

Ekstraksi Fitur Daun Tembakau Berbasis Discrete Cosine Transform (DCT)

Rosida Vivin Nahari^{1*}, Arda Surya Editya^{2*}, Riza Alfita^{3*}

* Fakultas Teknik, Universitas Trunojoyo Madura

rosida.nahari@trunojoyo.ac.id¹, ardasurya@gmail.com², riza.alfita@trunojoyo.ac.id³

Article Info

Article history:

Received 2019-11-26

Revised 2020-01-01

Accepted 2020-01-21

Keyword:

Tobacco leaf, DCT, feature extraction, YCbCr, Classification

ABSTRACT

The success of the tobacco leaf classification process is very dependent on the extraction of tobacco leaf features. Several stages of digital image processing can improve the ability to identify the best quality tobacco automatically through extracting leaf texture features. This study aims to apply the leaf texture feature extraction system using the Discrete Cosine Transform method. Classification results measure the accuracy of the success of the system in extracting the best texture features. The classification of tobacco leaves requires extensive knowledge and complex terminology, even professional graders require significant time in this field for mastery of the subject. This is because tobacco leaves are usually considered to have characteristics that are useful for identification of tobacco quality where the extraction of appropriate features through leaf images can be considered a research problem that plays an important role for classification. The proposed research aims to find a suitable extraction model for obtaining color features through YCbCr color space conversion and tobacco leaf texture obtained from the transformation of the Discrete Cosine Transform frequency space. On classification stage in this research uses the maximum likelihood method. The trial results show an accuracy of success in the classification of tobacco leaves by 90% through the extraction of 12 features.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. PENDAHULUAN

Pada tanaman tembakau bagian terpenting yang dipanen adalah daunnya. Oleh sebab itu, klasifikasi tanaman tembakau berdasarkan identifikasi daun menjadi tren yang populer. Daun tembakau membawa informasi penting yang dapat digunakan untuk mengidentifikasi dan mengklasifikasikan asal atau jenis tanaman tembakau.

Pengolahan citra digital untuk pengklasifikasian daun tembakau dapat mempermudah tugas *grader*. Berdasarkan sejarahnya pengolahan citra digital telah lama digunakan oleh pakar kesehatan untuk mendiagnosis penyakit dan metode ini telah terbukti handal selama bertahun-tahun. Dengan menggunakan metode yang sama seperti pakar kesehatan, peneliti mencoba mensimulasikan prinsip yang sama untuk mengenali grade daun tembakau yang menggunakan citra digital daun tembakau berkualitas tinggi dan formula matematika yang rumit berbasis komputer untuk menentukan grade tanaman tembakau. Beberapa

percobaan telah menghasilkan banyak kisah sukses di laboratorium, tetapi beberapa pendekatan telah gagal total ketika diuji di dunia nyata. Hal ini dikarenakan para peneliti mungkin telah mengabaikan fakta bahwa pengambilan sampel dunia nyata tidak memiliki ketepatan dan kesempurnaan model seperti di laboratorium. Apa yang ingin disampaikan oleh penelitian di sini adalah pendekatan kasus yang ideal dalam klasifikasi.

Sebagian besar pendekatan yang diusulkan memulai proses dengan pra-pemrosesan gambar. Sebagian besar teknik yang telah diimplementasikan selama pra-pemrosesan diantaranya adalah pengkonversian citra RGB ke skala abu-abu, konversi citra grayscale menjadi citra biner dan proses peningkatan kualitas citra. Tujuan dari penerapan proses ini adalah untuk meminimalkan *noise* pada gambar yang bisa mengganggu proses ekstraksi dan klasifikasi. Hampir semua pendekatan diimplementasikan dengan cara yang sama selama pra-pemrosesan citra [1]. Oleh karena itu, penelitian ini tidak akan menguraikan

banyak tentang Preprocessing citra. Penelitian ini akan fokus pada ekstraksi fitur daun yang diusulkan dan pendekatan klasifikasi oleh peneliti sebelumnya.

Beberapa pendekatan yang diusulkan sebelumnya berfokus pada metode pengenalan dan pengklasifikasian. Proses pengenalan biasanya terjadi selama preprocessing dan selanjutnya diikuti oleh proses ekstraksi. Setelah itu, proses klasifikasi akan melihat ke sebuah database untuk membandingkan fitur daun. Metode pengenalan bisa diterapkan melalui: (i) berbasis kontur dan (ii) pendekatan berbasis wilayah. Namun, pendekatan berbasis kontur memiliki kesulitan dalam menemukan titik kelengkungan yang tepat jika dibandingkan dengan pendekatan berbasis wilayah [2], [3]. Metode lainnya adalah momen invarian, yang diusulkan oleh [4] di mana ia bekerja dengan 10 jenis daun. Meskipun banyak pendekatan telah diusulkan dan diuji dan hampir seluruh fitur daun berhasil diekstraksi, masih ada keterbatasan pada pendekatan – pendekatan tersebut. Beberapa pendekatan ditemukan menjadi tidak akurat, terutama karena citra input mengandung *noise*. Selain itu, pemahaman berbeda tentang pembobotan untuk fitur yang diekstraksi juga mempengaruhi pengenalan karena perbedaan definisi fitur atau dataset yang berbeda selama pengujian [5], [6].

Dalam beberapa kasus, fitur warna dan bentuk telah digunakan untuk mendapatkan vektor fitur pada tahap klasifikasi [4], [7], [8]. Ada berbagai algoritma klasifikasi yang digunakan untuk mengklasifikasikan tanaman berdasarkan vektor fitur. Beberapa makalah membahas tentang *Support Vector Machine* (SVM)[9] dan *Algoritma Artificial Neural Network* [10]. Selain itu pada tahap preprocessing, distribusi Gaussian secara efektif mampu mengoptimalkan pengelompokan fitur daun [11]. Ekstraksi fitur tekstur pada beberapa penelitian sebelumnya berhasil dilakukan dalam ranah domain frekwensi [12]. Ekstraksi fitur teksturnya juga bisa dilakukan melalui pembagian subband citra hasil dari hasil transformasi *Discrete wavelet* [13]. Penelitian [12] mengklasifikasikan daun berdasarkan tekstur kasar dari daun dan sekaligus mereduksi data menggunakan *fourier descriptor*.

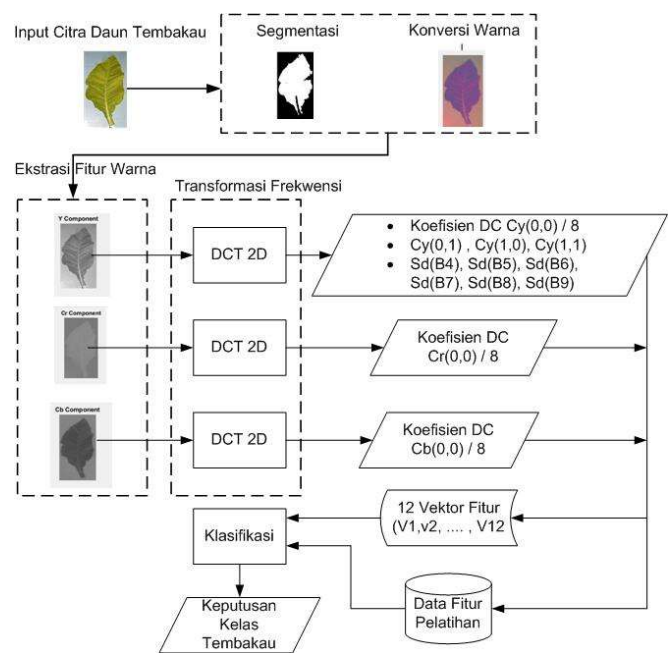
Tujuan dari penelitian ini adalah untuk mengeksplorasi koefisien *Discrete Cosine Transform* (DCT) dan fitur bentuk / morfologi untuk mendapatkan bentuk daun tembakau sehingga dapat digunakan untuk mengidentifikasi daun dengan akurasi tinggi yang didukung informasi lain seperti warna daun. Koefisien DCT dipilih berdasarkan kinerja yang baik dalam beberapa aplikasi diantaranya: klasifikasi protein [14], objek berbasis bentuk kendaraan [15] dan klasifikasi sidik jari [16]. Namun, beberapa fitur bentuk / morfologi lain dieksplorasi untuk mendukung koefisien DCT sehingga meningkatkan akurasi dalam mengidentifikasi daun tembakau.

II. METODE PENELITIAN

Pada penelitian ini, sistem citra digital memproses gambar warna 24-bit daun tembakau.. Fitur daun tembakau dapat

diekstraksi sesuai pemrosesan citra digital, yang meliputi fitur warna dan tekstur. Fitur-fitur ini disimpan dalam database di komputer. Fitur warna termasuk varian saluran merah, hijau dan biru dari citra digital daun tembakau. Fitur tekstur termasuk energi dan kontras tekstur citra digital daun tembakau yang diperoleh dengan mengimplementasikan transformasi DCT pada kanal Y, Cb dan Cr.

Tahapan dalam alur penelitian meliputi 3 langkah utama, yaitu : konversi warna, ekstraksi fitur dan terakhir tahap klasifikasi. Daun tembakau yang telah dicapture terlebih dahulu disegmentasi untuk mempermudah proses konversi. Selanjutnya dilakukan pemisahan kanal Y, Cb dan Cr untuk pengambilan fitur komponen DCT. Kedua belas fitur menjadi inputan untuk klasifikasi ke dalam 4 kelas seperti tampak pada gambar 1.



Gambar 1. Diagram Alur Penelitian

A. Ekstraksi Fitur Warna

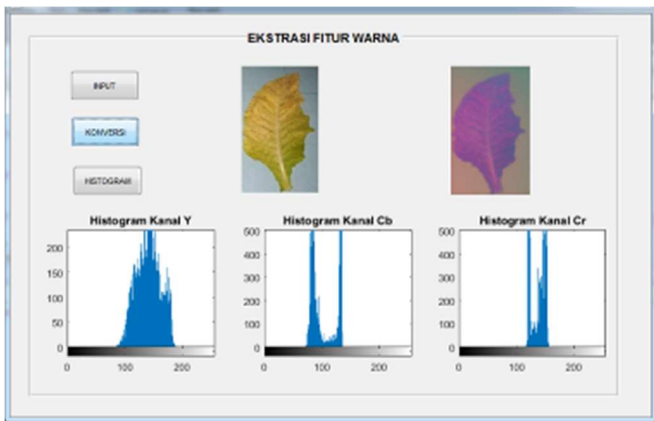
Warna adalah fitur penting dari daun tembakau karena hubungannya yang erat dengan kualitas yang dirasakan. Fitur ini merupakan parameter yang banyak digunakan dalam evaluasi kematangan, kesegaran, kondisi gizi dan faktor pertumbuhan. Ada sejumlah sistem warna, seperti RGB, HSV, YCbCr, dll [17], yang dapat digunakan pada computer vision untuk mengevaluasi warna. Setiap sistem warna memiliki kelebihan sendiri dan digunakan untuk itu memenuhi persyaratan yang berbeda. Warna dalam kebanyakan sistem warna berbeda dari penglihatan warna manusia dan tidak dapat dipahami secara intuitif. Agar mesin dapat mensimulasikan penglihatan warna manusia, maka sistem warna yang dipilih harus memiliki parameter spasial warna dan struktur yang mirip dengan sensasi warna manusia.

Perbandingan setiap ruang warna pada beberapa penelitian menunjukkan bahwa pada citra dengan Ruang warna YCbCr banyak digunakan dalam ekstraksi fitur [9]. Dalam format ini, informasi pencahayaan diwakili oleh komponen tunggal, Y, dan informasi warna disimpan sebagai dua komponen kroma warna, Cb dan Cr. Komponen Cb adalah perbedaan antara komponen biru dan nilai referensi, dan komponen Cr adalah perbedaan antara komponen merah dan nilai referensi. Y adalah komponen luma yang didefinisikan memiliki kisaran nominal 8-bit yaitu antara 16 - 235; Cb dan Cr adalah masing-masing komponen kroma tingkat perbandingan warna biru dan kroma tingkat perbandingan warna merah, yang didefinisikan memiliki kisaran nominal 16 - 240.

Sebelum penerapan transformasi RGB ke YCbCr, nilai rata-rata dari tiga gambar bidang R, G dan B dihapus. Penghapusan ini dikarenakan sebagian besar sinyal energi yang baru berubah menjadi citra YCbCr lebih banyak terkandung dalam bidang Y. Sehingga kita dapat mencapai efisien jumlah koefisien yang memungkinkan diekstraksi dalam Cb dan Cr. Hal ini menunjukkan bahwa penggunaan ruang warna YcbCr memberikan kinerja yang lebih baik daripada ruang RGB asli.

Transformasi yang digunakan untuk mengkonversi dari ruang warna RGB ke YCbCr ditunjukkan dalam persamaan (1):

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.144 \\ -0.16875 & -0.33126 & 0.5 \\ 0.5 & -0.41869 & -0.08131 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$



Gambar 1. Tahapan ekstraksi fitur pada ruang warna YcbCr

B. Discrete Cosine Transform (DCT)

Discrete Cosine Transform (DCT) telah banyak digunakan dalam literatur untuk efisiensi pemilihan fitur tekstur. DCT menggunakan cosinus dari berbagai frekuensi spasial sebagai fungsi dasar dan umumnya dikenal karena penerapannya dalam standar kompresi JPEG. Untuk tekstur citra, sebagian besar energi sinyal terletak pada komponen frekuensi rendah, yang mana muncul di sudut kiri atas DCT. DCT mengubah informasi spasial ke dalam domain frekuensi, fitur tekstur dapat didefinisikan sebagai spektrum energi di berbagai lokasi blok lokal. Karena koefisien DC

mewakili (hampir) nilai rata-rata skala abu-abu dari setiap blok $N \times N$, maka informasi tekstur yang didapat kurang spesifik detail tekstur yang diekstraksi.

Penelitian ini menggunakan *Discrete Cosine Transform* yang merupakan metode berbasis penampilan lokal untuk klasifikasi daun tembakau. DCT memiliki sifat-sifat seperti de-korelasi, Keterpisahan, pemadatan energi, simetris, Orthogonal yang membuatnya cocok untuk identifikasi daun tembakau. Komponen frekuensi rendah tersimpan di sudut kiri atas. Komponen-komponen ini dapat diekstraksi untuk pengklasifikasian daun tembakau. Definisi DCT paling umum dari urutan 1-D panjang N adalah

$$C(u) = \alpha(u) \sum_{x=0}^{N-1} f(x) \cos \left[\frac{(2x+1)\pi u}{2N} \right] \quad (2)$$

dimana $u=0,1,2,\dots,N-1$

Sedangkan *Inverse Discrete Cosine Transform* (IDCT) didefinisikan sebagai :

$$f(x) = \sum_{u=0}^{N-1} \alpha(u) C(u) \cos \left[\frac{(2x+1)\pi u}{2N} \right] \quad (3)$$

dimana $x=0,1,2,\dots,N-1$

Kedua persamaan di atas harus memenuhi persyaratan sebagai berikut :

Jika $u = 0$, maka $\alpha(u) = \sqrt{\frac{1}{N}}$ dan

Jika $u \neq 0$, maka $\alpha(u) = \sqrt{\frac{2}{N}}$

Sehingga dapat disederhanakan untuk $u=0$ menjadi :

$$C(u=0) = \sqrt{\frac{1}{N}} \sum_{x=0}^{N-1} f(x) \quad (4)$$

Koefisien transformasi yang pertama disebut sebagai koefisien DC. Sedangkan lainnya disebut sebagai koefisien AC. Transformasi 2D DCT dapat dilakukan secara langsung melalui model 1D sebagai berikut :

$$C(u, v) = \alpha(u)\alpha(v) \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x, y) \cos \left[\frac{(2x+1)\pi u}{2N} \right] \cos \left[\frac{(2y+1)\pi v}{2N} \right] \quad (5)$$

Sedangkan inverse 2D DCT (2D IDCT) adalah sebagai berikut:

$$f(x, y) = \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} \alpha(u)\alpha(v) C(u, v) \cos \left[\frac{(2x+1)\pi u}{2N} \right] \cos \left[\frac{(2y+1)\pi v}{2N} \right] \quad (6)$$

C. Ekstraksi Fitur Tekstur Berbasis DCT

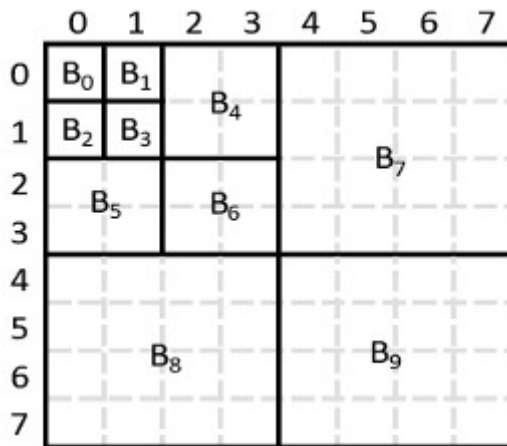
Penelitian ini mengekstraksi fitur tekstur dengan mengaplikasikan DCT dimana derajat keabuan setiap piksel dalam citra dinormalisasikan menjadi 128 melalui operasi subtraction. Selanjutnya, citra dibagi menjadi beberapa blok sehingga didapatkan koefisien DCT dari setiap blok.

[18] mengusulkan metode yang dimulai dengan mengubah citra dari RGB menjadi ruang warna YCbCr. Kemudian citra dibagi menjadi 8×8 blok, dan operasi DCT

seperti ditunjukkan oleh rumus (5) dilakukan pada masing-masing blok untuk saluran Y, Cb dan Cr. Hasil dari masing-masing blok koefisien DCT 8x8 dibagi menjadi sub-blok B_0, B_1, \dots, B_9 seperti tampak pada gambar 3.

Vektor fitur lokal yang berjumlah 12 fitur ($v_0, v_1, v_2, \dots, v_{11}$) didapatkan dari blok 8x8. Komponen vektor terdiri dari :

Vektor v_0, v_1, v_2 didapatkan dari nilai komponen DC pada sub-citra channel Y, Cb dan Cr dalam ruang DCT ($C_y(0,0)/8, C_{Cb}(0,0)/8, C_{Cr}(0,0)/8$) Vektor v_3, v_4, v_5 didapatkan dari nilai komponen B1, B2 dan B3 dari sub-citra Y ($C_y(0,1), C_y(1,0), C_y(1,1)$) Vektor v_6, v_7, v_8, v_9 didapatkan dari Standard deviasi blok B4, B5,.....,B9 sub-citra Y.



Gambar 3. Koefisien DCT dengan blok 8x8 yang dibagi menjadi 10 bagian [18].

D. Klasifikasi Daun Tembakau

Setelah mengekstraksi fitur warna dan tekstur untuk setiap daun tembakau, tahap selanjutnya menggunakan metode jaringan syaraf tiruan backpropagation untuk mengklasifikasi grade daun tembakau. Secara khusus, klasifikasi daun tembakau terdiri dari dua langkah yaitu tahap pelatihan dan tahap pengujian. Pada langkah pelatihan, setiap kelas dari citra daun tembakau diambil, dan fitur diekstraksi untuk setiap daun tembakau di kelas tersebut..

III. HASIL DAN PEMBAHASAN

Tahap pengujian kinerja sistem klasifikasi dilakukan dengan cara membandingkan hasil identifikasi pada citra daun tembakau yang diperoleh dari pengamatan langsung dengan hasil identifikasi melalui aplikasi yang telah dibuat. Sortasi daun tembakau dibagi menjadi empat yaitu:

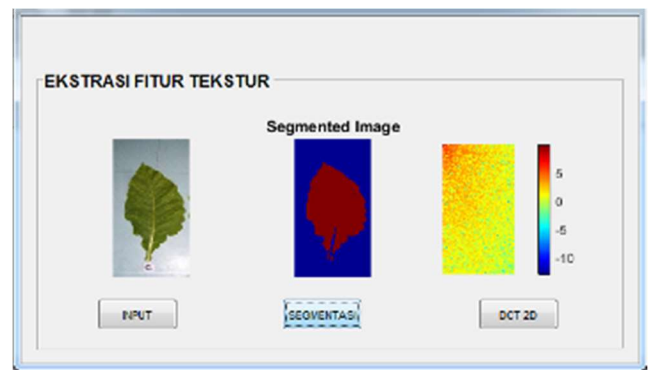
1. daun kurang masak ditandai dengan warna lebih hijau dan lebih segar
2. daun tepat masak berwarna hijau kekuningan
3. daun terlalu masak berwarna kekuningan dengan tepian berwarna coklat.
4. daun cacat.

Untuk setiap kelompok, 40 daun tembakau dipilih untuk mengevaluasi metode yang diusulkan. keempat jenis ini dipilih secara khusus dengan bantuan seorang grader. Penelitian ini mengaplikasikan DCT dan ruang warna YcrCb sebagai ekstraksi fitur. Citra yang digunakan adalah citra daun tembakau sebanyak 40 citra yang dicapture menggunakan kamera dengan pembagian sebanyak 30 citra untuk pelatihan dan 10 citra untuk uji coba. 12 Fitur didapatkan melalui ekstraksi berbasis DCT dengan blok 8x8.

Untuk mengevaluasi kinerja sistem ekstraksi, dilakukan analisa akurasi klasifikasi rata-rata dan standar deviasi. Tabel 1 menunjukkan hasil klasifikasi yang terbaik disertai dengan dalam bentuk confusion matrix. Kelas 1 merupakan data daun yang kurang masak. Kelas 2 adalah daun tepat masak. Kelas 3 untuk data daun yang terlalu masak. Sedangkan kelas 4 untuk daun tembakau yang cacat (umumnya terdapat lobang kecil di daun dengan tepian lubang berwarna coklat).

TABEL I
HASIL AKURASI KLASIFIKASI DAUN TEMBAKAU DENGAN 12 FITUR

| Prediksi / kelas real | 1 | 2 | 3 | 4 | \sum data | Prosentase error | Akurasi |
|-----------------------|---|---|----|----|-------------|------------------|---------|
| 1 | 9 | 1 | 0 | 0 | 10 | 1 % | 90 % |
| 2 | 1 | 7 | 1 | 1 | 10 | 3 % | |
| 3 | 0 | 0 | 10 | 0 | 10 | 0 % | |
| 4 | 0 | 0 | 0 | 10 | 10 | 0 % | |



Gambar 4. Hasil ekstraksi fitur tekstur pada domain DCT

Sistem dengan benar mengklasifikasikan 9 dari 10 sampel milik kelas 1 (akurasi 90 %), Sementara itu, tingkat kesalahan klasifikasi kelas 3 dan kelas 4 adalah 0,00, dan ini adalah keadaan yang sempurna. Di sisi lain, tingkat kesalahan klasifikasi tertinggi adalah di kelas 2 dengan kesalahan 3%. Menganalisa dari tingkat error pada tabel 3 di atas adalah terletak pada tiga sampel kelas 2 yang salah diklasifikasikan ke kelas 1, 3 dan 4.

IV. KESIMPULAN

Penelitian ini berhasil melakukan kombinasi fitur sebagai pondasi awal klasifikasi daun tembakau. Pemanfaatan ruang

warna YcbCr untuk mendapatkan fitur lokal terlebih pada kanal Y sebagai nilai informasi tingkat pencahayaan citra dan pemilihan fitur warna dari citra terhadap kroma merah dan biru. Pemilihan fitur selanjutnya adalah dengan cara melakukan transformasi frekwensi DCT yang diusulkan berbasis blok 8x8. Uji coba penelitian ini berhasil mengklasifikasikan daun tembakau dengan akurasi sebesar 90 %. Penelitian ini diharapkan dapat dikembangkan dengan mengintegrasikan ekstraksi fitur bentuk dari daun tembakau.

UCAPAN TERIMA KASIH

Kami sampaikan terima kasih kepada Universitas Trunojoyo atas bantuan dan kontribusinya pada penelitian ini.

DAFTAR PUSTAKA

- [1] R. Meena Prakash, G. P. Saraswathy, G. Ramalakshmi, K. H. Mangaleswari, and T. Kaviya, "Detection of leaf diseases and classification using digital image processing," in *Proceedings of 2017 International Conference on Innovations in Information, Embedded and Communication Systems, ICIECS 2017*, 2018.
- [2] K. B. Lee and K. S. Hong, "An implementation of leaf recognition system using leaf vein and shape," *Int. J. Bio-Science Bio-Technology*, 2013.
- [3] A. Hasim, Y. Herdiyeni, and S. Douady, "Leaf Shape Recognition using Centroid Contour Distance," in *IOP Conference Series: Earth and Environmental Science*, 2016.
- [4] Z. Zulkifli, P. Saad, and I. A. Mohtar, "Plant leaf identification using moment invariants & general regression neural network," in *Proceedings of the 2011 11th International Conference on Hybrid Intelligent Systems, HIS 2011*, 2011.
- [5] R. V. Nahari *et al.*, "Cow Weight Estimation Using Local Adaptive Thresholding Method And Connected Component Labelling," 2018.
- [6] R. V. Nahari and R. Alfitra, "Identification of Chlorophyll-A Distribution Using Landsat 8 in Madura," *Adv. Sci. Lett.*, 2018.
- [7] A. Salman, A. Semwal, U. Bhatt, and V. M. Thakkar, "Leaf classification and identification using Canny Edge Detector and SVM classifier," in *Proceedings of the International Conference on Inventive Systems and Control, ICISC 2017*, 2017.
- [8] C. Zhao, S. S. F. Chan, W. K. Cham, and L. M. Chu, "Plant identification using leaf shapes - A pattern counting approach," *Pattern Recognit.*, 2015.
- [9] S. Agrawal, N. K. Verma, P. Tamrakar, and P. Sircar, "Content based color image classification using SVM," in *Proceedings - 2011 8th International Conference on Information Technology: New Generations, ITNG 2011*, 2010.
- [10] J. Amara, B. Bouaziz, and A. Algergawy, "A Deep Learning-based Approach for Banana Leaf Diseases Classification," in *BTW*, 2017.
- [11] S. Van Wittenberghe, J. Verrelst, J. P. Rivera, L. Alonso, J. Moreno, and R. Samson, "Gaussian processes retrieval of leaf parameters from a multi-species reflectance, absorbance and fluorescence dataset," *J. Photochem. Photobiol. B Biol.*, 2014.
- [12] R. Roslan and N. Jamil, "Texture feature extraction using 2-D Gabor Filters," in *ISCAIE 2012 - 2012 IEEE Symposium on Computer Applications and Industrial Electronics*, 2012.
- [13] A. Ponomarev, H. S. Nalamwar, I. Babakov, C. S. Parkhi, and G. Buddhawar, "Content-based image retrieval using color, texture and shape features," in *Key Engineering Materials*, 2016.
- [14] X. Y. Song, Z. H. Chen, X. Y. Sun, Z. H. You, L. P. Li, and Y. Zhao, "An ensemble classifier with random projection for predicting protein-protein interactions using sequence and evolutionary information," *Appl. Sci.*, 2018.
- [15] M. Shehata, R. Abo-Al-Ez, F. Zaghlool, and M. T. Abou-Kreisha, "Vehicles Detection Based on Background Modeling," *Int. J. Eng. Trends Technol.*, 2018.
- [16] P. Sony, K. Kiran, and R. Bharat, "Latent Finger Print Matching using FFT and DCT," *Int. J. Comput. Appl.*, 2017.
- [17] A. Jurio, M. Pagola, M. Galar, C. Lopez-Molina, and D. Paternain, "A comparison study of different color spaces in clustering based image segmentation," in *Communications in Computer and Information Science*, 2010.
- [18] H. Nezamabadi-pour, H. Nezamabadi-pour, S. Saryazdi, and S. Saryazdi, "Object-Based Image Indexing and Retrieval in DCT Domain using Clustering Techniques," *Proc. World Acad. Sci. Eng. Technol.*, 2005.