

Implementasi Convolutional Neural Network Untuk Ekstraksi Fitur Citra Daun Dalam Kasus Deteksi Penyakit Pada Tanaman Mangga Menggunakan Random Forest

Riksa Adenia^{*1}, Agus Eko Minarno², Yufis Azhar³

^{1,2,3}Universitas Muhammadiyah Malang

rizkaadenia@webmail.umm.ac.id^{*1}, aguseko@umm.ac.id², yufis@umm.ac.id³

Abstrak

Penyakit daun mangga tidak hanya mempengaruhi pentingnya dari proses perkembangan pada tumbuhan, namun juga berdampak terhadap segala aspek, mulai dari hasil perkembangbiakan pada tanaman mangga seperti kualitas dari buah, daun, dan batang, namun juga berpengaruh terhadap nilai kuantitas Tujuan utama dari penelitian ini adalah untuk mengembangkan metode serta fitur-fitur yang akan mendukung proses dari penelitian. Seperti klasifikasi, memprediksi serta mendeteksi adanya suatu penyakit pada objek yang diteliti. Oleh karena itu, pada penelitian ini mengusulkan metode CNN untuk klasifikasi daun mangga yang terinfeksi oleh penyakit dan dikombinasi dengan metode Random Forest masuk dalam kategori Supervised Learning yang merupakan sebuah pendekatan yang dimana sudah terdapat data yang dilatih, dan terdapat variabel yang sudah ditargetkan sehingga tujuan dari pendekatan tersebut adalah mengelompokan sebuah data ke data yang sudah ada. Sehingga algoritma Random Forest akan digunakan sebagai pengujian algoritma pembelajaran mesin konvensional di dalam penelitian. Penelitian ini mendapatkan hasil yang baik pada hasil pengujian dengan tingkat akurasi sebesar 100% dari kedua metode yang digunakan CNN dan Random forest.

Kata Kunci: Fitur Selection, Klasifikasi, Daun Mangga, CNN, Random Forest, Deep Learning, Supervised Learning, Berpenyakit, Sehat

Abstract

Mango leaf disease not only affects the importance of the developmental process in plants, but also has an impact on all aspects, ranging from breeding results in mango plants such as the quality of fruit, leaves, and stems, but also affects the quantity value. is to develop methods and features that will support the research process. Such as classification, predicting and detecting the presence of a disease in the object under study. Therefore, this study proposes the CNN method for the classification of mango leaves infected by the disease and combined with the Random Forest method into the Supervised Learning category which is an approach in which there is already trained data, and there are variables that have been targeted so that the purpose of The approach is to group a data into existing data. So the Random Forest algorithm will be used as a tester of conventional machine learning algorithms in research. The study obtained good results on test results with an accuracy rate of 100% of both methods used CNN and Random Forest.

Keywords: Feature Selection, Classification, Mango Leaf, CNN, Random Forest, Deep Learning, Supervised Learning, Diseased, Healthy

1. Pendahuluan

Tanaman mangga adalah tanaman tropis, tapi saat ini budidaya tanaman mangga sudah banyak dilakukan di beberapa negara selain kawasan Asia Tenggara. [1] Penyakit paling populer pada daun mangga biasanya aspek dari serangga, suhu, bakteri, jamur, dan infeksi virus, lalu nantinya penyakit ini dapat mempengaruhi perkembangan pada tanaman itu sendiri. Biasanya tanaman juga akan menunjukkan beberapa spot atau titik tertentu pada bagian daun, bunga buah, dan batang ketika sudah terinfeksi oleh penyakit tanaman. Ini adalah salah satu faktor penting dalam perkembangan tanaman mangga karena dapat mempengaruhi hasil kualitas dan kuantitas. Hama ini rentan sekali terjadi pada banyak tanaman terutama pada daun mangga. Hal ini menyebabkan pada saat proses terjadinya fotosintesis pada sebuah tanaman, menjadi terganggu

[2]. Untuk mengatasi masalah yang dihadapi, para ahli patologi dan petani melakukan pendeteksian penyakit melalui mata mereka serta membuat landasan keputusan pada pengalaman mereka. Hal ini tentu saja tidak akan efektif mengingat metode yang mereka lakukan masih sangat tradisional [3]. Beberapa metode informasi dilakukan pada peningkatan basis pengetahuan yang ditinjau dari banyak artikel tentang hama pada tanaman manga [4]. Dan dalam waktu dekade terakhir, teknologi AI dan ML sudah mencapai atensi yang lebih kuat untuk mengidentifikasi beberapa penyakit pada tumbuhan, dengan tersedianya sejumlah perangkat dan prosesor komputasi dengan kapasitas tinggi. Oleh sebab itu diperlukannya alat untuk mendeteksi penyakit daun mangga agar bisa mencapai hasil yang lebih akurat [5]. Penelitian sebelumnya yang dilakukan oleh S. Madiwalar beserta tim sebelumnya [6], dengan judul penelitian "Plant Disease Identification A Comparative Study". Model yang diusulkan untuk mengidentifikasi penyakit pada daun mangga dengan gambar dan warna dari daun mangga. Model yang diusulkan dalam mengidentifikasi penyakit daun mangga adalah SVM, selama fase pengujian, YCbCr akan mengkonversi gambar serta mengerjakan bagian vektor fitur tekstural dan fitur warna melalui gambar input yang diberikan ke pengklasifikasian selama sesi pengujian. Lalu untuk basis warna teknik dan filter Gabor dipakai pada tekstur dan warna ekstraksi yaitu GLCM. Dengan jumlah dataset 86 gambar hanya untuk klasifikasi jarak minimum dan pendukung Vector Machine dengan hasil akurasi sebesar 79,16% dan 83,34%. Penelitian sebelumnya dilakukan oleh Ratnawati beserta tim sebelumnya [7], yang berjudul "Penerapan Random Forest untuk mengukur Tingkat Keparahan Penyakit Pada Daun Apel". Model yang diusulkan menggunakan metode klasifikasi Random Forest. Dengan Jumlah data citra 467 citra yang mampu menghasilkan kemampuan klasifikasi Random Forest untuk mengukur seberapa banyak keparahan pada penyakit daun apel dengan nilai akurasi sebesar 75.31%. Dalam penelitian sebelumnya, yang dilakukan oleh Pham beserta tim sebelumnya [3], dengan judul "Early Disease Classification of Mango Leaves Using Feed-Forward Neural Network and Hybrid Metaheuristic Feature Selection". Model yang diusulkan untuk mendeteksi penyakit dini pada daun yaitu CNN dengan arsitektur AlexNet, VGG16, RestNet-50 kemudian dibandingkan dengan model ANN. Dataset yang digunakan 450 gambar dengan resolusi gambar 256x256 pixel. Akurasi tertinggi sebesar 89,41%. Kelebihan dari model ANN hasil akhirnya lebih tinggi dibandingkan dengan model CNN untuk arsitektur VGG16 sebesar 78,64%, lalu AlexNet sebesar 79,92% dan RestNet-50 sebesar 84,88%.

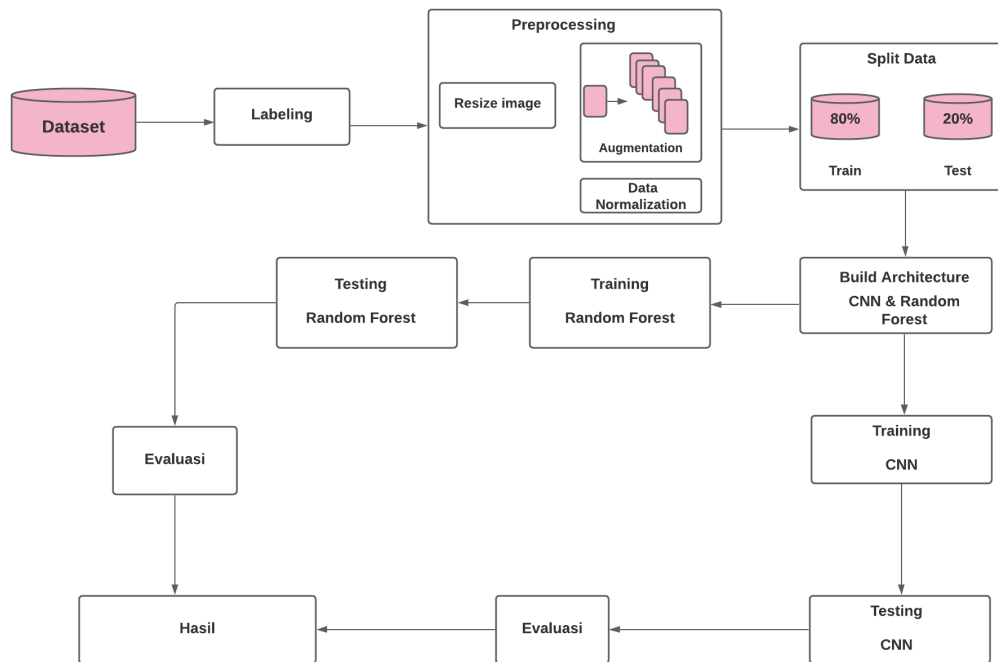
Dari penelitian sebelumnya, yang dilakukan oleh Ullagaddi beserta tim sebelumnya [8], dengan judul penelitian "Disease recognition in Mango crop using modified rotational kernel transform features". Model yang diusulkan untuk deteksi penyakit pada daun mangga adalah model set fitur arah MRKT. Dengan dataset yang digunakan 500 gambar yang dibagi menjadi 70% untuk set pelatihan dan sisanya 30% untuk set pengujian resolusi gambar 4320x3240 pixel. Dengan akurasi sebesar 98%. Penelitian sebelumnya, dilakukan oleh Singh beserta tim sebelumnya [5], dengan judul "Multilayer Convolutional Neural Network for the Classification of Mango Leaves Infected by Anthracnose Disease". Model yang diusulkan untuk mendiagnosis penyakit daun mangga dan mengklasifikasikan penyakitnya adalah CNN. Dengan jumlah dataset 2200 kemudian gambar dibagi kedalam pelatihan diambil dari masing-masing label kelas 0, kelas 1, kelas 2, dan kelas 3 menjadi 80% untuk set pelatihannya dan sisanya 20% untuk set pengujian dengan resolusi 128x128. Kemudian kumpulan data yang sudah berisi masing masing gambar dinormalisasi ke dalam MCNN untuk mengekstrak fitur. Model ini dilakukan untuk memprediksi label kelas dari setiap pelatihan. Dari hasil training dan testing yang telah dilakukan, label kelas 0 dengan training data sebesar 98,17% dan testing 97,12%, lalu kelas 1 untuk training data sebesar 98,59%, kemudian testing sebesar 96,78%, selanjutnya kelas 2 training data sebesar 98,34%, untuk testing sebesar 97,74%, terakhir kelas 3 dengan training data sebesar 98,20% dan testing sebesar 96,88%. Dapat kita simpulkan pada hasil testing akurasi tertinggi adalah pada kelas 2 sebesar 97,74. Kemudian pada fitur model CNN dengan akurasi tertinggi yaitu fitur MCNN sebesar 97,13.

Dalam penelitian ini, untuk mengidentifikasi penyakit pada daun mangga. Metode yang diusulkan adalah Klasifikasi Random Forest dengan menggunakan fitur CNN [3] yang mampu mengklasifikasikan antara penyakit dengan tidak berpenyakit. Dengan dipecahkan klasifikasi multi-kelas untuk jenis penyakit, kerangka kerja yang diusulkan bertujuan untuk mencapai hasil meningkat akurasi model ketika data model terbatas.

2. Metode Penelitian

Pada tahap ini membahas mengenai alur pengerjaan penelitian serta tahapan dalam pengerjaan yang sedang dijalankan Pada dasarnya metode yang digunakan adalah Convolutional

Network Network dan Random Forest. Penggambaran dari tahapan dalam pengerjaan penelitian yang dijalankan dapat di lihat Pada Gambar 1.



Gambar 1. Arsitektur CNN Gambar 1. Alur Penelitian Dalam Model Penerapan Random Forest Menggunakan CNN

2.1 Pengumpulan Dataset

Dataset yang digunakan merupakan data klasifikasi antara daun mangga berpenyakit dan daun mangga normal yang didapat dari author jurnal panduan utama, yang dibagikan melalui google drive. Dengan total jumlah data sebanyak 605 gambar yang dimana terbagi menjadi 2 kelas yaitu daun mangga berpenyakit dan daun mangga normal. Jumlah data yang digunakan pada kelas daun berpenyakit sebanyak 435 gambar dan daun normal sebanyak 170 gambar.

2.2 Labeling

Labeling Melakukan operasi manipulasi data seperti penggabungan, pembentukan kembali pada data yang diolah, kemudian pemilihan data, serta pembersihan, dan wriling data. Tujuannya untuk melihat jumlah data gambar pada masing – masing label. Selanjutnya proses yang dilakukan pada tahapan ini adalah melakukan augmentasi dari masing – masing label yang telah di bentuk, dan menambahkan masing – masing sebanyak 2000 untuk tiap - tiap kelasnya.

2.3 Preprocessing

1. Preprocessing yang dilakukan pada tahap ini adalah melakukan resize gambar, karena ukuran gambar awal sebelumnya sebesar 4000x6000 piksel kemudian dikonversi menjadi 200x200 piksel untuk tinggi serta lebar gambar, guna memenuhi kebutuhan pemrosesan pada gambar untuk menunjang proses penelitian lebih lanjut.
2. Melakukan sistem augmentasi. Augmentasi data bertujuan untuk menambah data yang masih kurang serta memaksimalkan dan menyeimbangkan proses pengerjaan pada program untuk mendapatkan hasil akurasi yang baik. Proses yang dilakukan pada augmentasi gambar adalah melakukan augmentasi secara bervariasi yaitu, melakukan model horizontal flip dan vertical flip, kemudian rotasi berlawanan arah jarum jam serta rotasi searah jarum jam, selanjutnya melakukan proses pergeseran melengkung pada gambar, lalu proses menambah efek terang pada gambar serta penambahan efek blur/buram pada gambar [1]. Setelah melakukan sistem augmentasi, data yang didapatkan bertambah meningkat menjadi 2465 untuk kelas daun berpenyakit dan 2170 untuk kelas daun normal dengan total keseluruhan 4635 data gambar.

3. Melakukan normalisasi data terhadap data yang telah dibagi dalam tiga bagian pada metode CNN yakni, data train sebesar 80%, data testing sebesar 10%, dan data validasi sebesar 10%. Kemudian untuk metode Random Forest yang telah dibagi menjadi dua bagian, untuk data train sebesar 80% dan data testing 20%.

Setelah melakukan proses sistem augmentasi, data yang didapatkan akan meningkatkan keragaman data untuk model pelatihan di dalam deep learning tanpa harus benar-benar mengumpulkan dataset baru.



Gambar 2. Proses Augmentasi Data yang Dilakukan 10 x 1 gambar

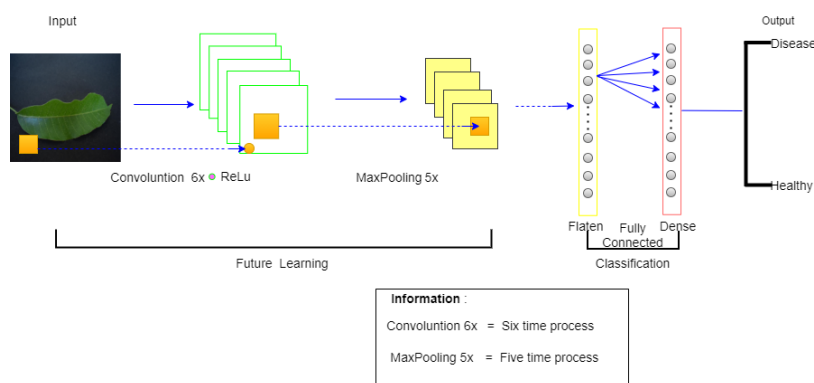
2.4 Splitting Data

Melakukan proses splitting data terhadap dataset yang sudah di augmentasi dari masing - masing kelas. Tiap metode pada netode CNN dibagi menjadi tiga bagian yaitu, 80% untuk data training, 10% data testing, dan 10% data validasi. Kemudian untuk metode Random Forest hanya melakukan pembagian data menjadi dua bagian yaitu, 80% data training dan 20% data testing, Karena dataset yang digunakan pada Random Forest data manual atau dataset asli yang diberikan oleh author jurnal panduan utama yang tidak melakukan proses augmentasi namun tetap melakukan proses resize pada gambar.

2.5 Membangun Arsitektur Dari Dua Metode Yang Digunakan

2.5.1 Metode Convolutional Neural Network

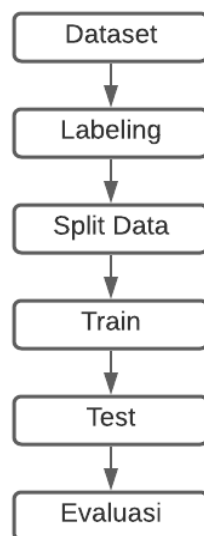
Rancangan model ialah struktur rancangan dari program yang akan dibentuk untuk sistem klasifikasi pada citra Daun Mangga. Model CNN menggunakan regulasi di fully connected layer. Skema dari arsitektur model CNN akan dilatih pada serangkaian dataset pada gambar daun yang telah di preprocessing. Model CNN melibatkan 214.402 parameter yang berisi 6 lapisan convolutional dengan bidang reseptif yang sangat kecil yaitu 3 x 3, lima lapisan max-pooling dengan ukuran 2 x 2 untuk melakukan pemberian jalan informasi yang nantinya diikuti oleh tiga lapisan fully connected bersama lapisan akhir sebagai lapisan softmax. Dalam mobilisasi/aktivasi 64 non linier (ReLU) gunakan pada semua bagian hidden layer[1] [17]. Proses convolution dan pooling, pada hasil perhitunga nilai, nilai – nilai perhitungan pada layer sebelumnya lanjutkan sampai ke fully connected untuk dilakukan pemrosesan prediksi pada data, guna menghasilkan output kelas. Skema ini dapat dilihat Pada Gambar 3.



Gambar 3. Arsitektur model CNN

2.5.2 Random Forest

Random Forest masuk dalam kategori Supervised Learning yang merupakan sebuah pendekatan yang dimana data telah dilatih lalu terdapat variabel yang sudah ditargetkan kemudian tujuan dari pendekatan tersebut adalah mengelompokkan sebuah data ke dalam data yang sudah ada. Sehingga algoritma Random Forest akan digunakan sebagai pengujian algoritma pembelajaran mesin konvensional di dalam penelitian [2]. Adapun struktur rancangan dari program yang sudah dibentuk dari mulai penginputan dataset yang akan digunakan pada proses dalam proses selanjutnya. Kemudian dilakukan pelabelan terhadap database yang telah diimplementasikan dengan membuat folder dataset gambar menjadi bentuk dataframe agar rapi dan terstruktur. Tahapan berikutnya melakukan pembagian data terhadap data train sebesar 80% dan data test sebesar 20%. Kemudian pada proses training melakukan prediksi data gambar yang dimana menunjukkan berapa banyak jumlah data dari dua kelas yang terdeteksi yaitu data kelas diseased dan data kelas health. Pada testing menampilkan hasil akurasi dari model Random Forest, dan terakhir pada evaluasi melihat hasil prediksi dari data berjenis daun apa yang akan terdeteksi, apakah sehat atau berpenyakit.



Gambar 4. Arsitektur Model Random Forest

2.6 Pengujian

Pada tahap ini adalah bagian akhir dalam tahap penelitian. Tahap yang digunakan dalam menentukan bagaimana kinerja keseluruhan metode yang dimasukkan dan digunakan untuk dilakukan perbandingan dari kedua metode akan menghasilkan nilai akurasi yang tertinggi, itulah metode yang terbaik yang digunakan pada penelitian ini.

3. Hasil dan pembahasan

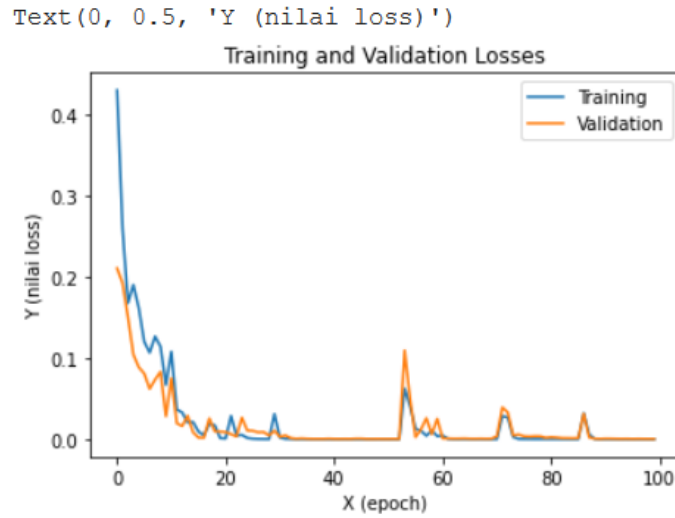
Pada tahap ini pengujian yang dilakukan secara bertahap dan terpisah antara metode CNN dengan Random Forest. Karena untuk Splitting Data CNN dibagi menjadi tiga yaitu training 80%, testing 10%, dan validation 10%. Kemudian untuk splitting Random Forest dibagi menjadi dua yaitu training 80% dan testing 20%. Dataset yang digunakan Random Forest bukan dari augmentasi tetapi data manual awalnya. Jadi hanya pada CNN data dilakukan system augmentasi dari dataset daun mangga yang didapat melalui author peneliti sebelumnya yang menjadi acuan dalam penelitian sekarang.

3.1 Metode CNN

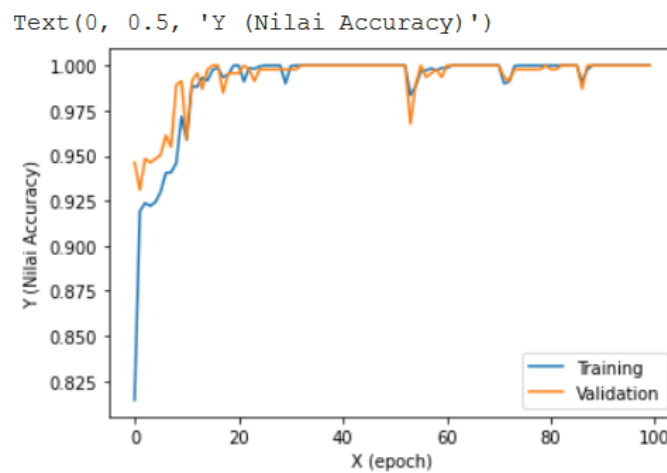
3.1.1 Training Model

Melatih model dan validasi dilakukan terlebih dahulu sebelum melakukan proses pengujian model. Hal ini bertujuan untuk melatih model dan dilakukannya validasi sekaligus untuk mengetahui mengevaluasi hasil akurasi training yang diperoleh sama baiknya dari akurasi validasi yang diperoleh. Pada Gambar 5, grafik dibawah menunjukkan Training dan validation loss. Jika diperhatikan grafik loss nya semakin turun dan dapat dikatakan hasil dari grafik nya sudah

bagus. Karena jika grafik loss nya semakin naik maka modelnya tidak bagus. Pada Gambar 6, untuk hasil akurasi training dan validation sudah sama-sama mencapai akurasi 100%.



Gambar 5. Hasil Akurasi Training Dan Validation Loss



Gambar 6. Hasil Pengujian dengan teknik preprocessing

3.1.2 Testing Model

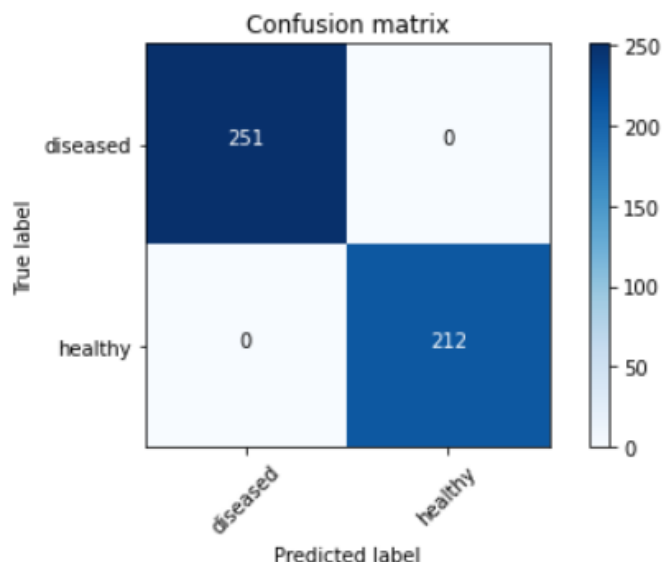
Pada Gambar 7, pengujian metode CNN menggunakan aktivasi hidden layer softmax dan dilakukan sebanyak 50 epoch. untuk dapat melihat konsistensi dari akurasi. dari hasil akurasi trainingnya sebesar 100% dan hasil akurasi testing yang didapatkan sebesar 100%.

	precision	recall	f1-score	support
diseased	1.000	1.000	1.000	251
healthy	1.000	1.000	1.000	212
accuracy			1.000	463
macro avg	1.000	1.000	1.000	463
weighted avg	1.000	1.000	1.000	463

Gambar 7. Hasil Testing Model

3.1.3 Plot Hasil Confusion Matrix

Pada hasil perhitungan confusion matrix berdasarkan data sesungguhnya dan data prediksi, bisa dilihat Pada Gambar 8, bahwa dalam gambar menunjukkan dari data kelas diseeded dan data kelas healthy. Pada data kelas deceased semua data terprediksi sebanyak 251 yang berpenyakit, dan sebanyak 212 yang sehat.



Gambar 8. Hasil Confusion Matrix Dan Data Prediksi

3.1.4 Hasil Prediksi Data Baru

Proses prediksi yang dilakukan menghasilkan output berupa kalimat "I am 100% percent confirmed that this is a Diseased leaves". dan dengan gambar daun mangga yang 100% tampak sehat jika dilihat Pada Gambar 9 dibawah ini.



Gambar 9. Hasil Prediksi Dari Data Baru

3.2 Metode Random Forest

3.2.1 Training Random Forest

Model Random Forest yang telah dibuat dengan menggunakan metode GridsearchCV untuk mencari parameter terbaik yang akan diberikan kepada model yang akan diuji, sekaligus melakukan cross validation sebanyak 10 fold. Untuk parameter model Random Forest yang digunakan, menggunakan parameter default dari modul sklearn. Namun untuk parameter "criterion", "max_depth", "n_estimators" akan dilakukan perbandingan terhadap nilai (value) dari masing – masing parameter tersebut. Contohnya pada parameter "max_depth" diberikan beberapa yang kemudian ditempatkan dalam bentuk list. Dari beberapa nilai tersebut akan dicari yang menghasilkan akurasi paling baik dari proses GridsearchCV.

3.2.2 Plot Hasil Confusion Matrix

Tabel 1 berikut hasil prediksi pada data yang sudah dilakukan prosesnya dari membuat parameter kemudian bisa mendapatkan hasil parameter terbaik untuk menunjang proses tahapan berikutnya. Pada Gambar 66, adalah hasil prediksi data gambar yang dimana menunjukkan dari dua kelas yaitu data kelas diseased dan data kelas healthy. Pada data kelas diseased semua data terprediksi sebanyak 53 dan healthy sebanyak 34. Dari hasilnya semua benar karena jumlahnya sama rata antara Expected results dengan Predicted results.

Tabel 1. Hasil Confusion Matrix Dan Data Prediksi

Leaf	Detected Disease	Detected Healthy
Disease	53	0
Healthy	0	34

3.2.3 Testing Random Forest

Pada Gambar 11 adalah hasil akurasi testing dari Random Forest dengan nilai sebesar 100%.

```

Classification report of model:
              precision    recall  f1-score   support

     0           1.00      1.00      1.00         53
     1           1.00      1.00      1.00         34

 accuracy              1.00              1.00              1.00         87
 macro avg              1.00              1.00              1.00         87
 weighted avg           1.00              1.00              1.00         87

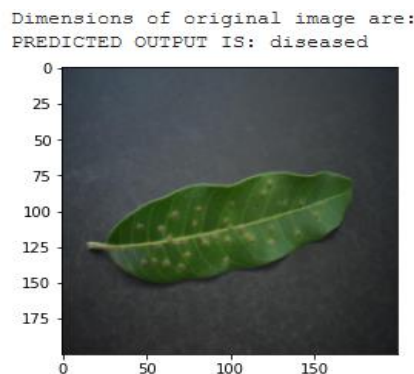
```

Accuracy score: 100.0

Gambar 10. Hasil Testing Model

3.2.4 Hasil Prediksi Data Baru

Proses prediksi yang dilakukan menghasilkan output diseased/berpenyakit. Mengapa demikian, karena gambar dataset yang dimasukkan secara manual dan secara acak kemudian diproses oleh system. Untuk hasil yang mendominasi yaitu daun berpenyakit/diseased.



Gambar 11. Hasil Prediksi Pada Data Baru

3.3 Analisa Dan Perbandingan

Pada penelitian ini, peneliti mengusulkan dua metode yaitu CNN dan Random Forest. Metode CNN untuk mengklasifikasikan antar dua jenis daun yaitu berpenyakit dan daun sehat. Sedangkan metode Random Forest untuk mendeteksi duan dari dataset normal menunjukkan hasil akurasi pada model training dan testing. Kemudian Gambar 9 dan Gambar 12 menunjukkan keakuratan data yang diprediksi dari kelas yang berbeda. Dan terakhir hasil output gambar dari prediksi apakah sesuai dari prediksi yang dilakukan pada Gambar 9 dan Gambar 12. Berikut adalah hasil akurasi dari 2 metode yang digunakan dalam eksperimen ini.

Tabel 2. Akurasi Eksperimen Terhadap 2 Metode

Scenario	Class/Label	Precision	Recall	F1-score	Akurasi	Support
CNN	Disease	100%	100%	100%	100%	251
	Healthy	100%	100%	100%		212
Random Forest	Disease	100%	100%	100%	100%	53
	Healthy	100%	100%	100%		34

Tabel 3. Perbandingan Nilai Akurasi Penelitian Terdahulu Dengan yang Sekarang

No.	Model	Dataset	Accuracy	Method
1.	Pham, Tan Nhat	450 picture	89.41%	ANN
	Tran, Ly Van	with resolution 256x256 pixel	84,88%	CNN
	Dao, Son Vu Truong [3]			
2.	Singh, Uday Pratap	2200 picture	97,74%	CNN
	Chouhan, Siddharth Singh [5]	with resolution 128x128 pixel	97,13%	MCNN
3.	Scenario 1	605 picture with resolution 200x200+(Augmentation)	100%	CNN
4.	Scenario 2	605 picture with resolution 200x200	100%	Random Forest

Dijelaskan hasil perbandingan antara penelitian terdahulu dengan penelitian yang dilakukan sekarang. Pada penelitian sebelumnya yang berjudul "Early Disease Classification of Mango Leaves Using Feed-Forward Neural Network and Hybrid Metaheuristic Feature Selection", yang dilakukan oleh Tan Nhat Pham, Ly Van Tran, dan Son Vu Truong Daon mendapatkan hasil akurasi yang paling baik dari metode CNN sebesar 84,88% dan dari metode ANN sebesar 89,41%. Pada penelitian sebelumnya peneliti menggunakan dataset sebanyak 450 gambar dari sampel yang diakumulasi dari setiap daerah yang ada di Provinsi An Giang, [14] tempat pembudidayaan tanaman mangga terbesar di Vietnam. Kemudian membuat 4 kelas, 3 kelas untuk nama penyakit dan 1 kelas untuk yang kelas sehat. Pada permasalahan yang sama penelitian yang dilakukan sekarang, peneliti merekomendasikan sistem yang berbeda pada preprocessing data serta model yang berbeda. Pada penelitian ini peneliti melakukan sistem augmentasi terhadap dataset gambar saat proses preprocessing data yang berguna untuk menambah data gambar dari beberapa proses seperti melipatgandakan data gambar. Teknik image preprocessing juga dilakukan dalam penelitian ini, dengan melakukan rescaling agar dapat menetapkan gambar training dan testing memiliki ukuran yang sama. Selain itu, mengubah resolusi gambar awal yang sangat besar menjadi resolusi lebih rendah dan dilakukan perhitungan ulang untuk memadatkan gambar asli dengan ukuran 200 × 200 piksel. Pertama gambar asli dilakukan pembagian dan dikonversi ke biner untuk menemukan kotak batas minim. Melakukan ke 200 piksel pada vertikal ukuran kotak batas yang dimana agar dapat memastikan daun atas dan bawah sama serupa ke atas dan bawah dari gambar yang diskalakan. Dan ukuran horizontal dari kotak batas difungsikan sebagai penggeseran gambar daun ke pusat yang tepat dari gambar berskala. Metode yang digunakan pada penelitian ini adalah CNN dan Random Forest, karena di penelitian sebelumnya metode CNN dan ANN yang digunakan mampu menghasilkan akurasi yang baik. Jadi penelitian yang sekarang mencoba untuk menggunakan metode CNN dan menggunakan Random Forest guna melakukan penelitian terbaru dari metode pada penelitian sebelumnya, apakah mampu untuk mendapatkan nilai akurasi yang lebih baik. Pengujian yang dilakukan menghasilkan akurasi paling baik pada metode Random Forest sebesar 100% pada pengujian.

4. Kesimpulan

Berdasarkan pada pengujian yang dilakukan, disimpulkan bahwa metode Convolution Neural Network dan Random Forest dalam penelitian ini mampu mendapatkan nilai akurasi yang baik pada penelitian yang dilakukan terhadap dataset daun mangga multi-kelas. Pada pengujian dilakukan preprocessing data dengan tujuan mengoptimalkan hasil akhir yaitu menghasilkan nilai akurasi yang baik. Pengujian yang dilakukan juga pada model CNN yang diusulkan dengan menambahkan *content loss* guna untuk mengetahui apakah nilai yang akan dihasilkan baik atau tidak dari penelitian sebelumnya. Jika nilai *loss* nya semakin turun maka bisa dikatakan fungsi *loss* semakin baik. Karena jika nilai *loss* nya semakin naik dan besar, sudah dipastikan hasil nilai akhir dan model yang digunakan tidak bagus. Dan keluaran dari model Random Forest yang dibuat menggunakan model GridsearchCV untuk mencari parameter terbaik yang akan diberikan kepada model yang akan diuji, sekaligus melakukan cross validation sebanyak 10 fold. Dari beberapa nilai tersebut akan dicari yang menghasilkan nilai akurasi paling baik dari proses model yang digunakan dalam parameter.

Referensi

- [1] L. C. Ngugi, M. Abelwahab, and M. Abo-Zahhad, "Recent advances in image processing techniques for automated leaf pest and disease recognition – A review," *Inf. Process. Agric.*, no. xxxx, 2020, doi: 10.1016/j.inpa.2020.04.004.
- [2] E. E. Saeed *et al.*, "Detection and management of mango dieback disease in the United Arab Emirates," *Int. J. Mol. Sci.*, vol. 18, no. 10, 2017, doi: 10.3390/ijms18102086.
- [3] T. N. Pham, L. Van Tran, and S. V. T. Dao, "Early Disease Classification of Mango Leaves Using Feed-Forward Neural Network and Hybrid Metaheuristic Feature Selection," *IEEE Access*, vol. 8, pp. 189960–189973, 2020, doi: 10.1109/access.2020.3031914.
- [4] C. Trongtorkid and P. Pramokchon, "Expert system for diagnosis mango diseases using leaf symptoms analysis," *3rd Int. Conf. Digit. Arts, Media Technol. ICDAMT 2018*, pp. 59–64, 2018, doi: 10.1109/ICDAMT.2018.8376496.
- [5] U. P. Singh, S. S. Chouhan, S. Jain, and S. Jain, "Multilayer Convolution Neural Network for the Classification of Mango Leaves Infected by Anthracnose Disease," *IEEE Access*, vol. 7, pp. 43721–43729, 2019, doi: 10.1109/ACCESS.2019.2907383.
- [6] S. C. Madiwalar and M. V. Wyawahare, "Plant disease identification: A comparative study," *2017 Int. Conf. Data Manag. Anal. Innov. ICDMAI 2017*, pp. 13–18, 2017, doi: 10.1109/ICDMAI.2017.8073478.
- [7] L. Ratnawati and D. R. Sulistyningrum, "Penerapan Random Forest untuk Mengukur Tingkat Keparahan Penyakit pada Daun Apel," vol. 8, no. 2, pp. A71–A77, 2019.
- [8] S. B. Ullagaddi and S. Viswanadha Raju, "Automatic Robust Segmentation Scheme for Pathological Problems in Mango Crop," *Int. J. Mod. Educ. Comput. Sci.*, vol. 9, no. 1, pp. 43–51, 2017, doi: 10.5815/ijmecs.2017.01.05.
- [9] B. Tymchenko, P. Marchenko, and D. Spodarets, "Deep learning approach to diabetic retinopathy detection," *ICPRAM 2020 - Proc. 9th Int. Conf. Pattern Recognit. Appl. Methods*, pp. 501–509, 2020, doi: 10.5220/0008970805010509.
- [10] T. Rui, J. Zou, Y. Zhou, J. Fei, and C. Yang, "Convolutional neural network simplification based on feature maps selection," *Proc. Int. Conf. Parallel Distrib. Syst. - ICPADS*, vol. 0, pp. 1207–1210, 2016, doi: 10.1109/ICPADS.2016.0161.
- [11] W. Setiawan, "Perbandingan Arsitektur Convolutional Neural Network Untuk Klasifikasi Fundus," *J. Simantec*, vol. 7, no. 2, pp. 48–53, 2020, doi: 10.21107/simantec.v7i2.6551.
- [12] R. Qian, Y. Yue, F. Coenen, and B. Zhang, "Visual attribute classification using feature selection and convolutional neural network," *Int. Conf. Signal Process. Proceedings, ICSP*, vol. 0, no. i, pp. 649–653, 2016, doi: 10.1109/ICSP.2016.7877912.
- [13] K. Trang, L. Tonthat, N. Gia Minh Thao, and N. Tran Ta Thi, "Mango Diseases Identification by a Deep Residual Network with Contrast Enhancement and Transfer Learning," *2019 IEEE Conf. Sustain. Util. Dev. Eng. Technol. CSUDET 2019*, pp. 138–142, 2019, doi: 10.1109/CSUDET47057.2019.9214620.
- [14] J. Sethupathy and S. Veni, "OpenCV based disease identification of mango leaves," *Int. J. Eng. Technol.*, vol. 8, no. 5, pp. 1990–1998, 2016, doi: 10.21817/ijet/2016/v8i5/160805417.