al have compared VGG-16 and CNN for coffee bean roasting level classification and found that CNN has an accuracy of 98.75% and a running time of 856 ms per step [9]. Meanwhile, Pakaya is using CNN with MobileNet architecture for the classification of coffee roasting level [10]. Firmansyah and Ontoum are also using MobileNet for the same cases [11][12].

In the context of coffee, Alrasyid uses RestNet50 to classify the types of Indonesian local coffee beans. Indonesia has many coffee-producing regions, each with its own unique flavor profile, and also different shapes of coffee beans. The research intention is to categorize three types of arabica coffee beans, and the accuracy is about 99.6% [13].

This study is driven by the need to identify the most effective deep learning architecture for accurately classifying coffee bean roasting levels from visual imagery. Specifically, it compares the performance of two prominent convolutional neural network architectures—ResNet50 and VGG16—by evaluating their accuracy, precision, recall, and training time. The ultimate objective is to recommend the most suitable model for automatic roast level classification, supporting the development of modern coffee roasting systems that integrate artificial intelligence for real-time roast profile determination. This research addresses a notable gap in the literature, where comparative analyses of ResNet50 and VGG16 for coffee roast classification remain scarce, thereby contributing both practical guidance and academic insight to the field.

3. Research Method

To facilitate the research process, a research framework needs to be designed. The stages of this framework begin with determining the type and approach of the research, obtaining the research dataset, performing data preprocessing, building the machine learning model architecture, training the model, analyzing model performance, and selecting appropriate tools and programming languages. Refer to Figure 1 below:

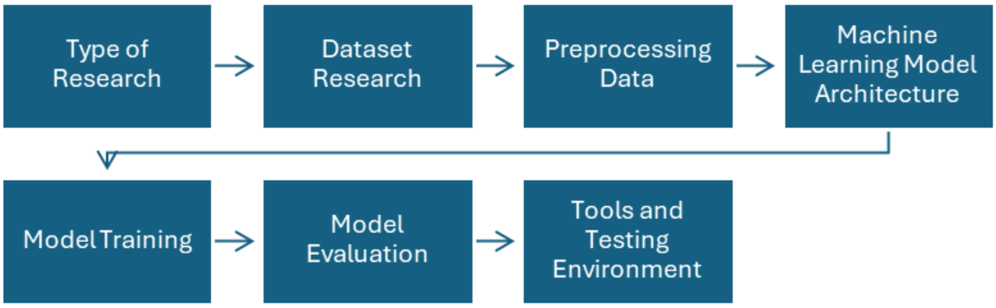

Figure 1. Research Process Framework

3.1. Type and Approach of the Research

This research is a quantitative experimental study that employs a computational approach to compare the performance of two deep learning models (ResNet50 and VGG16) in classifying coffee bean images based on roasting levels. The study is conducted in silico. An insilico study is a research approach that utilizes computer modeling and simulation to investigate biological, medical, or pharmacological phenomena. It involves using computational tools and algorithms to mimic real-world biological processes, often to predict or analyze outcomes that would be difficult, time-consuming, or expensive to study through traditional in vivo (in a living organism) or in vitro (in a controlled laboratory setting) experiments [14].

(i.e., research or experiments performed using computer simulations) using the Python programming platform and the TensorFlow/Keras libraries.

3.2. Research Dataset

The dataset used in this study consists of images of coffee beans categorized into four roasting level classes: Green Bean, Light Roast, Medium Roast, and Dark Roast. The dataset was obtained from: https://www.kaggle.com/datasets/gpiosenka/coffee-bean-dataset-resized-224-x-224.

The total number of images is 1,600, with a split ratio of 80% for training data and 20% for validation data, while the testing set contains 240 images.

3.3. Preprocessing Data

Before training the model, several preprocessing steps were performed, including:

Resizing the images to 256×256 pixels (the standard input size for VGG16 and ResNet50). Normalizing pixel values to the range [0, 1]. Applying real-time data augmentation (such as rotation, flipping, zooming, etc.) to increase data variation and prevent overfitting.

3.4. Machine Learning Model Architecture

1) VGG16 Model:

The VGG16 is a classification algorithm and used to classify images with high accuracy. VGG16 refers to 16 layers that have weights. In VGG16 there are thirteen convolutional layers, five Max Pooling layers, and three Dense layers which sum up to 21 layers [15]. We can see the architecture in figure 2 below.

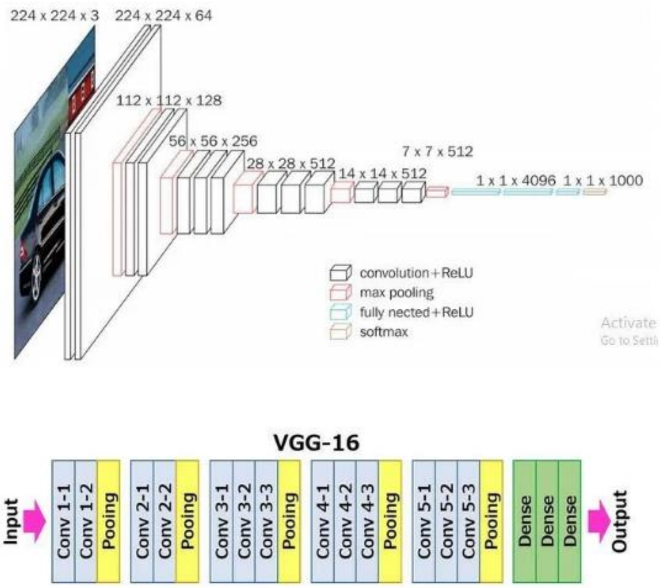

Figure 2. VGG16 Architecture

The architecture was used with pretrained weights from ImageNet. The final fully connected layer was modified to perform 4-class classification (Green Bean, Light, Medium, and Dark). Transfer learning was applied, with fine-tuning performed on selected convolutional layers.

2) ResNet50 Model:

The ResNet50 is the successor of ResNet which consists of 16 residual blocks, with each block consisting of several convolutional layers with residual connections [16]. The architecture is in figure 3.

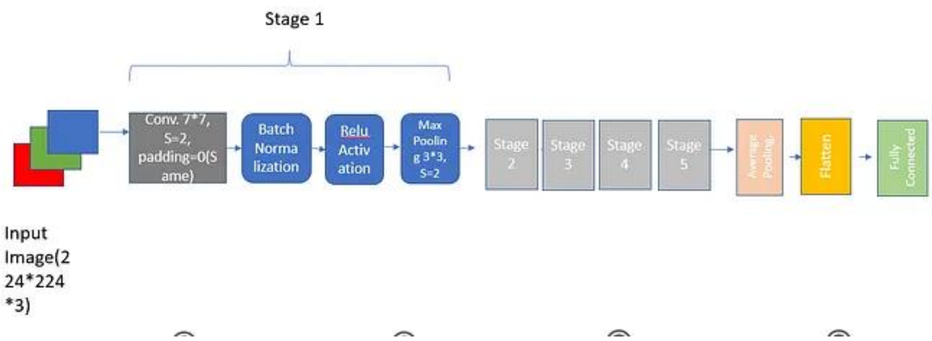

Figure 3. RestNet50 Architecture

RestNet50 architecture was also used with pretrained weights from ImageNet. The output layer was replaced to classify the same 4 classes. Fine-tuning was applied to selected residual blocks to optimize performance.

3.5. Model Training

To ensure a fair comparison, the model configuration and parameters for VGG16 were made consistent with those used for ResNet50, as follows:
Optimizer: Adam
Loss function: Categorical Crossentropy
Activation Dense Layer: Relu
Activation Dense Output: Sotfmax
Batch size: 32
Epoch: 10
Callback: EarlyStopping and ModelCheckpoint

3.6. Model Evaluation

The models were evaluated using the following metrics: Accuracy, Precision, Recall, F1-Score, and Confusion Matrix. The evaluation results of both models were then compared to determine the most optimal model for classifying coffee bean roasting levels.

3.7. Tools and Testing Environment

Here is the tools and testing environment setup for this research:
Programming Language: Python 3
Libraries: TensorFlow, Keras, NumPy, Matplotlib, scikit-learn
Platform: Google Colab with GPU support or a local machine with CUDA-enabled GPU

4. Results and Discussion

4.1. Data Exploration

The developed program is a computer-based simulation built using Python, along with the TensorFlow, Scikit-Learn, Pandas, and Matplotlib libraries. The first step involves exploring the research dataset, which consists of coffee bean images categorized into four classes: Green, Light, Medium, and Dark. Refer to the following figure 2. And then we split the data into 80% for training and 20% for testing.

Figure 2. Image Data of Coffee Bean Roasting Levels

4.2. Model Configuration and Training

The next step is to configure the model parameters for ResNet50, as follows:

```
resnet_50V2 = Sequential([
    base_model,
    GlobalAvgPool2D(),
    Dense(256, activation='relu'),
    Dropout(0.2),
    Dense(n_classes, activation='softmax')
])
```

After configuration, the next step is to train the model. Here are the results for RestNet50:

```
Epoch 1/10
30/30 ━━━━━━━━━━━━━━ 76s 2s/step - accuracy: 0.6222 - loss: 1.0498 - val_accuracy: 0.9250 - val_loss: 0.2236
Epoch 2/10
30/30 ━━━━━━━━━━━━━━ 69s 2s/step - accuracy: 0.9384 - loss: 0.1617 - val_accuracy: 0.9667 - val_loss: 0.1039
…
Epoch 10/10
30/30 ━━━━━━━━━━━━━━ 69s 2s/step - accuracy: 0.9837 - loss: 0.0411 - val_accuracy: 0.9625 - val_loss: 0.0806
```

For measurement, a classification report is used, which includes precision, recall, f1-score, and support. The results are shown in table 2 below for RestNet50:

Table 2. Classification Report Model RestNet50:

Class	Precision	Recal	F1-Score	Support
Dark	1.00	0.98	0.99	100
Green	0.99	1.00	1.00	100
Light	1.00	0.99	0.99	100
Medium	0.98	1.00	0.99	100
Accuracy			0.99	400

To evaluate performance in a more informative way, this study uses a confusion matrix as shown in the following Figure 3:

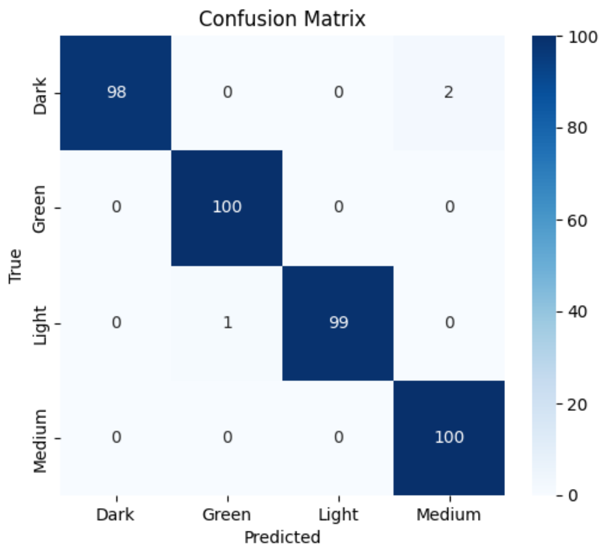

Figure 3. Confusion Matrix of the ResNet50 Model

Next, for VGG16, the same steps as ResNet50 are used. The following are the results:
Configuration for VGG16:

```
# Gunakan model VGG16 tanpa klasifikasi layer atas
base_model = VGG16(weights='imagenet', include_top=False,
input_shape=(256, 256, 3))

for layer in base_model.layers:
    layer.trainable = False

# tambahkan konfigurasi kustom layer
a = Flatten()(base_model.output)
a = Dense(128, activation='relu')(a)
a = Dense(4, activation='softmax')(a)
```

The following is the training output:

```
Epoch 1/10
30/30 ━━━━━━━━━━━━━ 192s 6s/step - accuracy: 0.4607 - loss: 1.2900 -
precision: 0.5849 - recall: 0.1641 - val_accuracy: 0.6375 - val_loss:
0.8556 - val_precision: 0.7606 - val_recall: 0.4500
Epoch 2/10
30/30 ━━━━━━━━━━━━━ 201s 7s/step - accuracy: 0.8924 - loss: 0.4505 -
precision: 0.9480 - recall: 0.8101 - val_accuracy: 0.7500 - val_loss:
0.5852 - val_precision: 0.8308 - val_recall: 0.6750
…
Epoch 10/10
30/30 ━━━━━━━━━━━━━ 193s 7s/step - accuracy: 0.9981 - loss: 0.0576 -
precision: 0.9981 - recall: 0.9981 - val_accuracy: 0.8958 - val_loss:
0.2856 - val_precision: 0.9254 - val_recall: 0.8792
```

The following is the classification report for VGG16:

Table 3. Classification Report VGG16:

Class	Precision	Recal	F1-Score	Support
Dark	0.99	0.98	0.98	100
Green	1.00	0.97	0.98	100
Light	0.95	0.99	0.97	100
Medium	0.97	0.97	0.97	100
Accuracy			0.98	400

The confusion matrix for VGG16 is as follows:

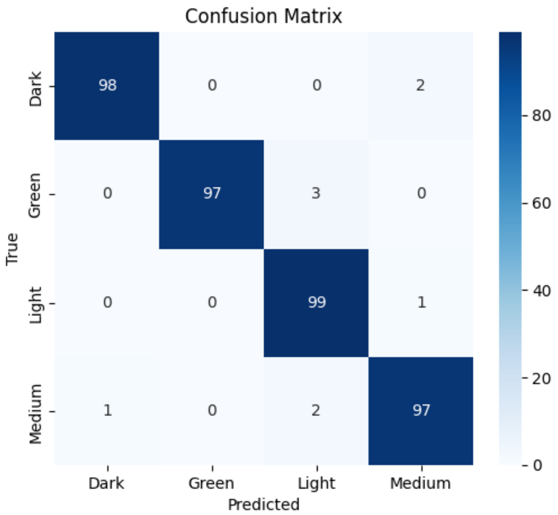

Figure 4. Confusion Matrix of the VGG16 Model

4.3. Evaluation of VGG16 and ResNet50 Models

Based on the training and prediction results of both models, the performance metrics are presented in Table 1 and Table 2, which contain the classification reports. The accuracy achieved by ResNet50 is 0.99, while VGG16 achieved an accuracy of 0.98, indicating that ResNet50 outperforms VGG16 by 0.01 point.

ResNet50 performed better in detecting the Dark and Light classes, with a precision of 1.00, whereas VGG16 excelled in the Green class, also achieving a precision of 1.00. The results of the confusion matrices can be seen in Figure 3 for ResNet50 and Figure 4 for VGG16.

In terms of training time performance, with each model trained for 10 epochs, ResNet50 required approximately 10 minutes for training, while VGG16 took more than 25 minutes. This is a significant difference, indicating that ResNet50 is considerably more efficient than VGG16 in terms of training time.

This efficiency can be attributed to the architectural differences between the two models: VGG16 consists of 21 layers, making it computationally heavier. In contrast, ResNet50 has 50 layers, but it utilizes only 5 main residual blocks, resulting in faster training while maintaining high performance.

The results indicate that ResNet50 slightly outperforms VGG16 in terms of classification accuracy (0.99 vs. 0.98) and demonstrates superior efficiency in training time, requiring less than half the time of VGG16 for the same number of epochs. These findings suggest that ResNet50 may be more suitable for practical implementations where both accuracy and computational efficiency are critical, such as in real-time coffee roasting monitoring systems.

In relation to previous research on machine learning algorithms to determine the roasting level of coffee conducted by [8], using VGG16 and CNN, where CNN is implemented with the DenseNet architecture. In DenseNet, depth is measured as the number of layers that form a dense block. A dense block consists of several convolutional layers that are sequentially connected with concatenation. DenseNet's performance is strong on classification tasks with diverse datasets due to feature reuse. It offers parameter efficiency and good generalization potential. Furthermore, in their study, the VGG16 model combined with the DenseNet architecture achieved an accuracy of 98.7%.

In [8], the aim is to reconstruct the VGG16 architecture using the DenseNet architecture. Meanwhile, this study aims to compare the VGG16 model with RestNet50 to determine which model is best for the coffee roasting level classification task without altering the VGG16 architecture. It was found that RestNet50 outperforms with an accuracy of 0.99%, while VGG16 achieves 0.98%. The VGG16 results are not far from [8]' findings, even though reconstruction of the architecture with DenseNet was not used. The results show that it's not necessary to reconstruct the architecture of the VGG16, but it is better to use Resnet50 instead of reconstructing VGG16 for the classification of coffee roasting level.

However, the performance gap between the two models (VGG16 and ResNet50), while present, is relatively small, which opens opportunities for further investigation. Future research could explore the following directions:

1. Dataset Expansion and Variability – Increasing the dataset size and including images from diverse lighting conditions, camera types, and bean origins could help evaluate the robustness of each model in more realistic and variable environments.
2. Model Optimization Techniques – Applying transfer learning fine-tuning strategies, data augmentation, or pruning techniques to VGG16 may help reduce its training time without significantly compromising accuracy. Similarly, hyperparameter tuning could be applied to both models to seek further performance gains.
3. Integration with IoT-Based Roasting Systems – Testing the models in real-time scenarios with IoT-enabled coffee roasting machines would assess their inference speed, latency, and suitability for embedded hardware deployment.
4. Comparisons with Other Architectures – Extending the comparison to include more recent architectures such as EfficientNet, DenseNet, or MobileNet could reveal whether newer models offer better trade-offs between accuracy, efficiency, and resource consumption.
5. Class Imbalance and Misclassification Analysis – A deeper examination of misclassified samples, especially in borderline roast levels, may provide insights into where each model struggles and how targeted improvements could be made.

By pursuing these directions, future research could not only validate the current findings but also enhance the applicability of deep learning models in the automation of coffee roasting profile classification.

5. Conclusion

Based on the research findings, the evaluation results show that the accuracy difference between the two models in identifying coffee roasting levels is only 0.01 point, with ResNet50 slightly outperforming VGG16 when tested on the same coffee bean roasting image dataset under identical configurations. Therefore, it can be concluded that both models are equally reliable in recognizing images, specifically in the context of roasted coffee bean classification.

However, from the performance perspective, ResNet50 clearly outperforms VGG16, requiring only 10 minutes of training time for a dataset of 1,800 images with a resolution of 244 × 244 pixels.

For future research, it is recommended that the VGG16 model be re-evaluated, particularly with regard to its large number of layers, which significantly increases training time.

References

[1] L. Poisson, I. Blank, A. Dunkel, and T. Hofmann, “The Chemistry of Roasting-Decoding Flavor Formation,” in *The Craft and Science of Coffee*, Elsevier Inc., 2017, pp. 273–309. doi: 10.1016/B978-0-12-803520-7.00012-8.

[2] K. F. Ting, J. C. Chen, and T. L. Chen, “Effect of coffee roasting on the cupping quality of coffee,” *Coffee Sci.*, vol. 19, 2024, doi: 10.25186/.v19i.2174.

[3] “The Ultimate Guide to Understanding Coffee Roast Levels,” coffeebros.com. Accessed: Jul. 17, 2025. [Online]. Available: https://coffeebros.com/blogs/coffee/the-ultimate-guide-to-understanding-coffee-roast-levels?srsltid=AfmBOoqzWtMDSq7ov_uDv-klLX1iqz35Jb_3CqoGbBMmuoWz97PvPAZH

[4] C. da C. Garcia *et al.*, “Relative importance and interaction of roasting variables in coffee roasting process,” *Coffee Sci.*, vol. 13, no. 3, pp. 379–388, 2018, doi: 10.25186/cs.v13i3.1483.

[5] W. B. Sunarharum, M. Nurminah, and N. G. Purba, “Effect of Different Roasting Levels and Manual Brewing Techniques on the Sensory Profile of Liberica Coffee with Honey Process,” *J. Coffee Sustain.*, vol. 01, no. 01, pp. 40–48, 2024, doi: 10.21776/ub.jcs.2024.01.01.05.

[6] N. Klinger, “ANN and CNN: Analyzing Differences and Similarities,” viso.ai. Accessed: Jul. 17, 2025. [Online]. Available: https://viso.ai/deep-learning/ann-and-cnn-analyzing-differences-and-similarities/

[7] S. D. Astuti *et al.*, “Electronic nose coupled with artificial neural network for classifying of coffee roasting profile,” *Sens. Bio-Sensing Res.*, vol. 43, p. 100632, 2024, doi: 10.1016/j.sbsr.2024.100632.

[8] A. F. Ihsani and M. Ichwan, “Kombinasi Arsitektur VGG16 dan DenseNet121 Untuk Klasifikasi Kualitas Biji Kopi Berdasarkan Tingkat Kematangan,” vol. 8, no. 1, pp. 1–12, 2023.

[9] M. N. Tama, A. B. Saptomo, D. K. Baroroh, A. P. Rifai, N. H. Tho, and A. Info, “Comparative Evaluation of Convolutional Neural Network Full Learning Model with Transfer Learning (VGG-16) for Coffee Bean Roasting Level Classification,” vol. 6, no. 2, pp. 332–343, 2025, doi: 10.59395/ijadis.v6i2.1398.

[10] I. M. Pakaya, R. Radi, and B. Purwantana, “Classification of Roasting Level of Coffee Beans Using Convolutional Neural Network with MobileNet Architecture for Android Implementation,” *J. Tek. Pertan. Lampung (Journal Agric. Eng.*, vol. 13, no. 3, pp. 924-933, 2024, doi: 10.23960/jtep-l.v13i3.924-932.

[11] T. Firmansyah *et al.*, “Klasfikasi Tingkat Kematangan Roasting Biji Kopi Berbasis Deep Learning dengan Arsitektur MobileNet,” *J. Inf. Syst. Res.*, vol. 6, no. 2, pp. 1433–1443, 2025, doi: 10.47065/josh.v6i2.6811.

[12] S. Ontoum, T. Khemanantakul, P. Sroison, T. Triyason, and B. Watanapa, “Coffee Roast Intelligence,” 2022, [Online]. Available: http://arxiv.org/abs/2206.01841

[13] M. A. Alrasyid, B. Rohmatullah, R. Damayanti, and D. F. Al-Riza, “ResNet-50 to classify the types of Indonesian local coffee beans,” *Proc. 1ST UNHAS Int. Conf. Agric. Technol.*

(UICAT 2021), vol. 2596, no. 1, p. 050008, 2023.
[14] O. Osakwe, "The Significance of Discovery Screening and Structure Optimization Studies," *Soc. Asp. Drug Discov. Dev. Commer.*, pp. 109-128, 2016.
[15] S. T. Musharraf, "Everything you need to know about VGG16- A Deep Learning Architecture," linkedin.com. Accessed: Jul. 17, 2025. [Online]. Available: https://www.linkedin.com/pulse/everything-you-need-know-vgg16-deep-learning-syed-talal-musharraf-pjkkf/
[16] T. Sharma, "Detailed Explanation of Resnet CNN Model," medium.com. Accessed: Jul. 17, 2025. [Online]. Available: https://medium.com/@sharma.tanish096/detailed-explanation-of-residual-network-resnet50-cnn-model-106e0ab9fa9e