

## KLASIFIKASI VARIETAS KOPI BERDASARKAN GREEN BEAN COFFEE MENGUNAKAN METODE MACHINE LEARNING

M. Agung Nugroho<sup>1)</sup>, Maria Mediatrice Sebatubun<sup>2)</sup>

<sup>1,2)</sup> Teknik Informatika, STMIK AKAKOM Yogyakarta  
email : [m.agung.n@akakom.ac.id](mailto:m.agung.n@akakom.ac.id)<sup>1)</sup>, [memey@akakom.ac.id](mailto:memey@akakom.ac.id)<sup>2)</sup>

### Abstraksi

Kopi memiliki jenis yang terdiri dari kopi robusta dan kopi arabika. Kopi arabika memiliki beragam varietas berdasarkan daerah asal atau dikenal dengan nama single origin. Petani dan coffee expert membedakan varietas kopi arabika melalui bentuk dan karakter dari green bean atau kopi mentah. Perbedaan ciri ini meliputi perbedaan warna, bentuk, atau tekstur. Setiap varietas kopi memiliki harga yang berbeda-beda tergantung dari jenis varietasnya. Walaupun demikian, tidak semua petani, dan pemilik coffee shop mampu mengenali varietas kopi dengan hanya melihat green bean. Sehingga, bisa terjadi kesalahan dalam mengenali varietas kopi jika pemilik coffee shop tidak mengetahui pengetahuan tentang kopi. Hal ini bisa disolusikan dengan pemodelan yang dapat mengidentifikasi varietas kopi arabika agar dapat digunakan sebagai second opinion untuk mengidentifikasi varietas kopi arabika. Salah satu metode yang digunakan adalah metode pencitraan. Penelitian ini dimulai pada tahap pra-proses yaitu cropping citra, tahapan selanjutnya adalah melakukan proses klasifikasi menggunakan metode Machine Learning. Metode ini dapat melakukan ekstraksi fitur secara otomatis sehingga tahap pengenalan varietas kopi dapat dilakukan dengan lebih cepat dan diharapkan mampu memberikan tingkat akurasi yang tinggi.

### Kata Kunci :

image processing, machine learning, raw bean coffee, CNN.

### Abstract

*Coffee has a type consisting of robusta coffee and arabica coffee. Arabica coffee has various varieties based on origin or known as single origin. Farmers and coffee experts differentiate Arabica coffee varieties through the shape and character of green beans or raw coffee. Differences in these characteristics include differences in color, shape, or texture. Each coffee variety has a different price depending on the type of variety. However, not all farmers, and coffee shop owners are able to recognize coffee varieties by just looking at green beans. So, there can be mistakes in recognizing coffee varieties if the coffee shop owner does not know the knowledge about coffee. This can be solved by building a system that is able to recognize varieties of Arabica coffee so that it can be used as a second opinion in recognizing Arabica coffee varieties. One way that can be done is with the imaging method. This research begins at the pre-process stage, namely image cropping, the next step is to carry out a classification process using the Machine Learning method. This method can automatically extract features so that the introduction of coffee varieties can be done more quickly and is expected to provide a high level of accuracy.*

### Keywords :

image processing, machine learning, raw bean coffee, CNN.

### Pendahuluan

Kopi memiliki jenis yang terdiri dari kopi robusta dan kopi arabika. Kopi arabika memiliki beragam varietas berdasarkan daerah asal atau dikenal dengan nama single origin. Petani dan coffee expert membedakan varietas kopi arabika melalui bentuk dan karakter dari green bean atau kopi mentah. Perbedaan ciri ini meliputi perbedaan warna, bentuk, atau tekstur. Setiap varietas kopi memiliki harga yang berbeda-beda tergantung dari jenis varietasnya. Walaupun demikian, tidak semua petani, dan pemilik coffee shop mampu mengenali

varietas kopi dengan hanya melihat green bean. Sehingga, bisa terjadi kesalahan dalam mengenali varietas kopi jika pemilik coffee shop tidak mengetahui pengetahuan tentang kopi. Hal ini bisa disolusikan dengan pemodelan yang dapat mengidentifikasi varietas kopi arabika agar dapat digunakan sebagai second opinion untuk mengidentifikasi varietas kopi arabika. Salah satu metode yang digunakan adalah metode pencitraan. Pada penelitian-penelitian sebelumnya terdapat metode dengan menggunakan beragam teknik pengolahan citra untuk proses identifikasi varietas kopi. Metode yang umum adalah *feature extraction*

yang bertujuan untuk mengidentifikasi karakteristik atau fitur dari objek. Metode ini digunakan untuk melakukan proses identifikasi fitur-fitur yang dimiliki green bean coffee atau raw bean agar sistem yang dibuat memiliki kemampuan dalam mengenali varietas-varietas kopi. Metode yang digunakan untuk proses tersebut, salah satunya adalah dengan melakukan pengambilan fitur tekstur, dimana fitur-fitur tekstur pada permukaan raw bean coffee dapat diperhitungkan secara statistic. Fitur tekstur berupa fitur tekstur orde satu dan fitur tekstur orde dua. Perhitungan statistika digunakan pada pengukuran tekstur pada orde pertama dengan menggunakan dasar nilai piksel dari citra asli yaitu varian, dan tidak melihat dari hubungan ketetangaan piksel. Sementara itu, GLCM menggunakan teknik perhitungan dengan hubungan antarpasangan dua piksel citra asli [1]. Penelitian lain yang membahas identifikasi varietas biji kopi dengan fitur tekstur orde dua adalah metode *Gray Level Co-occurrence Matrices* (GLCM) [2][3]. Selain itu, terdapat juga penelitian dengan menggunakan fitur tekstur orde satu dan orde dua [4] dengan tujuan agar hasil pengenalan dapat lebih akurat. Penggunaan fitur tekstur ini digunakan karena berdasarkan pada kenampakan objek pada citra hasil pemotretan, yaitu terdapat perbedaan tekstur untuk masing-masing varietas. Setelah diperoleh karakteristik fitur yang sesuai, proses berikutnya adalah menggunakan teknik klasifikasi agar dapat mengenali jenis-jenis varietas kopi.

Teknik klasifikasi terdiri dari berbagai metode yang berbeda dan memiliki tingkat kehandalan yang berbeda pula. Berdasarkan penelitian sebelumnya, proses klasifikasi menggunakan metode *Support Vector Machine* (SVM) [3][2], *Naïve Bayes Classifier* (NBC) [5], dan ada juga yang menggunakan *MultiLayer Perceptron* (MLP) [4]. Selain metode-metode tersebut, terdapat beberapa penelitian lain dalam bidang pengolahan citra yang juga melakukan klasifikasi menggunakan metode hasil pengembangan dari *Neural Network* yaitu *Deep Learning*. Hasil yang diperoleh dari metode tersebut menunjukkan tingkat akurasi yang cukup tinggi dengan menggunakan data citra yang berbeda-beda. Berdasarkan penelitian-penelitian tersebut, maka penelitian ini akan melakukan proses identifikasi pada tiga jenis varietas kopi arabika yaitu Red Bourbon, Catimor dan Sigararutang dengan menggunakan salah satu metode *machine learning* yang sedang populer yaitu *Deep Learning*.

## Tinjauan Pustaka

Adapun penelitian-penelitian yang membahas tentang proses identifikasi varietas kopi dengan menggunakan teknik ekstraksi fitur dan klasifikasi. Penelitian tersebut [2] melakukan proses

identifikasi raw bean coffee dengan contoh citra yang diambil adalah 3367 citra. Pada proses awal menggunakan teknik segmentasi sehingga dapat memisahkan *background* dan objek di citra menggunakan metode *contours algorithm*, proses selanjutnya adalah *feature extraction* dengan GLCM pada empat sudut arah seperti pada sudut  $0^0$ ,  $45^0$ ,  $90^0$  dan  $135^0$ . Pada penelitian ini, GLCM menggunakan fitur-fitur diantaranya adalah *contrast*, *dissimilarity*, *homogeneity*, *energy*, *entropy*, *correlation* dan *variance*. Pada proses akhir dilakukan proses klasifikasi menggunakan metode SVM. Diperoleh hasil dengan tingkat akurasi 86%.

Penelitian lain mengenai klasifikasi raw bean coffee dilakukan dengan menggunakan beberapa metode [3]. Pada proses pertama dilakukan teknik *pre-processing* pada citra menggunakan algoritma *Multi-scale Retinex with Color Restoration* (MSRCR) kemudian dilakukan proses perbandingan menggunakan metode lainnya yaitu *Histogram Equalization* dan *Contrast Limited Histogram Equalization* (CLAHE). Proses kedua adalah ekstraksi fitur menggunakan metode GLCM. Pada proses ketiga adalah proses klasifikasi dengan metode SVM. Dari penelitian tersebut diperoleh hasil bahwa metode MSRCR memiliki kualitas visual dan akurasi yang lebih unggul.

Penelitian lainnya yaitu pengenalan terhadap raw bean coffee berdasarkan warna dengan algoritma ANN [5]. Metode ini merupakan transform model dimana NBC digunakan dalam mengidentifikasi raw bean coffee, dimana karakteristik pembeda. Metode ANN memiliki nilai kesalahan senilai 1,15% dan nilai akurasi senilai 100% diperoleh dengan menggunakan NBC. Agar dapat menggunakan banyak fitur terdapat penelitian [4] dengan menerapkan beberapa metode fitur ekstraksi metode penggabungan fitur tekstur orde satu dan orde dua. Pada tahap klasifikasi dengan metode ANN diperoleh nilai akurasi mencapai 80%.

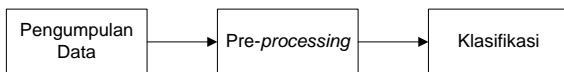
Metode lain seperti *machine learning* yang juga dapat melakukan klasifikasi dengan baik adalah *deep learning*. Penelitian-penelitian yang memanfaatkan metode ini melakukan proses klasifikasi dengan menggunakan beragam data set citra. Salah satu penelitian [6] yang dilakukan adalah melakukan otomasi penilaian kualitas batuan menggunakan resolusi citra yang berbeda. Penelitian ini dilakukan dengan mengusulkan teknik data-driven pertama yang diterapkan pada klasifikasi ubin granit. Pendekatan tersebut mencoba memahami pola intrinsik pada potongan citra yang kecil menggunakan *Convolutional Neural Network* (CNN). Hasil penelitian tersebut menunjukkan bahwa metode yang diusulkan mampu mengenali resolusi citra baik resolusi tinggi maupun rendah. Penelitian lain juga mengusulkan sistem *Computer-aided Diagnosis* (CAD)

menggunakan CNN [7] yang bertujuan untuk membantu radiologi dalam mengklasifikasikan massa lesi pada citra mammogram. Tingkat akurasi yang diperoleh cukup tinggi jika dibandingkan dengan metode-metode terbaru yaitu akurasi sebesar 98,94%. Kemudian metode tersebut juga diujikan pada dataset independen (MIAS) dan akurasi yang diperoleh sebesar 98,23%. Sementara terdapat penelitian yang juga memanfaatkan CNN untuk melakukan klasifikasi terhadap kualitas kayu berdasarkan pada citra [8]. Metode ini dibandingkan dengan kombinasi dari ekstraksi fitur berbasis tekstur dan teknik tradisional (*Decision Tree, Neural Networks, Nearest Neighbors* dan SVM). Penggunaan *deep learning* mampu mencapai performa prediksi yang sangat tinggi, terutama dalam skenario yang sangat kompleks. Penelitian lain yang menggunakan CNN untuk *medical image retrieval* terhadap citra medis sebagai dataset [9]. Penelitian ini menyimpulkan bahwa metode yang diusulkan sangat mampu untuk melakukan temu kembali citra dengan akurasi yang diperoleh sebesar 99,77%. Pada tahun 2017, terdapat sebuah penelitian menggunakan CNN untuk mengenali kecacatan pada biji kopi hijau di Timor Leste [10]. Hasil yang diperoleh yaitu akurasi sebesar 72,4% hingga 98,7% berdasarkan jenis cacat.

## Metode Penelitian

Pada penelitian ini digunakan beberapa tahapan penelitian diantaranya adalah:

### A. Tahap Penelitian

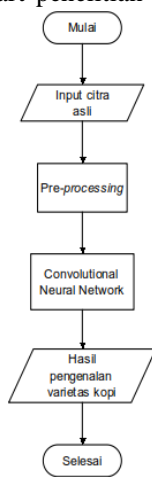


Gambar 1. Tahapan penelitian

Penelitian ini diawali dengan proses pengambilan dan collecting data citra, kemudian tahap pre-processing untuk menyiapkan citra sebelum proses selanjutnya. Setelah proses ini data citra akan langsung diklasifikasi untuk mengetahui kehandalan metode yang digunakan dalam mengenali varietas kopi arabika.

### B. Rancangan Penelitian

Berikut flowchart penelitian yang menggambarkan



metode-metode yang digunakan dalam penelitian.

Gambar 2. Flowchart Penelitian

Data yang digunakan sebagai masukan adalah data mentah yang diperoleh dari pengambilan citra secara manual menggunakan kamera. Selanjutnya *pre-processing* dilakukan untuk menyiapkan citra sebelum diklasifikasi. Tahap terakhir adalah klasifikasi menggunakan CNN untuk menguji kemampuan metode ini dalam mengenali varietas kopi.

## Hasil dan Pembahasan

Dalam proses pengenalan green bean kopi menggunakan metode Deep Learning dilakukan beberapa tahapan proses yang direpresentasikan melalui proses:

### A. Persiapan Data

Pada fase ini dilakukan proses collecting data-data yang digunakan untuk proses selanjutnya seperti training. Data-data ini berupa green bean arabika kopi yang terdiri dari Ciwangi Red Bourbon, Ciwangi Catimor, Rasuna Sigararutang. Ketiga varietas ini di photo posisi depan dan belakang dengan mengambil 30 sample. Hasil dari photo digital menjadi basis data untuk training pada mesin autoML google. Masing-masing varietas memiliki data lebih dari 50 dan 10 data untuk dilakukan proses validasi. Total jumlah sample yang digunakan adalah 188.

### B. Ekstrasi data

Proses ekstrasi data dilakukan untuk mengetahui ciri dari objek yang telah ditentukan dalam hal ini adalah green bean. Dari green bean terdapat ciri warna, kebulatan, tekstur dan ciri khas lainnya. Untuk membantu machine learning autoML, seluruh data yang ada dipisahkan dan diberikan label dari setiap green bean sehingga menjadi 3 jenis green bean yaitu red bourbon, catimor, sigararutang. Ketiga jenis varietas yang sudah diberikan label tersebut akan digunakan sebagai dataset yang kemudian di deploy ke Google AutoML.

### C. Seleksi fitur

Dataset terdiri dari training, validating, dan testing. Dengan tidak menentukan pemisahan, maka AutoML Vision secara otomatis menggunakan 80% dari gambar untuk training, 10% untuk validating, dan 10% untuk testing.

Training set, Mayoritas merupakan data set yang akan digunakan untuk training. Ini adalah data

yang "dilihat" oleh model Anda selama proses training. ini digunakan untuk mempelajari parameter-parameter model, yaitu bobot koneksi antara node-node dari jaringan saraf.

Validation Set, biasa disebut "dev" set, juga digunakan selama proses training. Setelah kerangka pembelajaran model menggabungkan data training pada setiap iterasi dari proses training, ia menggunakan kinerja model pada set validasi untuk menggunakan model hyperparameters, yang merupakan variabel yang menentukan struktur model.

Test set, Tidak berhubungan dengan proses training. Setelah model selesai melakukan proses training, test set digunakan untuk menguji ketangguhan dari model data yang telah ada.

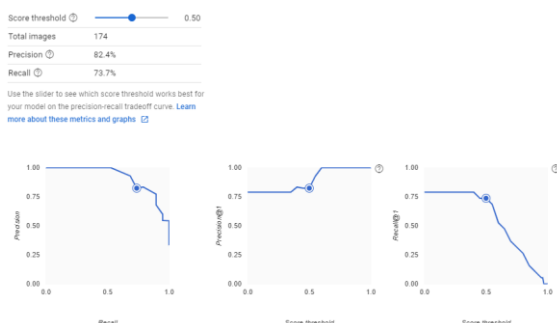
#### D. Proses identifikasi

Dalam proses identifikasi menggunakan google autoML, dapat dilakukan komparasi performance model dengan menggunakan label confusion matrix. Dalam model yang ideal, seluruh nilai yang bernilai tinggi pada garis biru diagonal, dan yang lain bernilai rendah. Ini menunjukkan identifikasi objek dapat teridentifikasi dengan baik.

True label	Predicted label		
	Ciwangi Redbourbon	Ciwangi catimor	Rasuna Sigararutang
Ciwangi Redbourbon	71.4%	28.6%	-
Ciwangi catimor	14.3%	85.7%	-
Rasuna Sigararutang	-	20.0%	80.0%

Gambar 3. Confusion Matrix

Pada gambar dapat dilihat bahwa google autoML dapat mendeteksi varietas ciwangi red bourbon sebesar 71.4%, ciwangi catimor 85.7%, dan rasuna sigararutang 80%. Setelah proses identifikasi ini dapat dilakukan proses prediksi dengan memberikan sample green bean pada google autoML. Dalam proses identifikasi ini terdapat mekanisme precision dan recall untuk membantu kita mengetahui seberapa baik model yang kita miliki dalam melakukan prediksi dan memberikan informasi. Precision memberikan informasi bahwa bagaimana model test yang sudah memiliki label sudah sesuai dengan prediksi kita. Sementara recall memastikan bahwa seluruh test telah memiliki label

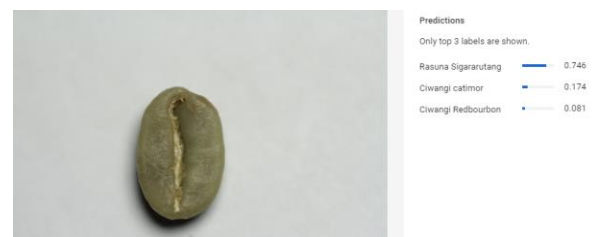


yang benar.

Gambar 5. Metriks google autoML

Nilai threshold memungkinkan untuk mengeksplorasi bagaimana ambang batas skor threshold yang dipilih mempengaruhi nilai precision dan recall. Dalam eksperimen ini digunakan dataset green bean coffee dengan nilai treshold pada 0,5 memiliki precision dan recall yang terukur untuk setiap jenis gambar.

Nilai threshold rata-rata adalah 0.5, sementara precision rata-rata adalah 0.8 dan recall 0.7. Parameter ini digunakan untuk klasifikasi green bean berdasarkan jenis varietasnya. Metrik yang berguna untuk akurasi model adalah area di bawah kurva precision-recall. Ini mengukur seberapa baik kinerja model di semua batas ambang nilai treshold. Pada google AutoML, metrik ini disebut Presisi Rata-Rata. Semakin dekat dengan skor ini, semakin baik kinerja model Anda pada set tes; model menebak secara acak untuk setiap label akan mendapatkan presisi rata-rata sekitar 0,5.



Gambar 6. Tes prediksi green bean kopi

Dalam proses prediksi 10% data yang dimasukkan ke dalam sistem google autoML dengan menggunakan model dataset sendiri. Sistem dalam mengenali dengan presisi sebesar 0.7 sesuai dengan data set yang digunakan untuk training. Walaupun demikian, proses prediksi ini juga memiliki kekurangan dikarenakan data varietas yang terbatas. Artinya jika memasukkan sample data bean berbeda dari ketiga varietas, masih ada kemungkinan terdeteksi sebagai salah satu varietas yang menjadi model.

#### Kesimpulan

Dengan menggunakan dataset model tiga varietas green bean kopi yaitu Ciwangi Catimor, Ciwangi Redbourbon, dan Rasuna Sigararutang. Pada hasil confusion matrix Google autoML dapat mendeteksi varietas ciwangi red bourbon sebesar 71.4%, ciwangi catimor 85.7%, dan rasuna sigararutang 80%. Dengan Nilai threshold rata-rata adalah 0.5, sementara precision rata-rata adalah 0.8 dan recall 0.7. Dalam proses prediksi 10% data yang

dimasukkan ke dalam sistem google autoML dengan menggunakan model dataset sendiri. Sistem dalam mengenali dengan presisi sebesar 0.7 sesuai dengan data set yang digunakan untuk training.

Walaupun demikian, proses prediksi ini juga memiliki kekurangan dikarenakan data varietas yang terbatas. Artinya jika memasukkan sample data bean berbeda dari ketiga varietas, masih ada kemungkinan terdeteksi sebagai salah satu varietas yang menjadi model

### Daftar Pustaka

- [1] M. M. S. Muhammad Agung Nugroho, "Ekstraksi Fitur Circularity untuk Pengenalan Varietas Kopi Arabika," *J. Teknol. Inf. Dan Ilmu Komput. JTIK*, vol. 4, no. 4, pp. 283–289.
- [2] R. H. M. Condori, J. H. C. Humari, C. E. Portugal-Zambrano, J. C. Gutiérrez-Cáceres, and C. A. Beltrán-Castañón, "Automatic classification of physical defects in green coffee beans using CGLCM and SVM," in *Proceedings of the 2014 Latin American Computing Conference, CLEI 2014*, 2014.
- [3] R. G. Apaza, C. E. Portugal-Zambrano, J. C. Gutiérrez-Cáceres, and C. A. Beltrán-Castañón, "An approach for improve the recognition of defects in coffee beans using retinex algorithms," in *Proceedings of the 2014 Latin American Computing Conference, CLEI 2014*, 2014.
- [4] Radi, M. Rivai, and M. H. Purnomo, "Combination of first and second order statistical features of bulk grain image for quality grade estimation of green coffee bean," *ARPN J. Eng. Appl. Sci.*, vol. 10, no. 18, pp. 8165–8174, 2015.
- [5] E. M. De Oliveira, D. S. Leme, B. H. G. Barbosa, M. P. Rodarte, and R. G. F. Alvarenga Pereira, "A computer vision system for coffee beans classification based on computational intelligence techniques," *J. Food Eng.*, vol. 171, pp. 22–27, 2016.
- [6] A. Ferreira and G. Giraldo, "Convolutional Neural Network approaches to granite tiles classification," *Expert Syst. Appl.*, vol. 84, pp. 1–11, 2017.
- [7] H. Chougrad, H. Zouaki, and O. Alheyane, "Deep Convolutional Neural Networks for Breast Cancer Screening," *Comput. Methods Programs Biomed.*, vol. 157, pp. 19–30, 2018.
- [8] C. Affonso, A. L. D. Rossi, F. H. A. Vieira, and A. C. P. de L. F. de Carvalho, "Deep learning for biological image classification," *Expert Syst. Appl.*, vol. 85, pp. 114–122, 2017.
- [9] A. Qayyum, S. M. Anwar, M. Awais, and M. Majid, "Medical image retrieval using deep convolutional neural network," *Neurocomputing*, vol. 266, pp. 8–20, 2017.
- [10] C. Pinto, J. Furukawa, H. Fukai, and S. Tamura, "Classification of Green coffee bean images basec on defect types using convolutional neural network (CNN)," *Proc. - 2017 Int. Conf. Adv. Inform. Concepts Theory Appl. ICAICTA 2017*, 2017.