

IJCIT

(Indonesian Journal on Computer and Information Technology)

Journal Homepage: <http://ejournal.bsi.ac.id/ejurnal/index.php/ijcit>

Klasifikasi Buah Segar dan Busuk Menggunakan Ekstraksi Fitur *Hu-Moment*, *Haralick* dan *Histogram*

Fani Nurona Cahya¹, Rangga Pebrianto², Tika Adilah M³

Magister Ilmu Komputer, STMIK Nusa Mandiri

Jakarta, Indonesia

e-mail: 14002395@nusamandiri.ac.id¹, 14002396@nusamandiri.ac.id², 14002364@nusamandiri.ac.id³

ABSTRAK

Banyak teknologi yang sudah ditemukan, salah satunya adalah pengolahan citra digital. Identifikasi pada sebuah citra sudah lama dikembangkan, salah satunya dengan membedakan tekstur pada citra tersebut. Tekstur citra dapat dibedakan oleh kerapatan, keseragaman, kekasaran dan keteraturan dari citra yang diteliti. *Agriculture* saat ini sedang ramai dibahas khususnya di Indonesia. Banyak sekali penelitian yang dilakukan dalam sektor pertanian guna memajukan sektor pertanian itu sendiri. Dalam penelitian kali ini yaitu ekstraksi fitur menggunakan *Hu-moment*, *Haralick* dan *Histogram* dan klasifikasi menggunakan algoritma *Random Forest*. Peneliti mencoba mengklasifikasi buah-buahan segar atau busuk, dengan algoritma yang digunakan yaitu algoritma *Random Forest*. Penelitian ini mendapatkan akurasi yang sangat tinggi yakni 99.6% sangat baik sekali. Namun guna memperbaharui penelitian bisa mencoba beberapa fitur dan algoritma yang berbeda agar mendapatkan perbandingan atau hasil yang lebih maksimal.

Kata Kunci: ekstraksi fitur, *hu-moment*, *haralick* dan *histogram*, *random forest*.

ABSTRACTS

Various technologies have been discovered, one of which is digital image processing. Identification in an image has been developed for a long time, one of which is by distinguishing the texture of the image. Image texture can be distinguished based on the density, uniformity, roughness and regularity of the image under study. Agriculture is currently a topic of discussion especially in Indonesia. A lot of research has been carried out in the agriculture sector to propose the agriculture sector itself. In this study, feature extraction using Hu-Moment, Haralick and Histogram and classification using the Random Forest algorithm. Researchers tried to classify fresh or rotten fruit, with the algorithm used, the Random Forest algorithm. This Research obtained a very high accuracy of 99.6% which is very good. But to update your research, you can try several different features and algorithms to get the best comparison or result.

Keywords: feature extraction, *hu-moment*, *haralick* and *histogram*, *random forest*

1. PENDAHULUAN

Buah-buahan merupakan salah satu kelompok komoditas pertanian yang penting di Indonesia. Buah-buahan memiliki tingkat permintaan yang tinggi. Perminatan domestik terhadap komoditas buah-buahan cukup tinggi, ditandai dengan banyaknya buah-buahan yang ada di pasar modern maupun tradisional Indonesia. Buah apel, pisang dan jeruk merupakan salah satu jenis buah yang ada di Indonesia dan sangat di gemari oleh masyarakat

umum, baik muda sampai tua suka mengkonsumsi. Hal ini menunjukkan bahwa buah apel, pisang dan jeruk sudah sangat dikonsumsi masyarakat secara luas dan memiliki daya saing juga. Tingginya tingkat produksi dan distribusi buah-buahan tersebut yang sangat luas mengharuskan para petani mampu mengklasifikasikan tingkat kesegaran buah-buahan tersebut dan mengurangi tingkat resiko pembusukan.



Mutu buah apel, pisang dan jeruk yang baik sangat ditentukan oleh tingkat ketuaan buah dan penampakannya, secara fisik sebenarnya mudah dilihat karena tanda-tanda yang ada pada buah segar atau busuk mudah diamati. Namun untuk mendapatkan hasil yang lebih akurat perlu dilakukan penelitian lebih lanjut mengenai klasifikasi buah yang segar dan busuk. Proses klasifikasi kematangan buah sangat penting, klasifikasi kematangan buah saat ini kebanyakan masih menggunakan metode manual, yaitu penilaian subjektif dari petani. Kelemahan dari metode ini adalah menghasilkan produk yang beragam karena adanya keterbatasan visual manusia, tingkat kelelahan dan perbedaan persepsi tentang mutu buah.

Seiring dengan perkembangan zaman, teknologi berkembang dengan pesat saat ini. Dengan perkembangan teknologi sekarang ini memudahkan semua orang mengakses apa saja. Banyak teknologi yang sudah ditemukan salah satunya adalah pengolahan citra digital. Identifikasi pada sebuah citra sudah lama dikembangkan salah satunya dengan membedakan tekstur pada citra tersebut. Tekstur citra dapat dibedakan oleh kerapatan, keseragaman, kekasaran dan keteraturan (Y. Garis K, I. Santoso, 2011). Sebuah komputer tidak dapat membedakan tekstur seperti halnya penglihatan manusia, maka digunakan analisis tekstur untuk mengetahui pola dari suatu citra digital. Analisis tekstur akan menghasilkan nilai dari ciri atau karakteristik tekstur yang nilai dari ciri atau karakteristik tekstur yang kemudian dapat diolah sebuah komputer untuk proses klasifikasi (I. Permatasari dan T. Sutojo, 2016). Pengertian klasifikasi yaitu proses pengelompokan, artinya memisahkan benda/entitas yang tidak sama. Klasifikasi menganalisis sifat numerik gambar fitur dan mengatur data ke dalam berbagai kategori (Syahid et al., 2016).

Pada penelitian kali ini akan melakukan proses pengklasifikasian antara buah segar dan busuk menggunakan Ekstraksi Fitur Warna menggunakan *Hu-moment*, *Haralick* dan *Histogram*, untuk mengetahui tingkat akurasi menggunakan algoritma *random forest*. Model warna dalam citra digital telah banyak dikembangkan oleh para ahli seperti model warna *RGB*. Penelitian ini mencoba mengangkat permasalahan baru dalam klasifikasi buah apel, pisang dan jeruk, yaitu mengklasifikasikan perbedaan buah segar atau busuk pada buah

apel, pisang dan jeruk yang direpresentasikan dari sisi tekstur dan warnanya. Manfaat dari penelitian ini yaitu dapat menghemat waktu dan memudahkan dalam pengklasifikasian sesuai dengan yang diinginkan oleh petani buah. Pengklasifikasian juga bisa diperoleh pada waktu yang tepat dan akurasi yang lebih tinggi oleh metode yang dilakukan pada penelitian ini.

Sektor pertanian terus menjadi instrumen fundamental pembangunan berkelanjutan dan pengentasan kemiskinan di abad ke-21, pertanian berkontribusi pada pembangunan sebagai kegiatan ekonomi. Setiap pembangunan pertanian harus dilakukan dengan memindahkan pekerja pertanian (sektor tradisional) ke sektor non pertanian (sektor modern) (Moeis et al., 2020).

Menurut (Sapto, 2008) Tantangan globalisasi hanya dapat dimanfaatkan dengan meningkatkan daya saing. Daya saing yang tinggi dicirikan dengan tingginya produktivitas, mutu dan efisiensi usaha. Sementara itu pesatnya pengembangan teknologi dan derasnyanya arus informasi teknologi, harus dapat ditangkap dan dimanfaatkan untuk pengembangan sistem usaha agribisnis.

Penelitian sebelumnya oleh (Bhargava & Bansal, 2020) menelaskan tentang klasifikasi berbagai jenis buah dan identifikasi grading buah merupakan tantangan yang memberatkan akibat produksi massal produk buah. Untuk membedakan dan mengevaluasi kualitas buah secara lebih tepat, penelitian tersebut menyajikan sistem yang membedakan empat jenis buah dan menganalisis peringkat buah berdasarkan kualitasnya. Pertama, algoritme mengekstrak nilai merah, hijau, dan biru dari gambar dan kemudian latar belakang gambar dipisahkan dengan algoritme *split-and-merge*. Selanjutnya, beberapa fitur (30 fitur) yaitu fitur warna, statistik, tekstur, dan geometri diekstraksi. Untuk membedakan jenis buah, hanya fitur geometri (12 fitur), fitur lain digunakan dalam evaluasi kualitas buah. Selanjutnya empat pengklasifikasi berbeda *k-neighbourhood (k-NN)*, *support vector machine (SVM)*, *sparse representative classifier (SRC)*, dan *artificial neural network (ANN)* digunakan untuk mengklasifikasikan kualitas.

Penelitian terkait selanjutnya pernah dilakukan oleh (Rahmad et al., n.d.) dalam penelitiannya mencoba mengidentifikasi tingkat kesegaran buah secara cepat dengan menggunakan mikrokontroler arduino uno dan ditambah dengan beberapa komponen seperti

sensor *TCS3200*, sensor *DHT11*, *LCD 16x2*, *Arduino IDE*, Regulator tegangan dan beberapa komponen lainnya yang dirancang untuk dapat menentukan kesegaran buah.

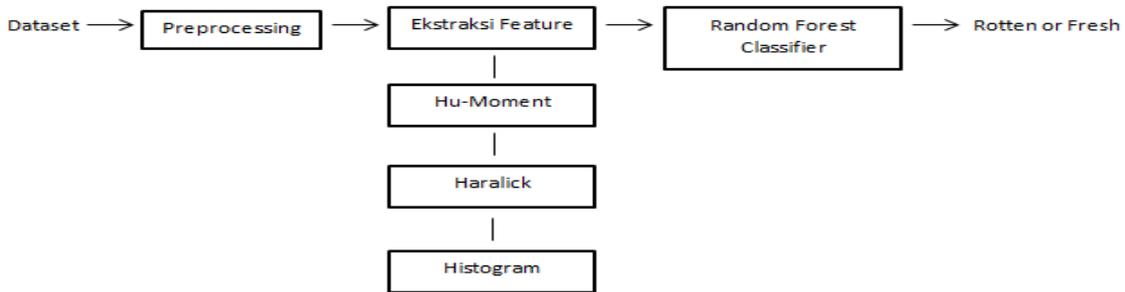
Penelitian terkait ketiga yang pernah dilakukan oleh (Ananthanarayana et al., 2020) menunjukkan hasil awal pada model deteksi objek untuk enam kategori buah yang sama. 000 gambar *RGB* dari enam kategori buah. Model ini mencapai akurasi 97% pada *platform edge* menggunakan kamera *13 MP ON Semiconductor* dengan sensor *AR1335*. Selain model klasifikasi gambar, pekerjaan sedang berlangsung untuk meningkatkan akurasi deteksi objek menggunakan *SSD* dan *SSDLite* dengan *MobileNetV2* sebagai ekstraktor fitur. Dalam makalah ini, kami menunjukkan hasil awal pada model deteksi objek untuk enam kategori buah yang sama. 000 gambar *RGB* dari enam kategori buah. Model ini mencapai akurasi 97% pada *platform edge* menggunakan kamera *13 MP ON Semiconductor* dengan sensor *AR1335*. Selain model klasifikasi gambar, pekerjaan sedang berlangsung untuk meningkatkan akurasi

deteksi objek menggunakan *SSD* dan *SSDLite* dengan *MobileNetV2* sebagai ekstraktor fitur.

Dalam penelitian ini, kami menunjukkan hasil awal pada model deteksi objek untuk enam kategori kelas buah segar dan busuk, kami mencoba memberikan kebaruan dengan nilai akurasi yang lebih tinggi dengan cara mencoba tiga *feature extraction* dan metode yang berbeda yaitu *Random Forest* dengan hasil akurasi 99,6% lebih tinggi dari penelitian terkait sebelumnya yang menggunakan dataset sama.

2. METODE PENELITIAN

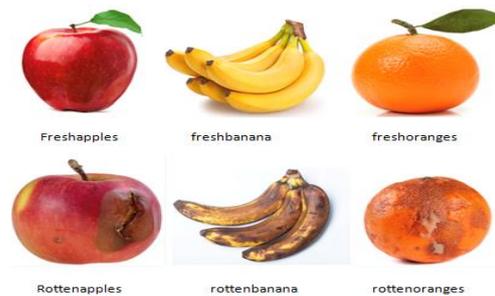
Untuk melakukan penelitian kali ini yang pertama harus dilakukan adalah pengumpulan dataset sebagai data latih dan data uji. Kemudian Segmentasi dan ekstraksi ciri dilakukan pada masing-masing citra latih dan citra uji. Kemudian untuk pengenalan buah segar atau busuk berdasarkan informasi ciri tersebut menggunakan ekstraksi fitur *Hu-moment*, *Harlick* dan *Histogram*. Dan untuk metode yang di pakai dalam menentukan akurasi adalah metode *Random Forest*.



Gambar 1. Metode Penelitian

Dataset ini melalui beberapa seleksi di antara beberapa dataset yang tersedia di *Kaggle.com*. alasan memilih dataset ini karena penelitian ini baru beberapa kali di kembangkan sehingga peneliti ingin mencoba untuk mengembangkannya. Penelitian ini memaparkan tentang klasifikasi perbedaan buah segar dan busuk dengan sample buah apel, pisang dan jeruk. Dataset ini berjumlah 10.901 dataset yang terbagi dari 6 kelas. Penelitian ini memaparkan klasifikasi perbedaan buah segar dan busuk dengan sample buah apel, pisang dan jeruk. Dimana pengklasifikasian ini akan sangat bermanfaat dan membantu petani juga konsumen buah. Masyarakat bisa mengetahui kualitas buah yang akan di konsumsi, apakah masih segar atau sudah busuk. Dari dataset yang

tersedia terdapat sample kelas citra yang tersedia dari buah apel, pisang dan jeruk. Dari jumlah dataset 10.901 terdapat 1.693 citra buah apel segar, 1.581 citra pisang segar, 1.466 citra jeruk segar, 2.342 citra buah apel busuk, 2.224 citra buah pisang busuk, dan 1.595 citra jeruk busuk.



Gambar 2. Contoh Dataset

Pre-processing Merupakan serangkaian proses untuk mempersiapkan citra sebelum proses segmentasi. Proses ini meliputi proses *cropping* dan proses *konversi* ruang warna. Proses ekstraksi fitur dilakukan proses pengambilan ciri dari citra buah apel, pisang dan jeruk yang menggambarkan karakteristik buah segar dan busuk. Fitur yang di ambil dari citra buah yang telah tersegmentasi adalah fitur warna, fitur luas dan fitur bentuk. Proses klasifikasi tingkat keparahan dilakukan menggunakan klasifikasi *Random Forest* (Ratnawati & Sulistyaningrum, 2019).

Ekstraksi fitur adalah salah satu tugas penting yang meningkat efesien seluruh sistem. Fitur menjelaskan beberapa properti yang dapat dihitung dari gambar masukan yang diberikan di dalam proses. Ekstraksi fitur menggunakan data citra hasil tahap segmentasi citra. Tahap ekstraksi fitur ini menggunakan fitur warna, luas dan bentuk.

pada (Zhang et al., 2020) Deskriptor Hog dapat dibuat dengan menghitung histogram kemiringan daerah setempat gambar. *Deskriptor* pertama-tama membagi gambar menjadi sel-sel kecil, dan menghitung gradien dan piksel di setiap sel. Dan kemudian, *HOG* menggabungkan banyak sel menjadi blok piksel. *Gradien horizoontal* dan vertikal dari setiap piksel abu-abu gambar (x, y) dapat dihitung menurut berikut ini, rumus :

$$G_x(x, y) = I(x + 1, y) - I(x - 1, y) \quad (7)$$

$$G_y(x, y) = I(x, y + 1) - I(x, y - 1) \quad (8)$$

Dimana $G_x(x, y)$, $G_y(x, y)$ mewakili *gradien* dalam arah horizontal dan vertikal. Berikutnya adalah besaran fradien dan arah gradien dihitung dengan persamaan. 9 dan persamaan 10.

$$\nabla G(x, y) = \sqrt{G_x(x, y)^2 + G_y(x, y)^2} \quad (9)$$

$$\theta(x, y) = \tan^{-1} \left(\frac{G_y(x, y)}{G_x(x, y)} \right) \quad (10)$$

Haralick, Fitur ini menceritakan korelasi intensitas antara kedua piksel dalam gambar pada jarak tertentu yang di tetapkan arah. *GLCM* tidak hanya menawarkan hubungan antar-piksel tetapi juga ketergantungan dan priodisitas tingkat abu-abu spasial (Gayathri et al., 2020).

Algorithm 1 Haralick Feature Extraction

Require: Load I with input images

START:

- 1) consider each image in I with pixel r_k location (s, t)
- 2) **Do for:**
- 3) compute the GLCM $G_{s,t}^k$ of current pixel r_k in a neighborhood of size $m \times m$
- 4) Extract Haralick features for each $G_{s,t}^k$
- 5) **end for**

Output: Haralick features for each GLCM

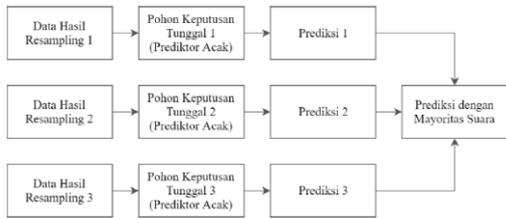
Gambar 3. *Algorithm Haralick*

Hu-Moment dapat mengurangi langkah-langkah pra-pemrosesan seperti normalisasi ukuran dan posisi pemutusan gambar yang akan dideteksi. Karena fitur invarian lokal digunakan, ini membuat pengenalan dan deteksi objek ditangani dengan lebih baik keadaan yang rumit ini. Ini adalah poin utama dari visi komputer saat ini dan memang begitu terutama digunakan untuk mendeskripsikan posisi hubungan piksel dalam sebuah gambar serta fitur deskripsi objek, termasuk tepi gambar, kontur dan sudut kemudian digabungkan dan diubah fitur-fitur ini sebelum membentuk deskriptor fitur stasioner yang mudah untuk dicocokkan (Wu et al., 2020).

Metode *Random Forest* merupakan metode klasifikasi yang dilakukan dengan mengembangkan metode *Decision Tree* berdasarkan pemilihan atribut acak pada setiap node untuk menentukan klasifikasi pada proses klasifikasinya didasarkan pada suara terbanyak dari pohon keputusan yang dikembalikan. *Random Forest* dapat dibangun dengan menggunakan bagging dengan pemilihan atribut acak. Metode *CART (classification an Regression Tree)* digunakan untuk menumbuhkan pohon keputusan. Pohon keputusan tersebut tumbuh hingga ukuran maksimum dan tidak akan di pangkas sehingga dihasilkan kumpulan pohon yang kemudian disebut hutan (*forest*) (Gorunescu, 2011). *Random Forest* adalah suatu metode klasifikasi yang terdiri dari kumpulan terstruktur pohon keputusan dimana vektor acak *independen* didistribusikan secara identik dan setiap pohon keputusan memberikan suara unit untuk kelas paling populer pada masukan x , seperti pada gambar 2 (E. Goel and E. Abhilasha, 2017).

Dalam klasifikasi *Random Forest* setiap pohon keputusan yang ditumbuhkan berdasarkan *resampling* data akan membawa eror tersendiri pada setiap pohon keputusan. Perhitungan eror yang diperoleh pada setiap

pohon keputusan yang dibangun merupakan hasil dari *misclassified observations*.



Gambar 4. Algoritma sederhana *Random Forest*

3. HASIL DAN PEMBAHASAN

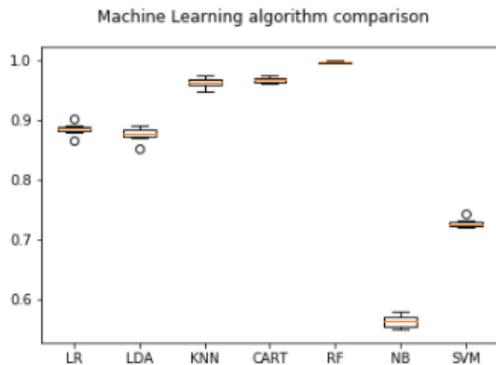
Hasil eksperimen dengan menggunakan tiga *feature extraction color histogram, Hu_moment* dan *Haralick*.

1) Nilai Akurasi

Tabel 1. Nilai Akurasi

Algoritma	Nilai Akurasi
LR	88.4%
LDA	87.6%
KNN	96.1%
CART	96.7%
RF	99.6%
NB	56.2%
SVM	72.6%

2) *Boxplot Metode*



Gambar 5. *Boxplot Metode*

Dari hasil komparasi dengan menggunakan tiga *feature extraction color histogram, Hu_moment* dan *Haralick* bahwa metode *Random Forest* merupakan metode terbaik karena menghasilkan *Accuracy* paling tinggi dengan nilai *accuracy* 99.6%.

3) *Confusion Matrix*

Confusion Matrix merepresentasikan prediksi dan kondisi sebenarnya (aktual) dari data yang dihasilkan oleh algoritma ML.

Berdasarkan *Confusion Matrix*, kita bisa menentukan *Accuracy, Precision, Recall* dan *Specificity*. *Confusion Matrix* dari penelitian terlihat pada tabel 2.

Tabel 2. *Confusion Matrix*

490	3	3	7	0	0
0	469	1	2	1	0
0	1	454	0	0	4
4	6	5	656	3	0
0	1	0	5	649	0
0	0	1	6	0	647

4) Nilai *Kappa statistic, precision tiap class, recall tiap class*

Precision adalah tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem. Sedangkan *recall* adalah tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi. Dan *Recall* (*Sensitifitas*) Merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif. Hasil nilai *Kappa statistic, precision tiap class, recall tiap class* terlihat pada tabel 5.

Tabel 5. Nilai *Kappa statistic, precision tiap class, recall tiap class*

<i>Kappa statistic</i>	0.9862126569296227
<i>Precision tiap class</i>	[0.99598394
	0.97319588
	0.98701299
	0.98652695
	0.99092284
	0.99784483]
<i>recall tiap class</i>	[0.9860835
	0.99788584
	0.99346405
	0.97774481
	1.
	0.97679325]

4. KESIMPULAN

Berdasarkan hasil uji coba dan pembahasan dapat diambil kesimpulan bahwa penelitian ini telah berhasil melakukan pengukuran identifikasi buah segar dan busuk pada buah apel, pisang dan jeruk dengan menerapkan *Random Forest* melalui beberapa proses, antara lain akuisisi citra, pra-pemrosesan citra, segmentasi citra menggunakan *Hu-moment, Haralick* dan *Histogram*, ekstraksi fitur warna, luas dan

bentuk serta klasifikasi menggunakan *Random Forest*. Ekstraksi fitur menggunakan *Hu-moment*, *Haralick* dan *Histogram* yang mendapatkan hasil akurasi sangat tinggi yaitu dengan menggunakan *algoritma Random Forest* yakni 99.6%. tingkat performa *Random Forest* untuk identifikasi buah segar dan busuk sangat baik digunakan. Untuk penelitian selanjutnya bisa mencoba menggunakan beberapa fitur dan *algorithma* yang berbeda.

5. REFERENSI

- Ananthanarayana, T., Ptucha, R., & Kelly, S. C. (2020). Deep Learning based Fruit Freshness Classification and Detection with CMOS Image sensors and Edge processors. *Electronic Imaging*, 2020(12), 172-1-172-177. <https://doi.org/10.2352/issn.2470-1173.2020.12.fais-172>
- Bhargava, A., & Bansal, A. (2020). Automatic Detection and Grading of Multiple Fruits by Machine Learning. *Food Analytical Methods*, 13(3), 751-761. <https://doi.org/10.1007/s12161-019-01690-6>
- E. Goel and E. Abhilasha. (2017). Random Forest: A Review. *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, Vol. 7, no.
- Gayathri, S., Krishna, A. K., Gopi, V. P., & Palanisamy, P. (2020). Automated Binary and Multiclass Classification of Diabetic Retinopathy Using Haralick and Multiresolution Features. *IEEE Access*, 8, 57497-57504. <https://doi.org/10.1109/ACCESS.2020.2979753>
- Gorunescu, F. (2011). *Data Mining concepts, models and techniques*.
- I. Permatasari dan T. Sutojo. (2016). *Pengenalan Ciri Garis Telapak Tangan Menggunakan Ekstraksi Fitur (GLCM) dan Metode (KNN)*.
- Moeis, F. R., Dartanto, T., Moeis, J. P., & Ikhsan, M. (2020). A longitudinal study of agriculture households in Indonesia: The effect of land and labor mobility on welfare and poverty dynamics. *World Development Perspectives*, 20(August 2019), 100261. <https://doi.org/10.1016/j.wdp.2020.100261>
- Rahmad, I. F., Gunawan, D. I., Teknik, J., Universitas, I., Utama, P., Jurusan, D., Informatika, T., Potensi, U., & Utama, U. P. (n.d.). *Perancangan Dan Implementasi Alat Pendeteksi Kesegaran Buah Berbasis Arduino*. 1(1), 368-379.
- Ratnawati, L., & Sulistyaningrum, D. R. (2019). Penerapan *Random Forest* untuk Mengukur Tingkat Keparahatan Penyakit pada Daun Apel. 8(2), A71-A77.
- Sapto, H. (2008). Membangun Sistem Keprofesian Penyuluh Pertanian. *Jurnal Ilmu-Ilmu Pertanian*, 4(1), 38-46.
- Syahid, D., Jumadi, & Nursantika, D. (2016). Sistem Klasifikasi Jenis Tanaman Hias Daun *Philodendron* Menggunakan Metode K-Nearest Neighbor (KNN) Berdasarkan Nilai Hue, Saturation, Value (HSV). *JOIN*, 1(1), 20-23.
- Wu, Z., Jiang, S., Zhou, X., Wang, Y., Zuo, Y., Wu, Z., Liang, L., & Liu, Q. (2020). Application of image retrieval based on convolutional neural networks and Hu invariant moment algorithm in computer telecommunications. *Computer Communications*, 150, 729-738. <https://doi.org/10.1016/j.comcom.2019.11.053>
- Y. Garis K, I. Santoso, dan R. I. (2011). *Klasifikasi Citra dengan Matriks Ko-Okurensi Aras Keabuan (Gray Level Co-Occurrence Matrix-GLCM) pada Limakelas Biji-Bijian*.
- Zhang, H., Jiang, M., & Kou, Q. (2020). Color Image Retrieval Algorithm Fusing Color and Principal Curvatures Information. *IEEE Access*, 8, 184945-184954. <https://doi.org/10.1109/access.2020.3030056>