

PENGUNAAN ALGORITMA RANDOM FOREST DALAM KLASIFIKASI BUAH SEGAR DAN BUSUK

Felix Santoso¹, Ery Hartati²

^{1,2}Universitas Multi Data Palembang

^{1,2}Jurusan Informatika, FIKR UMDP, Palembang

e-mail: ¹felixsantoso21@mhs.mdp.ac.id, ²EryHartati@mdp.ac.id

Abstrak

Buah-buahan merupakan salah satu makanan yang sering dikonsumsi oleh berbagai kalangan umur karena sumber berbagai mineral, vitamin dan serat pangan. Untuk memperoleh manfaat yang terdapat pada buah, masyarakat harus mengonsumsi buah yang segar dan belum busuk. Secara fisik, kesegaran buah dilihat dari ciri pada buah segar maupun buah busuk. LBP (Local Binary Pattern) adalah metode ekstraksi fitur tekstur yang sederhana, namun efisien dalam mempresentasikan ciri tekstur, sedangkan HSV (Hue, Value dan Saturation) merupakan ruang warna yang cocok untuk mengidentifikasi warna-warna dasar yang akan digunakan dalam penelitian sebagai warna identifikasi cahaya dan bisa menoleransi perubahan intensitas cahaya. Penelitian ini menggunakan public dataset buah segar dan buah busuk. Proses di mulai dari resize menjadi ukuran 300 x 300 pixel dan selanjutnya dilakukan ekstraksi fitur LBP dan dilanjutkan dengan ekstraksi fitur HSV. Hasil ekstraksi fitur LBP dan HSV di gunakan sebagai input klasifikasi menggunakan algoritma random forest dengan nilai *n_estimator* 500,1000,1500,dan 2000. Hasil pengujian menggunakan algoritma random forest menghasilkan nilai Accuracy tertinggi sebesar 95,92% dengan nilai *n_estimator* 2000.

Kata kunci : Buah, Local Binary Pattern, Hue, Value dan Saturation, Random Forest

Abstract

Fruits are one of the foods that are often consumed by people various age groups because it is a source of various minerals, vitamins and dietary fiber To get the benefits contained in the fruit, people must consume fresh and not rotten fruit. Physically, the freshness of the fruit can be seen because the signs on fresh fruit or rotten fruit are easy to observe. LBP (Local Binary Pattern) is a texture feature extraction method that is simple, but efficient in presenting texture characteristics, while HSV (Hue, Value and Saturation) is a color space that is suitable for identifying basic colors that will be used in research as light identification colors and can tolerate changes in light intensity. This study uses a public dataset of fresh fruit and rotten. First Procees resized to size of 300 x 300 pixels and then performed LBP feature extraction and continued with HSV feature extraction. The results of LBP and HSV feature extraction are used as input for classification using the random forest algorithm with *n_estimator* values of 500,1000, 1500, and 2000. Test results using the random forest algorithm produce the highest Accuracy value of 95.92% with an *n_estimator* value of 2000.

Keywords : Fruit, Local Binary Pattern, Hue, Value and Saturation,Random Forest.



This is an open-access article under the [CC-BY-CA](#) license

1. PENDAHULUAN

Buah-buahan merupakan sumber berbagai vitamin dan serat pangan. Buah-buahan menjadi jenis makanan yang dikonsumsi oleh berbagai kalangan umur karena memiliki banyak manfaat yang baik untuk kesehatan. Hal ini membuat buah-buahan memiliki tingkat permintaan yang tinggi sehingga menjadi salah satu kelompok komoditas yang penting di Indonesia. Jenis buah yang diminati oleh masyarakat umum seperti pisang, apel dan jeruk memiliki daya saing. Manusia harus mengonsumsi buah segar dari pada buah busuk untuk mendapatkan manfaat buahnya. Ciri fisik buah dapat menentukan kesegaran buah tersebut. Namun, penelitian tentang klasifikasi buah segar dan busuk diperlukan untuk mendapatkan hasil yang lebih tepat [1].

Dewasa ini banyak teknologi dikembangkan untuk mempermudah pekerjaan manusia salah satunya adalah *Computer vision*. *Computer vision focus* pada ilmu yang membuat mesin dapat mengidentifikasi dan mengekstrak informasi dari sebuah objek untuk menyelesaikan suatu masalah [2]. *Image classification* merupakan salah satu fitur yang ada dalam teknologi *computer vision*. Telah banyak dilakukan penelitian mengenai algoritma *image classification* diantaranya K-Means, KNN, *Random Forest*, *Support Vector Machine (SVM)* dan *Convolutional Neural Network (CNN)* [3].

Hue, Saturation, and Value (HSV) adalah alat untuk memodifikasi gambar dalam perangkat lunak pengeditan gambar yang dikembangkan untuk digunakan dalam aplikasi komputer (Madenda, 2015). Tiga komponen HSV adalah *Hue (H)*, *Saturation (S)*, dan Nilai (*V*). Pengukuran jenis warna, seperti merah, kuning, dan sebagainya, disebut *hue (H)*. Diwakili dalam derajat, dari 0 hingga 360. Saturasi warna (*S*) adalah hue-nya. *Value (V)* adalah nilai kecerahan suatu warna, dan semakin berwarna suatu warna, semakin tinggi nilai saturasinya. Nilai lebih tinggi pada warna cerah dari pada warna gelap [4].

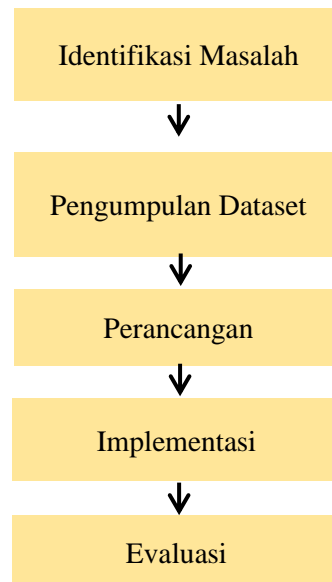
Salah fitur untuk ekstraksi tekstur adalah *Local Binary Pattern (LBP)*. LBP adalah fitur ekstraksi untuk mengidentifikasi tekstur yang sederhana, namun efisien dalam mempresentasikan ciri tekstur. Dengan perhitungan sederhana, operator LBP terdiri dari beberapa piksel tetangga. Karena LBP menggambarkan tekstur secara lokal, ini juga merupakan metode yang invarian skala abu-abu atau tidak terpengaruh oleh pencahayaan gambar yang tidak merata.[5].

2. PENELITIAN TERKAIT

Beberapa Penelitian Terhadap Buah telah dilakukan seperti klasifikasi citra buah dengan klasifikator SVM menghasilkan akurasi sebesar 94% [6], klasifikasi citra jenis daging menghasilkan akurasi 87,5% [7], *Automatic classification for fruits' types and identification of rotten ones using k-NN and SVM* menghasilkan akurasi 98% [8] dan implementasi SVM untuk klasifikasi jenis buah menghasilkan akurasi 99,120% [9]. Serta penelitian terkait seperti Klasifikasi Buah Segar dan Busuk Menggunakan Ekstraksi Fitur Hu-Moment, Haralick dan Histogram menggunakan algoritma *Random Forest* mendapatkan akurasi 99,6% [1], Implementasi *Random Forest* Untuk Klasifikasi Motif Songket Palembang Berdasarkan SIFT mendapatkan rata-rata akurasi 92% [10].

3. METODE PENELITIAN

Tahapan-tahapan yang dilakukan oleh peneliti dalam melakukan penelitian ini digambarkan dalam bentuk diagram kerangka metode penelitian yang ditunjukkan pada gambar 1.



Gambar 1 Kerangka Kerja Penelitian

3.1. Identifikasi Masalah

Tahap ini dimulai dengan mencari informasi dari jurnal, buku, artikel dan teori yang terkait dengan klasifikasi buah segar dan busuk. Kemudian mengumpulkan literatur sebagai referensi relevan, guna memberikan kontribusi penelitian yang baru

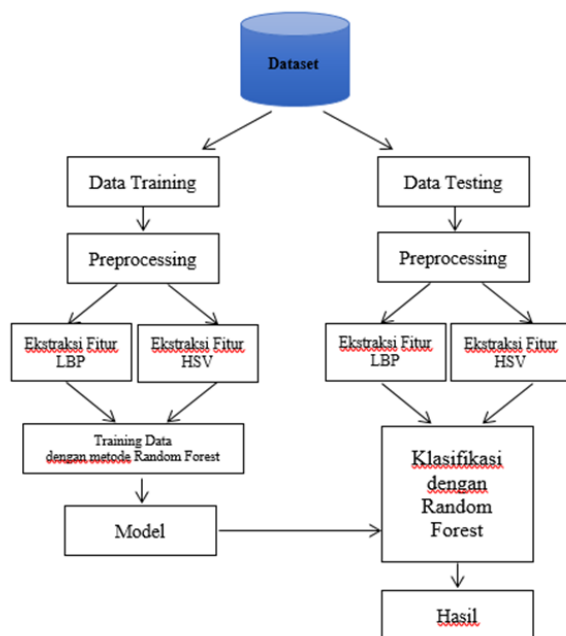
3.2. Pengumpulan Data

Pada tahap ini, dataset yang dikumpulkan berupa objek gambar buah-buahan menggunakan public dataset yang terdapat pada *website* kaggle (<https://www.kaggle.com/sriramr/fruits-fresh-and-rotten-for-classification>). Dari dataset yang tersedia terdapat sample kelas citra buah apel, pisang, dan jeruk. Dataset berjumlah 13.599 gambar yang terbagi menjadi 10.901 data train dan 2.696 data test yang memiliki dimensi gambar yang beragam.

3.3. Perancangan Sistem

Setelah mengumpulkan dataset, maka dilakukan preprocessing. *Pre-processing* adalah proses mempersiapkan citra sebelum masuk pada proses segmentasi. Proses ini meliputi proses *cropping* pada citra. Setelah semua citra memiliki dimensi yang sama, maka langkah selanjutnya adalah melakukan ekstraksi fitur dengan melihat ciri pada citra buah, yang menunjukkan karakteristik buah segar dan busuk. Fitur yang di ambil pada citra buah adalah fitur tekstur dan fitur warna. Pada proses pelatihan (*training*), citra buah akan dilakukan tahap ekstraksi fitur tekstur LBP dan HSV. Hasil dari proses ekstraksi akan digunakan untuk proses traning yang menggunakan algortima *Random Forest*. Setelah proses *training* , maka akan didapatkan data model yang akan digunakan untuk pengujian data. Pada proses pengujian (*testing*), citra buah apel, pisang dan jeruk akan dilakukan tahap ekstraksi fitur tekstur LBP dan HSV. Hasil dari citra akan digunakan untuk

proses pengujian. Setelah pengujian selesai, maka akan didapatkan data yang mendekati atau sama dengan model yang sudah dibuat. Skema perancangan yang sudah dijelaskan dapat dilihat pada Gambar 2 Perancangan sistem.



Gambar 2. Perancangan Kerja Penelitian

3.4. Implementasi

Tahap ini mengimplementasikan perancangan sistem yang sudah dibuat menggunakan bahasa pemrograman Python.

3.4.1. *Resize Citra*

Pada tahap ini, seluruh dimensi citra RGB pada dataset akan di-*resize* menjadi ukuran 300 x 300 *pixel*.

3.4.2. Ekstraksi Fitur Tekstur LBP

Pada tahap ini, dilakukan ekstraksi fitur tekstur dari citra RGB yang sudah di-*resize* menjadi 300 × 300 *pixel* menjadi citra *grayscale*. Selanjutnya citra *grayscale* akan dikonversi menjadi citra LBP dengan nilai *feature vector*, dimana parameter yang digunakan adalah :

$$\begin{aligned} \text{SizeImage} &= 300 \times 300 \\ \text{CellSize} &= 300 \times 300 \\ \text{numNeighbors} &= 8 \end{aligned}$$

Nilai feature vector LBP akan dihitung dengan Persamaan (1), (2) dan (3) :

$$\begin{aligned} \text{numCells} &= \text{prod}(\text{SizeImage}/\text{CellSize}) & (1) \\ \text{numBins} &= \text{numNeighbors} \times (\text{numNeighbors} - 1) + 3 & (2) \\ N &= \text{numCells} \times \text{numBins} & (3) \end{aligned}$$

Maka,

$$\begin{aligned} \text{numCells} &= \text{prod}(300 \times 300 / 300 \times 300) = 1 \\ \text{numBins} &= 8 \times (8 - 1) + 3 = 59 \\ N &= 1 \times 59 \end{aligned}$$

Sehingga didapatkan nilai feature vector (N) untuk LBP adalah 1×59 .

3.4.3. Ekstraksi Fitur Warna HSV

Pada tahap ini, dilakukan ekstraksi fitur warna dari citra RGB yang sudah di *resize* menjadi 300×300 *pixel* menjadi citra HSV. Untuk *feature vector* HSV didapatkan dari nilai statistik yang meliputi mean dari seluruh nilai mean, modus dari seluruh nilai modus dan standar deviasi dari seluruh dari nilai standar deviasi yang didapatkan dari nilai citra HSV.

3.4.4. Klasifikasi dengan *random forest*

Hasil dari *feature vector* LBP dan HSV digabungkan sebagai input klasifikasi menggunakan algoritma *random forest* dengan *n_estimator* 500,1000,1500,2000 untuk mendapatkan hasilnya.

3.5. Pengujian

Setelah tahap implementasi selesai dilakukan, hasil klasifikasi akan dihitung untuk mendapatkan tingkat keberhasilan dari metode yang sudah ditentukan, dimana akan menghitung nilai *Precision*, *Recall*, *Accuracy*.

$$Presision = \frac{TP}{TP+FP} \quad (4)$$

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

Keterangan :

TP = Total citra buah positif yang diklasifikasi benar.

TN = Total citra buah negatif yang diklasifikas benar.

FN = Total citra buah negatif yang diterklasifikasi salah.

FP = Total citra buah positif yang diterklasifikasi salah.

4. HASIL DAN PEMBAHASAN

4.1. Implementasi *Preprocessing*

Pada tahap ini, dilakukan proses *preprocessing* pada setiap citra data *training* dan data *testing*. di- *resize* menjadi ukuran 300×300 piksel. Ukuran tersebut dipilih agar citra yang dipakai lebih fokus.

4.2. Implementasi Ekstraksi LBP & HSV

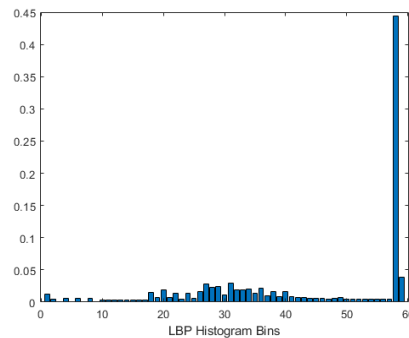
Pada tahap ini, dilakukan ekstraksi fitur tekstur dari citra RGB yang sudah di-*resize* menjadi 300×300 *pixel* (a) menjadi citra grayscale (b) yang ditampilkan pada Gambar Selanjutnya citra



(a) Citra RGB. (b) Citra

Gambar 3. Transformasi *Grayscale*

grayscale (Gambar 3 (b)) akan dikonversi menjadi citra LBP dan menampilkan histogram 59 bins yang dapat dilihat pada Gambar 4.

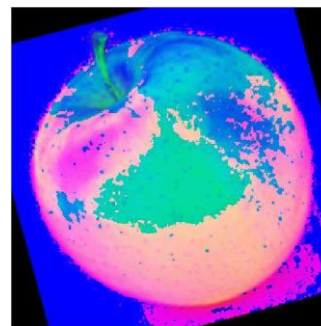


Gambar 4 Histogram LBP

Pada tahap ini, dilakukan ekstraksi fitur warna dari citra RGB yang sudah *resize* menjadi 300×300 pixel (a) menjadi citra HSV (b) yang dapat dilihat pada Gambar 5.



(a) Citra RGB



(b) Citra HSV

Gambar 5 Transformasi Warna HSV

4.3. Implementasi Random Forest

Pada tahap ini, nilai *feature vector* LBP dengan ukuran 1×59 digabungkan dengan *feature vector* HSV dengan ukuran 1×9 , maka akan didapatkan nilai *feature vector* 1×68 . Selanjutnya hasil penggabungan nilai *feature vector* tersebut dilakukan sebagai input klasifikasi menggunakan *random forest* dengan $n_estimator = 500, 1000, 1500$ dan 2000 .

3.4 Pengujian

Pada tahap ini, pengujian data uji terhadap data latih dan hasil dari pengujian tersebut akan ditampilkan dalam bentuk *confusion matrix*. pengujian untuk klasifikasi buah segar dan busuk menggunakan metode *random forest* dengan nilai $n_estimator = 500, 1000, 1500, 2000$. Hasil Pengujian *Precision*, *Recall* dan *Accuracy* dapat di lihat pada Tabel 1.

Tabel 1 Hasil Pengujian

Kelas	N_estimator	Precision	Recall	Accuracy
Segar	500	95,03%	96,19%	95,77%
	1000	95,26%	95,02%	95,81%
	1500	95,27%	95,10%	95,85%
	2000	95,43%	95,10%	95,92%
Busuk	500	96,34%	96,22%	95,77%
	1000	96,22%	96,41%	95,81%
	1500	96,29%	96,41%	95,85%
	2000	96,29%	96,54%	95,92%

Berdasarkan Tabel 1, diketahui bahwa nilai accuracy tertinggi pada kelas segar sebesar 95,92% pada n_estimator 2000 dan nilai accuracy tertinggi pada kelas busuk sebesar 95,92% pada n_estimator 2000. Sedangkan untuk nilai accuracy terendah pada kelas segar sebesar 95,77% pada n_estimator 500 dan nilai accuracy terendah pada kelas busuk sebesar 95,77% pada n_estimator 500.

5. KESIMPULAN

1. Algoritma random forest dengan penggunaan fitur ekstraksi tekstur LBP dan fitur ekstraksi warna HSV dapat diterapkan untuk klasifikasi buah segar dan busuk.
2. Berdasarkan hasil pengujian terhadap kelas segar dan kelas busuk menggunakan algoritma *random forest* mendapatkan nilai *Accuracy* tertinggi pada kelas segar sebesar 95,27% pada n_estimator 1000 dan nilai *Accuracy* tertinggi pada kelas busuk sebesar 96,54% pada n_estimator 2000.

6. SARAN

1. Menerapkan fitur ekstraksi yang berbeda dengan metode klasifikasi yang sama dengan penelitian terkait.
2. Menerapkan fitur ekstraksi yang sama dengan metode klasifikasi berbeda dari penelitian terkait.

DAFTAR PUSTAKA

- [1] Cahaya, F. N., Pebrianto, R., & M, T. A. (2021). Klasifikasi Buah Segar dan Busuk Menggunakan Ekstraksi Fitur Hu-Moment , Haralick dan Histogram. *IJCIT*, 6(1), 57–62. <https://creativecommons.org/licenses/by-sa/4.0/>.
- [2] Feriawan, J., & Swanjaya, D. (2020). Perbandingan Arsitektur Visual Geometry Group dan MobileNet Pada Pengenalan Jenis Kayu. Seminar Nasional Inovasi Teknologi, 185-190. <https://proceeding.unpkediri.ac.id/index.php/inotek/article/view/84>.
- [3] Stephen, Raymond, & Santoso, H. (2019). Aplikasi Convolution Neural Network Untuk Mendeteksi Jenis-Jenis Sampah. *Jurnal Sistem Informasi Dan Telematika*, 10(2), 122–130.
- [4] Priyanto. (2017). *Pengolahan Citra Digital Teori dan Aplikasi Nyata*. Informatika Bandung.
- [5] Ojala, T., Pietikäinen, M., & Mäenpää, T. (2000). Gray scale and rotation invariant texture classification with local binary patterns. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 1842, 404–420. https://doi.org/10.1007/3-540-45054-8_27.
- [6] Meiriyama, M. (2018). Klasifikasi Citra Buah berbasis fitur warna HSV dengan klasifikator SVM. *Jurnal Komputer Terapan*, 4(1 SE-), 50–61. <https://jurnal.pcr.ac.id/index.php/jkt/article/view/1678>.
- [7] Neneng, N., Adi, K., & Isnanto, R. (2016). Support Vector Machine Untuk Klasifikasi Citra Jenis Daging Berdasarkan Tekstur Menggunakan Ekstraksi Ciri Gray Level Co-Occurrence Matrices (GLCM). *Jurnal Sistem Informasi Bisnis*, 6(1), 1. <https://doi.org/10.21456/vol6iss1pp1-10>
- [8] Nosseir, A., & Ashraf Ahmed, S. E. (2019). Automatic classification for fruits' types and identification of rotten ones using k-NN and SVM. *International journal of online and biomedical engineering*, 15(3), 47–61. <https://doi.org/10.3991/ijoe.v15i03.9832>.
- [9] Yohannes, Y., Pribadi, M. R., & Chandra, L. (2020). Klasifikasi Jenis Buah dan Sayuran Menggunakan SVM Dengan Fitur Saliency-HOG dan Color Moments. *Elkha*, 12(2), 125. <https://doi.org/10.26418/elkha.v12i2.42160>
- [10] Devella, S., Yohannes, & Rahmawati, F. N. (2020). Implementasi Random Forest Untuk Klasifikasi Motif Songket Palembang Berdasarkan SIFT. *Jurnal Teknik Informatika dan*