



PREDIKSI PENYAKIT MANGGA MENGGUNAKAN MODEL KLASIFIKASI MULTI-LABEL BERDASARKAN DATA CUACA

Hafidz Sanjaya*, Fajar Maula Hidayat

Program Studi Ilmu Komputer, Universitas Yayasan Pendidikan Imam Bonjol Majalengka, Indonesia

Abstrak: Penyakit pada tanaman mangga merupakan salah satu faktor utama yang dapat menurunkan produktivitas buah tropis, sehingga diperlukan pendekatan prediksi dini yang efektif berbasis data. Penelitian ini bertujuan mengembangkan model klasifikasi multi-label untuk memprediksi risiko beberapa penyakit mangga secara simultan berdasarkan data cuaca historis. Dataset yang digunakan terdiri dari data harian suhu, kelembapan relatif, dan curah hujan di Kabupaten Majalengka, Jawa Barat selama periode 2011–2023. Proses pelabelan dilakukan menggunakan pendekatan berbasis aturan untuk menghasilkan indikator biner tiga penyakit utama yaitu antraknosa, embun tepung, dan bercak daun. Model Multi-Output Random Forest dilatih menggunakan 80% data dan diuji pada 20% data dengan evaluasi menggunakan akurasi, presisi, recall, F1-score, dan hamming loss. Hasil penelitian menunjukkan bahwa model mampu mencapai akurasi sebesar 0.8811, presisi 0.8920, recall 0.9227, F1-score 0.9062, serta hamming loss yang rendah sebesar 0.0396. Model menunjukkan performa sangat baik dalam mendeteksi embun tepung dan bercak daun, namun masih menghadapi tantangan moderat pada prediksi antraknosa akibat kemiripan pola kondisi cuaca. Penelitian ini menyimpulkan bahwa integrasi pelabelan berbasis pengetahuan domain dengan pembelajaran mesin pada data cuaca efektif untuk mendukung prediksi multi-penyakit secara simultan. Temuan ini berimplikasi pada peningkatan sistem peringatan dini dan pengambilan keputusan berbasis data dalam pengelolaan penyakit tanaman mangga.

Kata kunci: data cuaca, klasifikasi multi-label, penyakit mangga, prediksi penyakit tanaman, random forest.

I. PENDAHULUAN

Mango (*Mangifera indica* L.) merupakan salah satu komoditas buah tropis utama yang dibudidayakan secara luas di Asia, dengan Indonesia sebagai salah satu produsen penting (Liang dkk., 2024). Sebagai komoditas hortikultura strategis, mangga memiliki peran signifikan dalam ketahanan pangan, peningkatan pendapatan petani, serta potensi ekspor nasional (Kiloes dkk., 2024). Namun demikian, produktivitas mangga masih

menghadapi berbagai tantangan, terutama serangan penyakit daun seperti antraknosa (*Colletotrichum gloeosporioides*), embun tepung (*Oidium mangiferae*), dan bercak daun (Umar dkk., 2022). Penyakit-penyakit tersebut dapat menyebabkan kerugian ekonomi yang signifikan, mulai dari penurunan hasil panen hingga kegagalan panen total, khususnya pada wilayah dengan kelembapan tinggi dan pola curah hujan yang tidak menentu (Nega, 2025).

Berbagai penelitian menunjukkan bahwa dinamika serangan penyakit tanaman mangga sangat dipengaruhi oleh kondisi lingkungan, terutama suhu, kelembapan relatif, dan curah hujan (Dianda dkk., 2023). Antraknosa,

*) hafidzsanjaya@lecturer.univypib.ac.id

Diterima: 18 April 2026

Direvisi: 18 Mei 2026

Disetujui: 23 Mei 2026

DOI: 10.23969/infomatek.v28i1.45396

misalnya, berkembang optimal pada kelembapan relatif di atas 80% dan suhu antara 25–30°C, terutama selama musim hujan (Kumari dkk., 2025; Raju dkk., 2024). Embun tepung cenderung muncul pada kondisi kering namun lembap, sedangkan bercak daun sering dipicu oleh kombinasi suhu hangat dan curah hujan sedang (Raju dkk., 2024). Hubungan kuat antara variabel cuaca dan kemunculan penyakit ini menunjukkan bahwa data meteorologi memiliki potensi besar untuk dimanfaatkan dalam sistem prediksi penyakit tanaman berbasis data.

Perkembangan teknologi kecerdasan buatan dalam bidang pertanian telah membuka peluang baru dalam membangun sistem prediksi penyakit yang lebih akurat dan skalabel. Berbagai pendekatan *machine learning*, seperti *Random Forest* dan *deep learning*, telah berhasil digunakan untuk mendeteksi penyakit tanaman dengan memanfaatkan data citra, sensor, maupun data lingkungan (Iqbal dkk., 2024; Kusri dkk., 2022; Pansy & Murali, 2023). Namun, sebagian besar penelitian yang ada masih berfokus pada pendekatan klasifikasi *single-label*, yaitu memprediksi satu jenis penyakit pada satu waktu, padahal dalam kondisi lapangan, beberapa penyakit dapat muncul secara bersamaan akibat kesamaan kondisi lingkungan.

Disamping itu, terdapat pendekatan klasifikasi multi-label yang memungkinkan satu *instance* memiliki lebih dari satu label secara simultan (Han dkk., 2023). Metode ini telah banyak diterapkan dalam berbagai bidang seperti diagnosis medis dan klasifikasi teks (Chai dkk., 2023; Rajput & Grover, 2022), tetapi implementasinya dalam prediksi penyakit tanaman, khususnya pada komoditas tropis seperti mangga, masih sangat terbatas.

Beberapa penelitian telah menerapkan pendekatan multi-label pada tanaman lain seperti apel (Zhou dkk., 2024), mentimun (Yang dkk., 2024), dan tomat (Yadav & Tewari, 2025). Namun, hingga saat ini, kajian yang secara khusus mengembangkan model klasifikasi multi-label untuk penyakit mangga berbasis data cuaca masih belum banyak dilakukan.

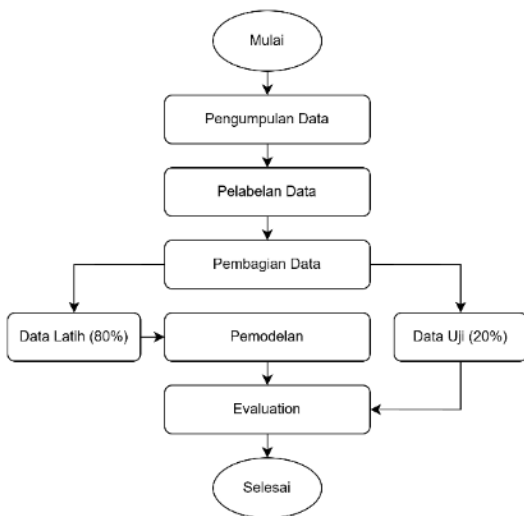
Di sisi lain, Indonesia memiliki keunggulan dalam ketersediaan data cuaca historis yang cukup lengkap melalui Badan Meteorologi, Klimatologi, dan Geofisika (BMKG), yang dapat dimanfaatkan untuk pengembangan sistem prediksi berbasis data. Studi sebelumnya menunjukkan bahwa data cuaca dapat digunakan secara efektif untuk memprediksi penyakit tanaman. (Matengu dkk., 2024) melaporkan bahwa model berbasis cuaca mampu memprediksi risiko *fusarium head blight* pada gandum dengan akurasi lebih dari 70%. (Z. Liu dkk., 2022) menunjukkan bahwa parameter lingkungan seperti suhu, kelembapan, dan curah hujan dapat digunakan untuk memprediksi penyakit daun pada tanaman teh. Sementara itu, (Dai dkk., 2023) mengombinasikan data cuaca dengan citra dalam model *deep learning* dan memperoleh performa yang sangat tinggi dalam identifikasi penyakit daun. Meskipun demikian, penelitian-penelitian tersebut masih belum mengeksplorasi pendekatan multi-label berbasis data cuaca secara khusus pada tanaman mangga, sehingga masih menyisakan gap penelitian yang signifikan.

Berdasarkan analisis tersebut, terdapat tiga kesenjangan utama yang menjadi dasar penelitian ini, yaitu terbatasnya penerapan model klasifikasi multi-label pada komoditas hortikultura tropis, khususnya mangga, serta belum optimalnya pemanfaatan data cuaca historis sebagai sumber utama dalam prediksi

beberapa penyakit secara simultan. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan model klasifikasi multi-label berbasis data cuaca yang mampu memprediksi beberapa penyakit pada tanaman mangga, yaitu antraknosa, embun tepung, dan bercak daun secara bersamaan. Model yang dikembangkan diharapkan dapat menjadi dasar bagi sistem peringatan dini (*early warning system*) yang ringan, adaptif, dan relevan dengan kondisi agroklimat, sehingga dapat mendukung pengambilan keputusan yang lebih tepat dalam manajemen penyakit tanaman mangga di wilayah tropis.

II. METODOLOGI

Penelitian ini menggunakan pendekatan *machine learning* berbasis klasifikasi multi-label untuk memprediksi penyakit mangga berdasarkan data cuaca historis. Alur penelitian terdiri dari lima tahapan utama, yaitu pengumpulan data, pelabelan data, pembagian data, pemodelan, dan evaluasi, sebagaimana ditunjukkan pada Gambar 1.



Gambar 1. Alur Penelitian

Gambar 1 menunjukkan alur sistematis penelitian dalam membangun model klasifikasi

multi-label. Proses dimulai dari pengumpulan data cuaca, dilanjutkan dengan pelabelan berbasis aturan, pembagian data, pemodelan, hingga evaluasi performa model. Setiap tahapan dirancang untuk memastikan model mampu melakukan prediksi secara akurat dan dapat digeneralisasi.

2.1 Pengumpulan Data

Dataset yang digunakan dalam penelitian ini diperoleh dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) Indonesia dengan fokus pada data stasiun cuaca dari Kabupaten Majalengka, Jawa Barat. Meliputi rentang waktu yang komprehensif dari 1 Januari 2011 hingga 31 Desember 2023. Dataset ini terdiri dari tiga variabel data cuaca seperti suhu ($^{\circ}\text{C}$), kelembaban relatif (%), dan curah hujan (mm). Ketiga variabel ini dipilih berdasarkan peran pentingnya dalam memicu terjadinya penyakit pada tanaman mangga, seperti antraknosa, embun tepung, dan bercak daun. Contoh data dalam dataset ditunjukkan pada Gambar 2.

| time | temperature | relative_humidity | precipitation |
|------------|-------------|-------------------|---------------|
| 2011-01-01 | 25.367654 | 85.186615 | 1.135348 |
| 2011-01-02 | 25.447899 | 86.173603 | 0.953230 |
| 2011-01-03 | 25.746970 | 89.058265 | 1.808085 |
| 2011-01-04 | 25.846479 | 85.285847 | 0.974780 |
| 2011-01-05 | 24.982169 | 90.513475 | 4.499832 |
| ... | ... | ... | ... |
| 2023-12-27 | 26.567116 | 79.941143 | 16.638808 |
| 2023-12-28 | 27.114086 | 82.850835 | 5.567098 |
| 2023-12-29 | 27.531671 | 84.496434 | 6.310406 |
| 2023-12-30 | 27.492492 | 85.934284 | 6.495102 |
| 2023-12-31 | 27.153203 | 84.011095 | 8.906113 |

Gambar 2. Contoh data pada dataset

Gambar 2 menampilkan contoh data pada dataset yang digunakan pada penelitian ini dengan empat struktur kolom yaitu *time* (waktu), *temperature* (suhu), *relative humidity* (kelembaban relatif), dan *precipitation* (curah hujan). Kolom waktu menunjukkan tanggal

spesifik dari setiap pengamatan. Kolom suhu memberikan suhu udara rata-rata harian, kolom kelembaban relatif menunjukkan persentase kelembaban, dan kolom curah hujan menunjukkan total curah hujan harian. Sementara itu, pembulatan dua digit di belakang koma dilakukan pada data cuaca sehingga sesuai dengan standar pelaporan data iklim (World Meteorological Organization (WMO), 2024). Misalnya, pada tanggal 1 Januari 2011, suhu yang tercatat adalah 25,37°C, kelembaban relatif 85,19%, dan curah hujan mencapai 1,14 mm. Dataset berlanjut dengan catatan harian serupa hingga 31 Desember 2023, di mana nilai-nilai tersebut mencerminkan tren iklim menjelang akhir periode pengamatan.

2.2. Pelabelan Data

Pada tahapan ini, multi-label biner untuk tiga kategori penyakit (antraknosa, embun tepung, dan bercak daun) diberikan pada setiap baris data di dataset. Pemberian label dilakukan berbasis aturan yang berasal dari pengetahuan agronomi tentang kondisi cuaca optimal yang memicu setiap penyakit (Umar dkk., 2022). Secara spesifik, antraknosa diberi label ketika kelembaban relatif melebihi 80% dan curah hujan lebih dari 5 mm (milimeter). Embun tepung diberi label ketika kelembaban melebihi 75% dan curah hujan kurang dari 1 mm. Bercak daun diidentifikasi ketika suhu berkisar antara 25°C sampai 30°C dan curah hujan antara 3 dan 10 mm. Aturan-aturan ini diimplementasikan dan diterapkan secara sistematis di semua baris data pada dataset sehingga menghasilkan dataset multi-label dimana setiap entri data dapat menunjukkan satu atau lebih risiko penyakit.

2.3. Pembagian Data

Setelah proses pelabelan, dataset dibagi menjadi dua bagian, yaitu data pelatihan (*training set*) dan data pengujian (*testing set*)

dengan rasio 80:20. Proporsi ini dipilih untuk memastikan model memperoleh data yang memadai dalam proses pembelajaran, sekaligus menyediakan subset data yang cukup untuk evaluasi. Pembagian ini memungkinkan model dilatih menggunakan sebagian besar data yang tersedia, sementara kinerjanya diuji pada data yang tidak pernah dilihat sebelumnya guna menilai kemampuan generalisasi model secara objektif.

2.4. Pemodelan

Model yang digunakan dalam penelitian ini adalah *multi-output random forest classifier*, yang memungkinkan prediksi beberapa label penyakit mangga secara simultan dalam satu kerangka pemodelan. Algoritma Random Forest dipilih karena memiliki sejumlah keunggulan, antara lain ketahanan yang baik terhadap *overfitting*, kemampuan dalam memberikan interpretasi melalui analisis *feature importance*, serta kesesuaian yang tinggi untuk mengolah data tabular terstruktur seperti data meteorologi (Elbeltagi dkk., 2023). Model dilatih menggunakan data latih, kemudian dievaluasi menggunakan data uji untuk mengukur kinerja prediktifnya.

2.5. Evaluasi

Evaluasi kinerja model dilakukan dengan menggabungkan metrik klasifikasi single label dan metrik khusus multi-label. Untuk setiap label penyakit, performa model diukur menggunakan *precision*, *recall*, dan *F1-score* guna menggambarkan tingkat ketepatan, sensitivitas, serta keseimbangan antara keduanya. Dalam konteks klasifikasi multi-label, *hamming loss* digunakan untuk mengukur proporsi rata-rata kesalahan dalam penetapan label pada seluruh prediksi. Nilai *hamming loss* yang lebih rendah menunjukkan kemampuan model yang lebih baik dalam memprediksi beberapa label secara simultan.

Selain itu, analisis *feature importance* diekstraksi dari model Random Forest untuk mengidentifikasi variabel meteorologi yang paling berkontribusi terhadap hasil prediksi. Analisis ini tidak hanya meningkatkan interpretabilitas model, tetapi juga memberikan wawasan praktis terkait faktor lingkungan yang paling berpengaruh terhadap kemunculan penyakit mangga.

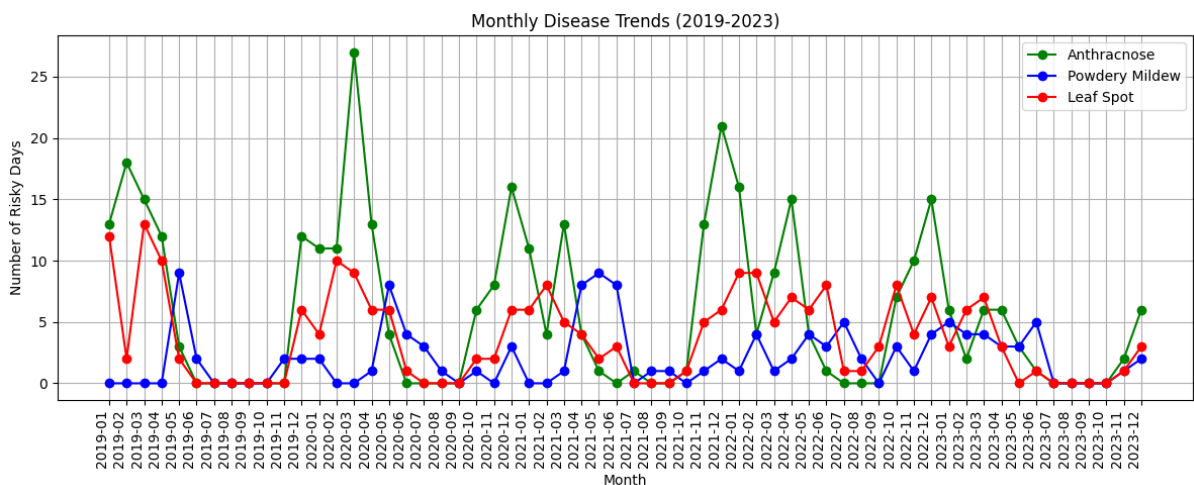
III. HASIL DAN PEMBAHASAN

Sebelum proses pelatihan model dilakukan, tahap pelabelan data berbasis aturan diterapkan terlebih dahulu untuk menetapkan indikator penyakit secara biner (0 atau 1) pada setiap baris data. Proses ini menghasilkan dataset dengan karakteristik multi-label, di mana setiap entri dapat merepresentasikan tidak adanya penyakit atau terdapat lebih dari satu penyakit sesuai dengan kondisi cuaca yang terjadi. Ringkasan hasil pelabelan data ditampilkan pada Gambar 3.

| time | temperature | relative humidity | precipitation | anthracnose | powdery mildew | leaf spot |
|------------|-------------|-------------------|---------------|-------------|----------------|-----------|
| 2011-01-01 | 25.367654 | 85.186615 | 1.135348 | 0 | 0 | 0 |
| 2011-01-02 | 25.447899 | 86.173603 | 0.953230 | 0 | 1 | 0 |
| 2011-01-03 | 25.746970 | 88.056265 | 1.808085 | 0 | 0 | 0 |
| 2011-01-04 | 25.846479 | 85.285847 | 0.974780 | 0 | 1 | 0 |
| 2011-01-05 | 24.982169 | 90.513475 | 4.499832 | 0 | 0 | 1 |
| ... | ... | ... | ... | ... | ... | ... |
| 2023-12-27 | 26.567116 | 79.941143 | 16.638808 | 0 | 0 | 0 |
| 2023-12-28 | 27.114086 | 82.850935 | 5.567098 | 0 | 0 | 0 |
| 2023-12-29 | 27.531671 | 84.496434 | 6.310406 | 1 | 0 | 0 |
| 2023-12-30 | 27.492492 | 85.934284 | 6.405102 | 1 | 0 | 1 |
| 2023-12-31 | 27.153203 | 84.011095 | 8.906113 | 1 | 0 | 0 |

Gambar 3. Ringkasan hasil pelabelan data

Untuk memberikan gambaran yang lebih komprehensif dari hasil pelabelan data, distribusi temporal penyakit mangga divisualisasikan dalam bentuk distribusi bulanan sebagaimana ditunjukkan pada Gambar 4. Visualisasi pada Gambar 4 menampilkan pola tren musiman serta frekuensi relatif masing-masing penyakit selama periode pengamatan setiap bulannya.



Gambar 4. Pola tren bulanan penyakit mangga pada dataset berlabel

Gambar 4 menyajikan tren bulanan tiga penyakit mangga selama periode 2019 hingga 2023. Di antara ketiganya, Antraknosa menunjukkan pola paling dominan dan fluktuatif, dengan puncak tertinggi terjadi pada Maret 2020 yang melebihi 25 hari berisiko

dalam satu bulan, mengindikasikan kondisi iklim yang sangat mendukung perkembangan penyakit. Embun Tepung muncul relatif lebih stabil dengan intensitas lebih rendah, namun menunjukkan kecenderungan peningkatan pada periode transisi menuju musim hujan,

khususnya di akhir tahun. Sementara itu, Bercak Dau memiliki pola yang lebih sporadis, dengan beberapa lonjakan signifikan yang umumnya terjadi pada awal dan akhir tahun, yang mengindikasikan adanya pengaruh faktor musiman. Meskipun terdapat tumpang tindih pola musiman antar penyakit, perbedaan intensitas dan waktu puncak menunjukkan karakteristik distribusi label yang berbeda. Analisis ini memberikan pemahaman penting terhadap struktur dataset berlabel, termasuk distribusi label, potensi ketidakseimbangan data, serta pola temporal yang mendasari kemunculan penyakit. Berdasarkan karakteristik tersebut, model random forest multi-output dipilih karena kemampuannya dalam menangani data multi-label secara simultan, mengakomodasi hubungan non-linear antar fitur cuaca dan kemunculan penyakit, serta relatif *robust* terhadap *noise* dan ketidakseimbangan data.

Setelah proses pelabelan data untuk tiga penyakit mangga dilakukan, model random forest multi-output dilatih dan dievaluasi untuk mengukur kinerjanya dalam memprediksi kemunculan penyakit berdasarkan data cuaca harian. Dalam konfigurasi model, jumlah estimator (*n_estimators*) ditetapkan sebanyak 100 untuk memungkinkan pembentukan ensemble pohon keputusan yang cukup beragam sehingga dapat meningkatkan kemampuan generalisasi. Selain itu, parameter *random_state* diatur ke nilai 42 guna memastikan reproduibilitas hasil pada setiap proses pelatihan dan evaluasi model.

Hasil evaluasi model menunjukkan bahwa model klasifikasi multi-label yang dikembangkan memiliki kinerja yang sangat baik dalam memprediksi risiko penyakit mangga. Sebagaimana ditunjukkan pada Tabel 1.

Berdasarkan tabel 1, hasil evaluasi model mencapai *accuracy* sebesar 0,8811, dengan

nilai *precision* 0,8920, *recall* 0,9227, dan *F1-score* 0,9062 yang menunjukkan performa kuat dan seimbang di seluruh metrik pengujian. Nilai *accuracy* sebesar 0,8811 mengindikasikan bahwa secara keseluruhan model mampu melakukan prediksi kombinasi label dengan benar, baik dalam mengonfirmasi kemunculan penyakit (*true positive*) maupun bukan penyakit (*true negative*).

Tabel 1. Hasil Evaluasi Model

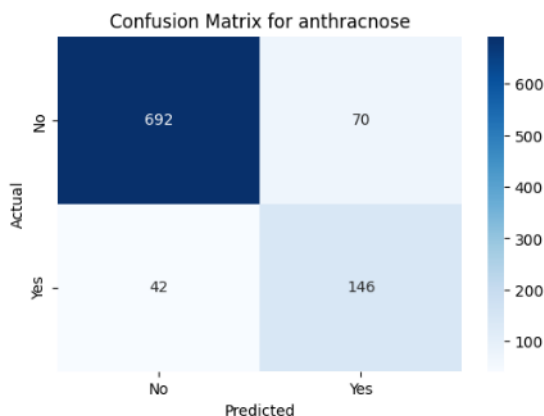
| Metrik | Nilai |
|---------------------|--------|
| <i>accuracy</i> | 0.8811 |
| <i>precision</i> | 0.8920 |
| <i>recall</i> | 0.9227 |
| <i>f1-score</i> | 0.9062 |
| <i>hamming lose</i> | 0.0396 |

Kinerja deteksi model juga diperkuat oleh nilai *recall* yang relatif tinggi, yaitu sebesar 0,9227. Nilai *recall* mengukur efektivitas kemampuan model dalam mengidentifikasi seluruh instans positif yang aktual melalui rasio *true positive* terhadap total kejadian positif yang sebenarnya (moola & Yakubreddy, 2024). Tingginya nilai ini membuktikan bahwa model mampu mengidentifikasi sebagian besar label penyakit dengan baik, sekaligus meminimalkan risiko luputnya deteksi atau munculnya kasus negatif palsu (*false negative*). Kemampuan identifikasi yang tinggi tersebut juga diimbangi oleh nilai *precision* yang kuat sebesar 0,8920. Nilai *precision* mencerminkan akurasi dari prediksi positif model dengan membandingkan instans yang benar-benar positif terhadap total seluruh instans yang diprediksi positif (moola & Yakubreddy, 2024). Dengan demikian, nilai ini menjamin bahwa model memiliki tingkat salah deteksi atau positif palsu (*false positive*) yang rendah. Keseimbangan yang optimal antara sensitivitas deteksi (*recall*) dan ketepatan prediksi (*precision*) ini direfleksikan secara akurat oleh nilai *f1-score* yang menyentuh

angka 0,9062. Nilai *f1-score* merepresentasikan rata-rata harmonik yang menggabungkan nilai *precision* dan *recall* dengan bobot yang sama (moola & Yakubreddy, 2024). Tingginya nilai *f1-score* ini menegaskan performa keseluruhan model yang solid dan tidak timpang sebelah dalam memprediksi keterkaitan antar-label kondisi tanaman.

Selain itu, nilai *hamming loss* yang rendah, yaitu sebesar 0,0396, menunjukkan bahwa proporsi kesalahan prediksi antar label sangat kecil, sehingga model tergolong *robust* dalam menangani klasifikasi multi-label secara simultan. Nilai *hamming loss* mengukur fraksi dari kesalahan prediksi, baik kasus luput deteksi (*false negative*) maupun salah deteksi (*false positive*) yang dinormalisasi terhadap total jumlah label dan instans data (J. Liu dkk., 2023). Secara keseluruhan, hasil ini mengindikasikan bahwa pendekatan random forest multi-output efektif dalam menangkap pola kompleks antara variabel cuaca dan kemunculan penyakit mangga, serta dapat digunakan sebagai sistem prediksi berbasis data.

Sementara itu, *confusion matrix* pada Gambar 5.

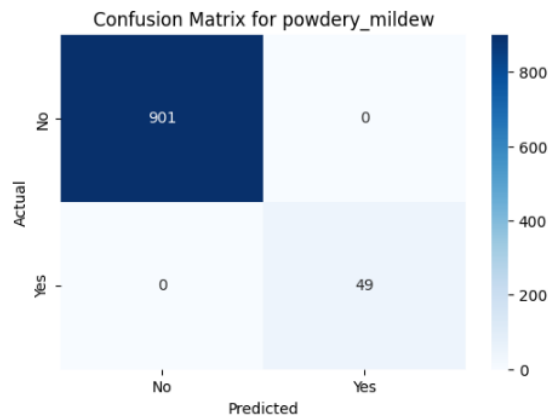


Gambar 5. *Confusion matrix* penyakit antraknosa

Gambar 5 memperlihatkan bahwa model mampu memprediksi sebagian besar kasus antraknosa dengan benar yang ditunjukkan dengan 146 *true positive* dan 692 *true negative*.

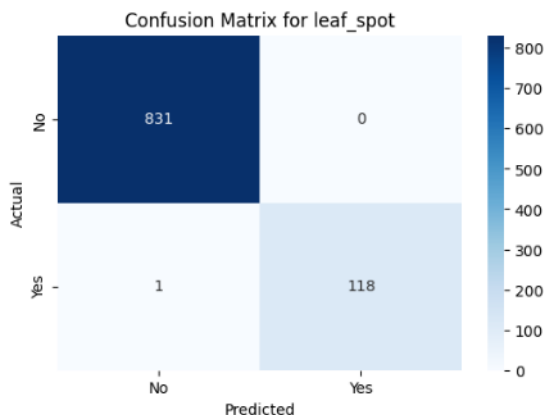
Meskipun masih terdapat beberapa *false positive*, kesalahan klasifikasi ini kemungkinan disebabkan oleh adanya data tumpang tindih yang tidak selalu berujung pada kejadian penyakit secara aktual. Meskipun demikian, model tetap mampu menangkap pola umum dengan baik, sehingga mendukung penggunaannya sebagai sistem peringatan dini.

Pada Gambar 6, tidak terdapat kesalahan klasifikasi untuk penyakit embun tepung yang ditunjukkan dengan 49 *true positive* dan 901 *true negative*. Hal ini menunjukkan bahwa embun tepung merupakan kelas yang sangat terpisah (*separable*) dalam ruang fitur dataset.



Gambar 6. *Confusion matrix* penyakit embun tepung

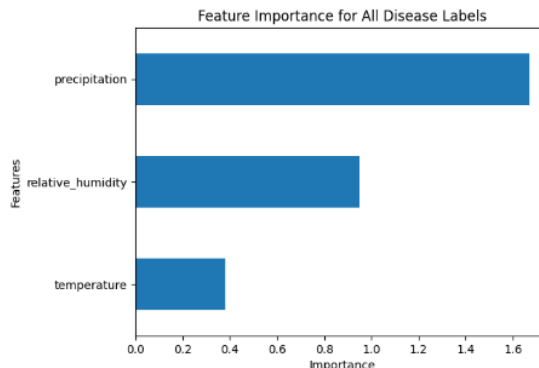
Demikian pula pada Gambar 7, model menunjukkan kinerja yang sangat baik dalam memprediksi penyakit bercak daun. Hal itu ditunjukkan hanya ada satu kasus *false negative* dalam melakukan klasifikasi.



Gambar 7. Confusion matrix penyakit bercak daun

Selain evaluasi kinerja model, analisis *feature importance* juga dilakukan untuk memahami kontribusi relatif masing-masing parameter cuaca terhadap prediksi model. Seperti ditunjukkan pada Gambar 8, curah hujan merupakan fitur paling berpengaruh untuk semua label penyakit, diikuti oleh kelembapan relatif dan suhu. Hasil ini menunjukkan bahwa curah hujan memiliki peran dominan dalam membentuk profil risiko penyakit mangga pada penelitian ini. Model sangat bergantung pada pola curah hujan, hal itu karena perubahan curah hujan secara langsung memengaruhi durasi kelembapan daun, yang merupakan faktor penting dalam perkembangan jamur (Umar dkk., 2022). Kelembapan relatif menempati urutan kedua, sejalan dengan pengetahuan agronomi bahwa tingkat kelembapan yang tinggi mendukung perkecambahan spora jamur dan penyebaran penyakit (Umar dkk., 2022). Meskipun suhu memiliki kontribusi paling rendah di antara ketiga fitur, variabel ini tetap relevan, khususnya dalam prediksi bercak daun yang memiliki ambang suhu tertentu. Secara keseluruhan, peringkat *feature importance* ini menegaskan adanya interaksi kompleks antara kondisi lingkungan dan kemunculan

penyakit, sekaligus memvalidasi pemilihan ketiga variabel tersebut dalam model prediktif.



Gambar 8. Feature importance untuk semua label penyakit mangga

IV. KESIMPULAN

Penelitian ini menunjukkan bahwa pendekatan klasifikasi multi-label berbasis data cuaca efektif dalam memprediksi beberapa penyakit mangga secara simultan. Model Random Forest Multi-Output yang dikembangkan mampu mencapai kinerja tinggi dengan *accuracy* 0,8811, *precision* 0,8920, *recall* 0,9227, dan *F1-score* 0,9062, serta *hamming loss* yang rendah sebesar 0,0396 yang menunjukkan kemampuan model dalam menangani prediksi multi-label secara robust. Model juga menunjukkan performa sangat baik pada embun tepung dan bercak daun, meskipun antraknosa masih memiliki tantangan akibat tumpang tindih kondisi cuaca. Penelitian ini menunjukkan integrasi pelabelan berbasis aturan dan pemodelan *machine learning* yang mampu menangkap pola kompleks antara variabel cuaca dan penyakit. Namun, keterbatasan utama terletak pada penggunaan dataset yang masih terbatas pada satu wilayah geografis, sehingga berpotensi membatasi generalisasi model. Oleh karena itu, penelitian selanjutnya disarankan untuk memperluas cakupan data lintas wilayah, menambahkan variabel lingkungan lainnya, serta membandingkan

dengan metode lain guna meningkatkan kinerja dan generalisasi model.

DAFTAR PUSTAKA

- Chai, Y., Liu, H., Xu, J., Samtani, S., Jiang, Y., & Liu, H. (2023). A Multi-Label Classification with an Adversarial-Based Denoising Autoencoder for Medical Image Annotation. *ACM Transactions on Management Information Systems*, 14(2), 1–21. <https://doi.org/10.1145/3561653>
- Dai, G., Fan, J., Tian, Z., & Wang, C. (2023). PPLC-Net: Neural network-based plant disease identification model supported by weather data augmentation and multi-level attention mechanism. *Journal of King Saud University - Computer and Information Sciences*, 35(5), 101555. <https://doi.org/10.1016/j.jksuci.2023.101555>
- Dianda, O. Z., Wonni, I., Ouédraogo, L., Sankara, P., Tollenaere, C., Del Ponte, E. M., & Fernandez, D. (2023). Lasiodiplodia species associated with mango (*Mangifera indica* L.) decline in Burkina Faso and influence of climatic factors on the disease distribution. *Physiological and Molecular Plant Pathology*, 126, 102041. <https://doi.org/10.1016/j.pmpp.2023.102041>
- Elbeltagi, A., Pande, C. B., Kumar, M., Tolche, A. D., Singh, S. K., Kumar, A., & Vishwakarma, D. K. (2023). Prediction of meteorological drought and standardized precipitation index based on the random forest (RF), random tree (RT), and Gaussian process regression (GPR) models. *Environmental Science and Pollution Research*, 30(15), 43183–43202. <https://doi.org/10.1007/s11356-023-25221-3>
- Han, M., Wu, H., Chen, Z., Li, M., & Zhang, X. (2023). A survey of multi-label classification based on supervised and semi-supervised learning. *International Journal of Machine Learning and Cybernetics*, 14(3), 697–724. <https://doi.org/10.1007/s13042-022-01658-9>
- Iqbal, S., Atiq, M., Fayyaz, M., Zakria, M., Rajput, N. A., Kachelo, G. A., Ahmad, I., Usman, M., Mehmood, A., & others. (2024). Powdery Mildew of Mango: Current Status, Perspective and Emerging Tools for Management. *Agricultural Sciences Journal*, 6(1), 92–101. <https://doi.org/10.56520/asj.v6i1.365>
- Kiloos, A. M., Joyce, D. C., & Abdul Aziz, A. (2024). Exploring the Challenges and Opportunities of Mango Export from Indonesia: Insights from Stakeholder Interviews. *The Qualitative Report*, 29(3), 811–830. <https://doi.org/10.46743/2160-3715/20246343>
- Kumari, N., Shukla, P. K., Singh, H., Singh, H. S., Pandey, S., & Singh, A. (2025). Identification of fungi causing pre-harvest anthracnose and anthracnose-like symptoms on mango fruits (*Mangifera indica*). *Canadian Journal of Plant Pathology*, 47(3), 210–226. <https://doi.org/10.1080/07060661.2025.2461015>
- Kusrini, K., Suputa, S., Setyanto, A., Agastya, I. M. A., Priantoro, H., & Pariyasto, S. (2022). A comparative study of mango fruit pest and disease recognition. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, 20(6), 1264–1275. <https://doi.org/10.12928/telkomnika.v20i6.21783>
- Liang, Q., Pan, H., He, X., Wang, S., Hou, Y., Xiao, H., Xu, G., Yi, R., Lin, D., & Yang, Z. (2024). Population structure and genetic diversity of mango (*Mangifera indica* L.) germplasm resources as revealed by single-nucleotide polymorphism markers. *Frontiers in Plant Science*, 15. <https://doi.org/10.3389/fpls.2024.1328126>
- Liu, J., Xu, H., Peng, X., Wang, J., & He, C. (2023). Reliable composite fault diagnosis of hydraulic systems based on linear discriminant analysis and multi-output hybrid kernel extreme learning machine. *Reliability Engineering & System Safety*, 224, 108747. <https://doi.org/10.1016/j.ssci.2023.108747>

- 234, 109178.
<https://doi.org/https://doi.org/10.1016/j.res.s.2023.109178>
- Liu, Z., Bashir, R. N., Iqbal, S., Shahid, M. M. A., Tausif, M., & Umer, Q. (2022). Internet of Things (IoT) and Machine Learning Model of Plant Disease Prediction—Blister Blight for Tea Plant. *IEEE Access*, *10*, 44934–44944.
<https://doi.org/10.1109/ACCESS.2022.3169147>
- Matengu, T. T., Bullock, P. R., Mkhabela, M. S., Zvomuya, F., Henriquez, M. A., Ojo, E. R., & Fernando, W. G. D. (2024). Weather-based models for forecasting Fusarium head blight risks in wheat and barley: A review. *Plant Pathology*, *73*(3), 492–505.
<https://doi.org/10.1111/ppa.13839>
- moola, swathi, & Yakubreddy, K. (2024). Predicting Earthquake Parameters Using Random Forest: A Multi-Output Approach With P-Arrival Data. *SSRN Electronic Journal*.
<https://doi.org/10.2139/SSRN.5055638>
- Nega, A. (2025). Climate change impacts on agriculture: A review of plant diseases and insect pests in Ethiopia and East Africa, with adaptation and mitigation strategies. *Advances in Agriculture*, *2025*(1), 5606701.
<https://doi.org/10.1155/aia/5606701>
- Pansy, D. L., & Murali, M. (2023). UAV hyperspectral remote sensor images for mango plant disease and pest identification using MD-FCM and XCS-RBFNN. *Environmental Monitoring and Assessment*, *195*(9), 1120.
<https://doi.org/10.1007/s10661-023-11678-9>
- Rajput, N. K., & Grover, B. A. (2022). A multi-label movie genre classification scheme based on the movie's subtitles. *Multimedia Tools and Applications*, *81*(22), 32469–32490.
<https://doi.org/10.1007/s11042-022-12961-6>
- Raju, C., Pazhanivelan, S., Perianadar, I. V., Kaliaperumal, R., Sathyamoorthy, N. K., & Sendhilvel, V. (2024). Climate Change as an Existential Threat to Tropical Fruit Crop Production—A Review. *Agriculture; Basel*, *14*(11).
<https://doi.org/10.3390/agriculture14112018>
- Umar, U. D., Ahmed, N., Zafar, M. Z., Rehman, A., Naqvi, S. A. H., Zulfiqar, M. A., Malik, M. T., Ali, B., Saleem, M. H., & Marc, R. A. (2022). Micronutrients foliar and drench application mitigate mango sudden decline disorder and impact fruit yield. *Agronomy*, *12*(10), 2449.
<https://doi.org/10.3390/agronomy12102449>
- World Meteorological Organization (WMO). (2024). *Guide to Instruments and Methods of Observation (WMO-No. 8) 2024 Edition*.
- Yadav, S., & Tewari, A. S. (2025). CONF-RCNN: a conformer and faster region-based convolutional neural network model for multi-label classification of tomato leaves disease in real field environment. *Journal of Plant Diseases and Protection*, *132*(2), 61.
<https://doi.org/10.1007/s41348-024-01057-y>
- Yang, B., Li, M., Li, F., Wang, Y., Liang, Q., Zhao, R., Li, C., & Wang, J. (2024). A novel plant type, leaf disease and severity identification framework using CNN and transformer with multi-label method. *Scientific Reports*, *14*(1), 11664.
<https://doi.org/10.1038/s41598-024-62452-x>
- Zhou, C., Zhao, Z., Chen, W., Feng, Y., Song, J., & Xiang, W. (2024). Residual attention based multi-label learning for apple leaf disease identification. *Journal of Agricultural Engineering*, *55*(4).
<https://doi.org/10.4081/jae.2024.1595>