

PENGESANAN PENYAKIT TUMBUHAN MENGGUNAKAN PEMBELAJARAN MENDALAM

¹Yong Zi Xian, ¹Afzan Adam

¹Fakulti Teknologi & Sains Maklumat

43600 Universiti Kebangsaan Malaysia

Abstrak

Projek ini bertujuan untuk membangunkan sistem pengesanan penyakit tumbuhan menggunakan teknik pembelajaran mendalam bagi menangani cabaran yang dihadapi oleh petani dalam mengenal pasti dan mengurus penyakit tepat pada masanya. Penyakit tumbuhan adalah isu biasa dan kritikal dalam pertanian yang boleh memberi kesan teruk kepada pengeluaran dan kualiti hasil tanaman. Ia akan membawa kepada isu keselamatan makanan dan kerugian ekonomi jika tiada tindakan awal diambil. Oleh itu, untuk mengatasi isu-isu ini, satu sistem pengesanan penyakit tumbuhan menjadi penting dalam pertanian moden kerana ia membantu untuk mengelas dan mengesan penyakit tumbuhan. Sebelum ini, teknik pembelajaran mesin tradisional yang melibatkan pengekstrakan ciri, dan pengelasan telah digunakan secara meluas. Kaedah ini mengekstrak ciri tumbuhan daripada imej, seperti warna, tekstur dan bentuk untuk melatih model pengelas bagi membezakan antara tumbuhan yang sihat dan berpenyakit. Pada masa kini, model pembelajaran mendalam seperti Rangkaian Neural Konvolusi (CNN) juga sesuai untuk pengesanan penyakit tumbuhan kerana ia boleh melakukan analisis automatik pada jumlah data yang kompleks dan besar, belajar daripada data sedia ada, dan kemudian mengenal pasti corak dan meramalkan hasil dengan ketepatan yang lebih tinggi. Kajian ini memberi tumpuan kepada melatih dan mengoptimumkan model pengesanan penyakit tumbuhan yang mantap dan cekap. Kajian kami adalah untuk meneroka model yang lebih pantas dan lebih cekap dengan mengubah suai model pembelajaran mendalam sedia ada iaitu model CNN, kepada model CNN yang lebih pantas dan ringan untuk pengkomputeran yang lebih pantas dan cekap. Seterusnya, kami boleh melatih model dengan lebih banyak data untuk meningkatkan ketepatan model. Metodologi kami juga melibatkan augmentasi data, latihan model, pembelajaran pemindahan dan penilaian untuk mencapai prestasi yang tinggi. Set data yang digunakan ialah set data *Rice Leaf Diseases Images* dengan 5932 imej. Untuk model CNN ringan, model *MobileNetV2* dibina dengan memanfaatkan pembelajaran pemindahan pada model pra-latihan. Pada akhirnya, model ini berjaya mencapai ketepatan yang tinggi dengan ketepatan ujian 98% dan web dinamik mudah alih dibangunkan dengan antara muka yang mudah dan mesra pengguna untuk mengesan penyakit tumbuhan.

Kata Kunci: Pengesanan Penyakit Tumbuhan, Pembelajaran Mendalam, CNN, CNN ringan, Pembelajaran Pemindahan, Augmentasi Data

Abstract

This project aims to develop a plant disease detection system using deep learning techniques to address the challenges faced by farmers in identifying and managing diseases in a timely manner. Plant diseases are common and critical issues in agriculture that can severely affect crop production and quality. They will lead to food safety issues and economic losses if no early action is taken. Therefore, to overcome these issues, a plant disease detection system has become crucial in modern agriculture as it helps to classify and detect plant diseases. Previously, traditional machine learning techniques involving feature extraction, and classification have been widely used. These methods extract plant features from images, such as color, texture, and shape to train a classifier model to distinguish between healthy and diseased plants. Nowadays, deep learning models such as Convolutional Neural Networks (CNN) are also suitable for plant disease detection because they can perform automatic analysis on large and complex amounts of data, learn from existing data, and then identify patterns and predict outcomes with higher accuracy. This study focuses on training and optimizing robust and efficient plant disease detection models. Our study is to explore faster and more efficient models by modifying the existing deep learning model, namely CNN model, to a faster and lighter CNN model for faster and more efficient computing. Next, we can train the model with more data to improve the accuracy of the model. Our methodology also involves data augmentation, model training, transfer learning and evaluation to achieve high performance. The dataset used is the Rice Leaf Diseases Images with 5932 images. For the lightweight CNN model, the MobileNetV2 model is built by leveraging transfer learning on the pretrained model. By the end, this model successfully achieved high accuracy with a test accuracy of 98% and a dynamic web was developed with a simple and user-friendly interface for detecting plant diseases.

Keywords: Plant Disease Detection, Deep Learning, CNN, Lightweight CNN, Transfer Learning, Data Augmentation

1.0 PENGENALAN

Penyakit tumbuhan adalah sebahagian daripada alam semula jadi apabila tumbuhan menghadapi proses fisiologi yang tidak normal, yang akan menyebabkan gangguan pada struktur, pertumbuhan dan fungsi tumbuhan. Walau bagaimanapun, penyakit tumbuhan mempunyai kesan yang besar kepada sektor pertanian kerana ia mengurangkan pengeluaran dan kualiti hasil tanaman, membawa kerugian ekonomi kepada para petani dan mengancam keselamatan makanan. Mengesan dan melindungi tumbuhan daripada penyakit secara manual oleh kepakaran manusia dan pemerhatian melalui teknologi pendekatan tradisional merupakan proses yang memakan masa, kurang tepat dan memerlukan intensif buruh yang banyak. Oleh itu, pengesanan awal penyakit tumbuhan melalui sistem automasi adalah teknologi yang penting untuk dilaksanakan dalam pertanian bagi mengurangkan kerugian hasil. Dengan kemajuan terkini dalam teknologi moden, pengesanan penyakit tumbuhan dapat dilakukan dengan menganalisis imej daun tumbuhan menggunakan pembelajaran mesin. Model pembelajaran mesin, terutamanya kaedah pembelajaran mendalam seperti Rangkaian Neural Konvolusi (CNN) yang terdiri daripada pelbagai lapisan, mempunyai keupayaan pembelajaran

autonomi yang baik dan kebolehan pengekstrakan ciri, secara automatik mengekstrak ciri imej untuk klasifikasi dan pengecaman imej. CNN juga menunjukkan ketepatan yang tinggi dalam mengenal pasti penyakit dalam pelbagai tanaman. Kepentingan membangunkan sistem ini adalah untuk membantu para petani dengan pengesanan awal penyakit tumbuhan. Ini akan mengelakkan kerugian tanaman dan meningkatkan kualiti hasil, sekali gus meningkatkan produktiviti pertanian. Selain itu, sistem yang dibina adalah mesra pengguna dan mudah digunakan pada peranti mudah alih. Sebagai hasilnya, ini membolehkan pengguna seperti petani untuk diagnosis penyakit semasa tanpa memerlukan pengetahuan pakar. Pengguna hanya perlu memuat naik imej tumbuhan ke sistem, kemudian mereka akan menerima hasil ramalan dengan segera. Sistem ini juga dapat membantu sektor pertanian untuk mengurangkan pembaziran tenaga manusia melalui teknologi pendekatan tradisional kerana mereka tidak perlu mengesan secara manual.

2.0 KAJIAN LITERATUR

Dari kajian lepas yang saya kaji, kelima-lima kajian menggunakan kaedah klasifikasi, set data dan teknikal yang berbeza. Dalam kertas kerja Pembinaan Model Pengesanan Penyakit Tumbuhan Berdasarkan Pembelajaran Mendalam yang ditulis oleh Minah Jung, Jong Seob, Song, Ah-Young Shin, Beomjo Choi, Sangjin Go, Suk-Yoon Kwon, Juhan Park, Sung Goo Park, Yong-Min Kim yang diterbitkan pada 2023, memfokuskan pada membangunkan pengesanan penyakit tumbuhan yang cekap dan tepat pada model menggunakan teknik pembelajaran mendalam untuk mengautomasikan pengenalan penyakit awal dalam tumbuhan. Kertas kerja ini menggunakan set data *PlantVillage* yang terdiri daripada lebih 50000 imej daun tumbuhan dari pelbagai spesies dan kategori penyakit. Kaedah klasifikasi yang digunakan ialah ia menggunakan imej daun berpenyakit dan sihat berpasangan serta algoritma CNN yang terdiri daripada 5 model pralatih seperti *ResNet50*, *AlexNet*, *GoogLeNet*, *VGG19* dan *EfficientNet* untuk pengesanan dan klasifikasi penyakit. Selain itu, ia juga menggunakan pembelajaran pemindahan dan augmentasi data untuk meningkatkan keteguhan dan ketepatan model. (Jung et al. 2023) Sebagai hasilnya, semua model berjaya mengelas imej mengikut spesies dan jenis penyakit daun dan mencapai metrik prestasi tinggi dari segi ketepatan, kepersisan, dan kebolehesanan. Antara model pra-latihan CNN, *EfficientNet* menunjukkan ketepatan tertinggi iaitu 99.30%.

Seterusnya, dalam kertas kerja Pengesanan Penyakit dan Perosak Tumbuhan Berdasarkan Pembelajaran Mendalam: Ulasan yang ditulis oleh Jun Liu, Xuewei Wang diterbitkan pada 2021 memfokuskan pada mengenali dan mengklasifikasikan penyakit daun tumbuhan menggunakan teknik pembelajaran mendalam menggunakan Rangkaian Neural Konvolusi (CNN). Set data yang digunakan ialah set data *PlantVillage* dan set data *IP102* yang terdiri daripada 102 jenis perosak daripada pelbagai spesies tanaman. Manakala kaedah

klasifikasi yang digunakan dalam kertas kerja ini ialah Rangkaian Neural Konvolusi (CNN). Model ini menyepadukan buku residual daripada seni bina *ResNet* dan *EfficientNetV2* untuk mencipta model yang lebih cekap dari segi pengkomputeran untuk pengesanan penyakit dan perosak. Selain itu, kertas kerja ini juga menggunakan rangkaian segmentasi seperti *Mask R-CNN*, yang membahagikan lesi penyakit pada tahap piksel. Model dilatih menggunakan pembelajaran pemindahan yang membolehkan rangkaian mendapat manfaat daripada pengetahuan pra-latihan pada set data lain. (Liu & Wang 2021)

Seterusnya, dalam kertas kerja Pengesanan Penyakit Tumbuhan Menggunakan Model Pembelajaran Mesin yang ditulis oleh Praadeep Gupta, RS Jadon diterbitkan pada 2024 memfokuskan pada keberkesanan pembelajaran mesin dan pembelajaran mendalam dalam pengesanan penyakit tumbuhan daripada pelbagai tanaman. Set data yang digunakan ialah set data *PlantVillage* dan *Embrapa*. Kaedah klasifikasi menggunakan teknik termasuk Mesin Vektor Sokongan (SVM), Rangkaian Neural Konvolusi (CNN), *K-Means*, Model Pembelajaran Pemindahan, CNN Ringan (*VGG-ICNN*), *UNet*, *DenseNet201* dan *MobileNet*. (Jadon & Gupta 2023) Hasil menunjukkan bahawa model pembelajaran mendalam seperti *VGG-ICNN* mempunyai ketepatan paling tinggi iaitu mencapai 99.16% ketepatan manakala CNN mempunyai ketepatan 98.13%. SVM dan *K-Means* pula dihasilkan ketepatan yang lebih rendah, dengan SVM mencapai ketepatan 98.97%.

Seterusnya, dalam kertas kerja Pengesanan Penyakit Tumbuhan Menggunakan Pemprosesan Imej dan Pembelajaran Mesin yang ditulis oleh Pranesh Kulkarni, Atharva Karwande, Tejas Kolhe, Soham Kamble, Akshay Joshi, Medha Wyawahare diterbitkan pada 2021 memfokuskan pada pembangunan sistem yang cekap untuk mengesan penyakit tumbuhan menggunakan penglihatan komputer dan teknik pembelajaran mesin. Set data tersuai yang terdiri daripada imej pelbagai penyakit tumbuhan telah diguna. Manakala kaedah klasifikasi yang digunakan ialah Pengelasan hutan rawak untuk tugas pengelasan dan pengesanan. Selain itu, gabungan pelbagai Pohon Keputusan juga digunakan untuk mencapai ketepatan yang lebih tinggi dan untuk mengatasi *overfitting*. (Kulkarni et al. 2021) Hasil daripada kajian ini mencapai 93% ketepatan dalam mengesan 20 penyakit merentasi 5 jenis tumbuhan.

Akhir sekali, dalam kertas kajian Pengesanan Penyakit Daun Menggunakan Pembelajaran Mesin dan Pembelajaran Mendalam: Ulasan dan Cabaran yang ditulis oleh Chittabarni Sarkar, Deepak Gupta, Uesh Gupta, Barennya Bikash Hazarika diterbitkan pada 2023 memfokuskan pada pelbagai teknik pembelajaran mesin (ML) dan pembelajaran mendalam (DL) untuk pengesanan penyakit daun, menganalisis prestasinya dan mengenal pasti cabarannya. Kaedah klasifikasi yang digunakan ialah pembelajaran mesin dan teknik klasifikasi pembelajaran mendalam seperti Mesin Vektor Sokongan (SVM), k-Jiran terdekat (KNN), Rangkaian Neural

Buatan (ANN), Rangkaian Neural Konvolusi (CNN), CNN Mendalam (DCNN), Hutan Rawak, Pohon Keputusan. Pada masa yang sama, teknik pembelajaran pemindahan juga digunakan untuk mewakili ciri daripada model yang telah dilatih tanpa perlu melatih semula model tersebut. Model seperti *Inception*, *VGG*, *ResNet*, dan *Xception* digunakan secara meluas sebagai model pembelajaran pemindahan. (Sarkar et al. 2023) Dalam kajian ini, pembelajaran mendalam (DL) mencapai ketepatan lebih tinggi berbanding kaedah pembelajaran mesin (ML) tradisional.

3.0 METODOLOGI

3.1 Keperluan Pengguna

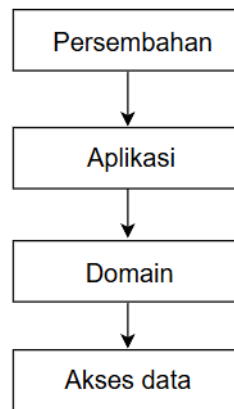
Keperluan pengguna ialah keperluan dan jangkaan yang pengguna miliki untuk produk atau perkhidmatan yang merangkumi fungsi asas, kepuasan emosi, matlamat, nilai, aspirasi dan keinginan. Untuk projek ini, sistem ini direka bentuk untuk memenuhi keperluan pengguna, terutamanya petani. Bagi para petani, sistem ini harus melibatkan penangkapan imej tumbuhan pada daun dan menerima diagnosis yang tepat dalam masa nyata. Web dinamik mesti menyokong antara muka berbilang bahasa dan berfungsi walaupun di luar talian. Selain itu, sistem ini harus meliputi pelbagai jenis tanaman dan penyakit biasa untuk memastikan ia mempunyai kebolegunaan yang meluas. Fungsi utama yang diperlukan dalam sistem ini ialah:

- i. Memuat naik imej tumbuhan
Sistem ini membenarkan pengguna memuat naik imej dalam format jpg/jpeg ke sistem.
- ii. Pengesanan penyakit tumbuhan
Selepas pengguna memuat naik imej daun tumbuhan, sistem mula mengenal pasti jenis dan penyakit tumbuhan tersebut.

3.2 Reka Bentuk Senibina

Reka bentuk senibina yang digunakan dalam sistem ini ialah Seni Bina Berlapis (*Layered Architecture*) untuk kebolehskalaan, kebolehselenggaraan, fleksibiliti dan 35 modulariti yang lebih baik. Terdapat beberapa lapisan dalam seni bina berlapis, iaitu lapisan persembahan, lapisan aplikasi, lapisan domain dan lapisan akses data. Setiap lapisan corak seni bina berlapis mempunyai peranan dan tanggungjawab khusus masing-masing.

Menggunakan Rajah Perhubungan Modul, data mengalir dari lapisan data ke lapisan domain untuk diproses. Kemudian ramalan dan keputusan kebolehtafsiran dihantar ke lapisan aplikasi. Akhir sekali, input pengguna diuruskan melalui lapisan pembentangan.

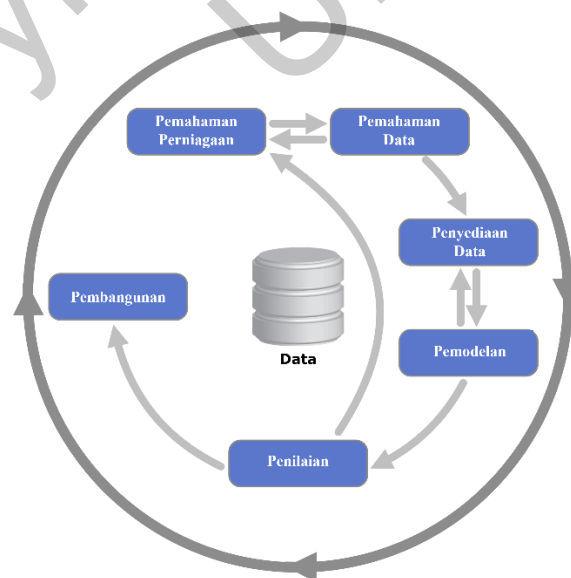


Rajah 1: Rajah Perhubungan Modul

3.3 Reka Bentuk Algoritma

Reka bentuk algoritma ialah proses yang melibatkan penciptaan prosedur langkah demi langkah atau set arahan untuk diikuti oleh komputer semasa melaksanakan tugas atau menyelesaikan masalah.

Dalam reka bentuk algoritma ini, *Cross-Industry Standard Process for Data Mining*, juga dikenali sebagai CRISP-DM menyediakan pendekatan berstruktur untuk merancang projek pengesanan penyakit tumbuhan ini. Terdapat enam fasa iaitu pemahaman perniagaan, pemahaman data, penyediaan data, pemodelan, penilaian dan pembangunan. (Wirth & Hipp 2000)



Rajah 2: Model CRISP-DM

Sumber: CRISP-DM Process diagram.png- Wikimedia Commons. (2012, April 26).

3.3.1 Fasa Pemahaman Perniagaan

Fasa Pemahaman Perniagaan memberi tumpuan kepada pemahaman objektif dan keperluan projek.

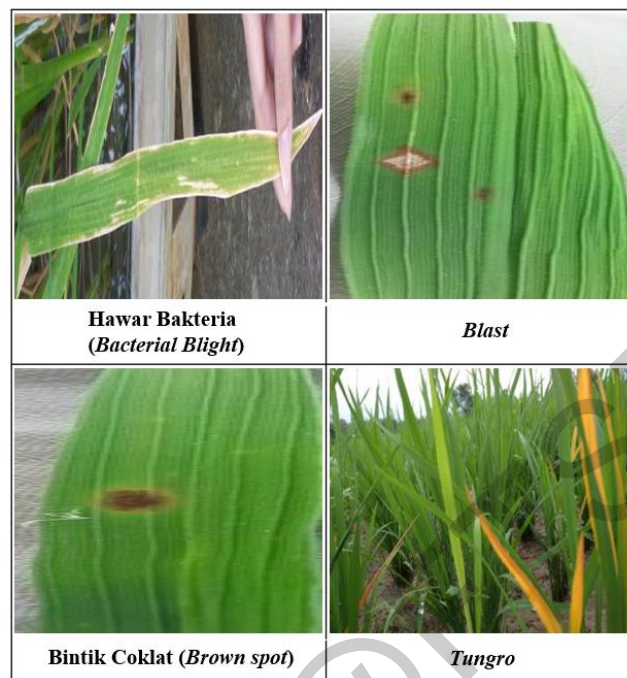
3.3.2 Fasa Pemahaman Data

Fasa Pemahaman Data memberi tumpuan kepada mengenal pasti, mengumpul dan menganalisis set data yang membolehkan pembangun mencapai objektif projek. Set data yang digunakan untuk latihan ialah set data *Rice Leaf Diseases Images*. (Sankalana n.d.) Ia ialah set data umum yang tersedia dan merangkumi koleksi 5932 imej. Set data ini memfokuskan kepada empat penyakit berbeza yang menjejaskan tanaman padi, iaitu Hawar Bakteria (*Bacteria Blight*), *Blast*, Bintik Coklat (*Brown Spot*) dan *Tungro*. Imej daun padi ini dalam format jpg, setiap satu menunjukkan simptom penyakit yang dinyatakan. Ini membolehkan diagnosis visual dan proses pengecaman imej.

Sebab memilih set data ini adalah kerana set data *Rice Leaf Diseases Images* cukup besar untuk latihan model. Model CNN memerlukan set data yang besar untuk mengelakkan masalah *overfitting* supaya dapat menghasilkan ketepatan dan generalisasi yang lebih baik.

Jadual 1: *Bilangan dalam setiap kategori penyakit*

Kategori Penyakit	Bilangan data
Hawar Bakteria	1584
<i>Blast</i>	1440
Bintik Coklat	1600
<i>Tungro</i>	1308



Rajah 3: Sampel imej setiap kategori

3.3.3 Fasa Penyediaan Data

Fasa penyediaan data menyediakan data yang bersih untuk set data terakhir untuk pemodelan selepas mengenal pasti isu kualiti dalam fasa sebelumnya. Semua imej yang berkaitan dengan hawar bakteria, *blast*, bintik coklat dan *tungro* dipilih. Imej diubah saiz kepada saiz standard, dan nilai piksel setiap imej dinormalkan.

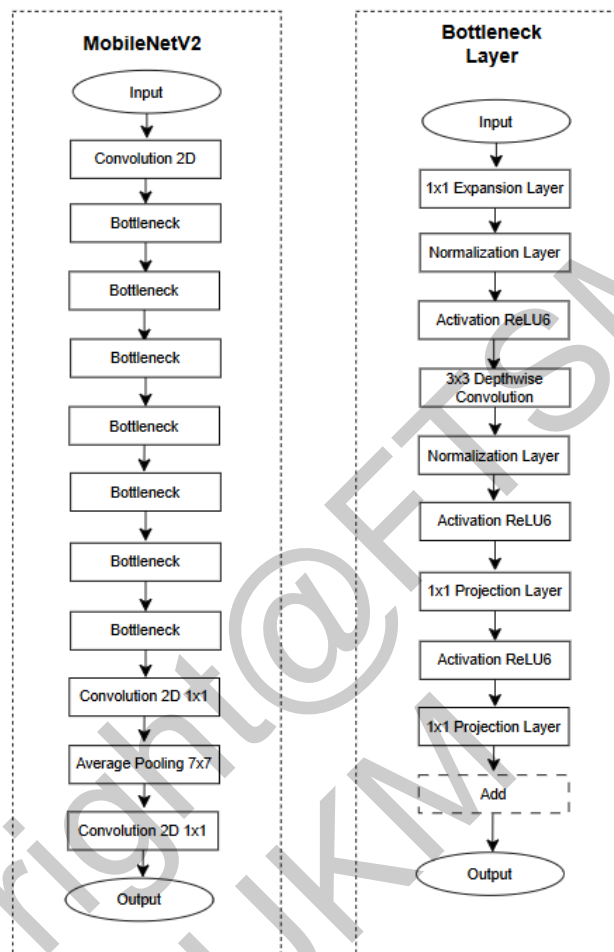
3.3.4 Fasa Pemodelan

Dalam projek ini, model CNN ringan dan pembelajaran pemindahan akan digunakan untuk membina model ketepatan yang tinggi dan teguh.

a) Model CNN ringan

Model CNN yang ringan digunakan untuk pengesanan penyakit tumbuhan kerana ia mempunyai bilangan lapisan yang agak kurang daripada model CNN, yang membolehkan model tersebut berprestasi lebih pantas dan mengambil storan pengiraan yang rendah. *MobileNet v2* (Girmaw & Taye 2025), *ShuffleNet* (Ma et al. 2018), *EfficientNet* (Tan & Le 2019), dan *VGG-ICNN* (Thakur, Sheorey & Ojha 2022) ialah contoh model CNN ringan yang menghasilkan ketepatan yang tinggi. Dalam model kami, Dalam model kami, *MobileNetV2* dipilih kerana saiznya jauh lebih kecil yang

membolehkan proses latihan dan pembangunan yang lebih pantas. (Girmaw & Taye 2025)



Rajah 4: Reka Bentuk *MobileNetV2*

Rajah di atas menunjukkan reka bentuk *MobileNetV2*. Lapisan dalam *MobileNetV2* termasuk lapisan input, lapisan konvolusi awal, lapisan *inverted residual bottleneck*, lapisan konvolusi akhir, *Global Average Pooling* (GAP) dan 42 lapisan *Dense* yang bersambung sepenuhnya.

b) Pemindahan Pembelajaran

Model pemindahan pembelajaran juga digunakan untuk pengesanan penyakit tumbuhan kerana ia boleh mengurangkan masa latihan, meningkatkan ketepatan dan memastikan prestasi yang mantap dengan memanfaatkan pengetahuan yang sedia ada. Dalam projek ini, pemindahan pembelajaran mempelajari ciri imej umum daripada set data yang besar seperti *ImageNet*. Kemudian, model pralatihan selanjutnya dilatih pada set data *Rice Leaf Diseases Images*, yang dikenali sebagai *fine-tuning*.

c) Augmentasi Data

Augmentasi data juga digunakan semasa latihan untuk meningkatkan generalisasi model di bawah pelbagai kondisi. Teknik augmentasi data yang akan saya laksanakan ialah putaran rawak 30 darjah dan 20% zum masuk atau keluar pada imej untuk meningkatkan variasi imej.

MobileNetV2 digunakan untuk pra-latihan pada *ImageNet*. Pada mulanya, keseluruhan model asas *MobileNetV2* pra-latihan dibekukan semasa fasa pertama latihan. Proses ini boleh memanfaatkan ciri yang dipelajari daripada set data yang besar (*ImageNet*) dan meningkatkan keupayaan untuk menggeneralisasi dan membuat ramalan yang tepat.

Selepas latihan awal, bahagian atas *MobileNetV2* dinyahbeku untuk penalaan halus dan meneruskan proses latihan dengan kadar pembelajaran yang lebih kecil.

Jadual 2: Konfigurasi Model

Komponen	Model Pra-latihan	Penalaan Halus
Model Asas	<i>MobileNetV2</i> dengan <i>include_top=False</i> , pralatihan pada <i>ImageNet</i>	<i>MobileNetV2</i> tetapi lapisan sebahagiannya tidak dibekukan untuk penalaan halus
Saiz Input Imej	(160, 160, 3)	(160, 160, 3)
Augmentasi Data	Putaran (30 darjah), Zum (0.2), <i>Fill Mode: 'nearest'</i>	Putaran (30 darjah), Zum (0.2), <i>Fill Mode: 'nearest'</i>
Penskalaan Semula	1./255	1./255
Lapisan Atas Tersuai	<i>GlobalAveragePooling</i> → <i>Dense(64, relu)</i> → <i>Dropout(0.3)</i> → <i>Dense(4, softmax)</i>	Sama
Lapisan Beku	Semua lapisan model asas dibekukan (<i>trainable=False</i>)	Lapisan 100 dan seterusnya tidak dibekukan
Pengoptimum	<i>Adam</i>	<i>Adam</i>
Kadar Pembelajaran	1e-4	1e-5
Saiz Kelompok	32	32
<i>Epoch</i>	10	10

3.3.5 Fasa Penilaian

Fasa penilaian adalah penting untuk menguji sama ada model telah memenuhi objektifnya. Semasa fasa penilaian, model pengesanan penyakit tumbuhan yang terlatih dinilai menggunakan metrik seperti ketepatan, *precision*, *recall* dan skor F1 dikenal pasti dan

digunakan untuk memperhalusi hiperparameter sehingga hasil yang dijangkakan dicapai. Formula setiap metrik adalah seperti berikut:

$$\text{Ketepatan} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{Skor F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

3.3.6 Fasa Penilaian

Web dinamik mudah alih dibangunkan dengan antara muka yang mudah dan mesra pengguna untuk mengesan penyakit tumbuhan. Pengguna boleh memuat naik imej tumbuhan, dan boleh mendapatkan ramalan penyakit tumbuhan selepas beberapa saat. Data penyakit tumbuhan baharu perlu dikemas kini secara berterusan kepada sistem. Selain itu, keputusan ramalan hendaklah dalam output yang jelas dan berformat untuk memudahkan pengguna.

4.0 HASIL

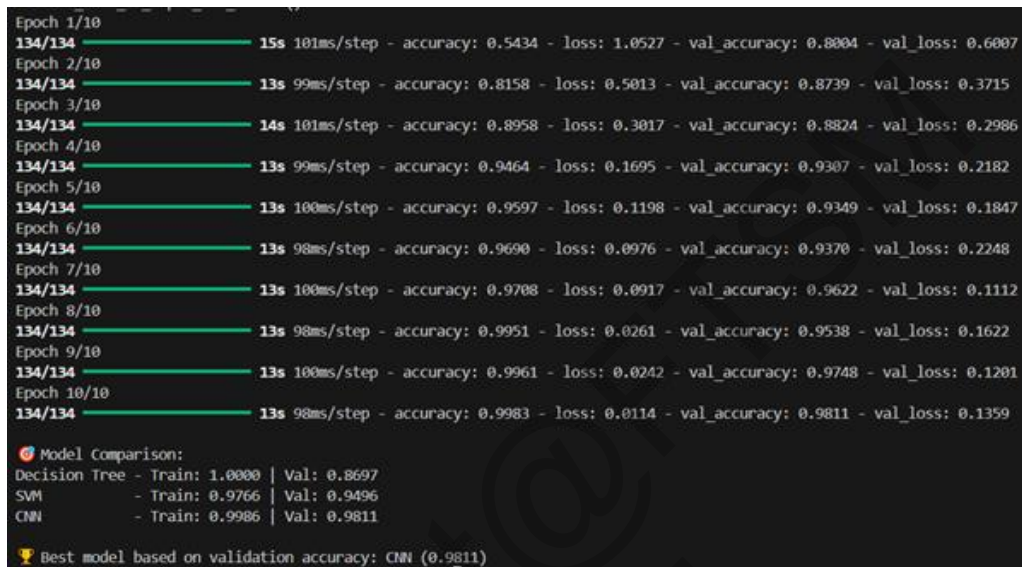
Model terlatih mobilenetv2_finetuned_model.h5 telah dinilai menggunakan evaluate_test.py. Output yang dijana dalam proses ini ialah ketepatan ujian, laporan klasifikasi dan matriks kekeliruan. Kemudian, pengujian keperluan kefungsiian telah dijalankan pada *UI Streamlit* dengan imej sampel yang berbeza.

4.1 Keputusan Objektif 1

Output Pemilihan Model:

Ketepatan latihan dan validasi untuk ketiga-tiga model (Pohon Keputusan, SVM dan CNN) dicetak dan perbandingan tiga model telah dilakukan untuk memilih model yang terbaik untuk

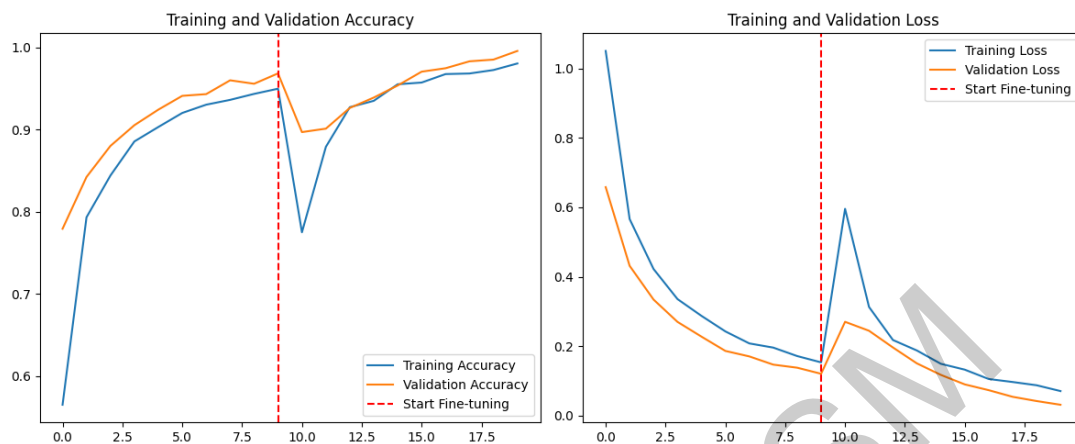
membina sistem selanjutnya. Ketepatan latihan mengukur prestasi model pada data terlatih, maksudnya apabila model mempunyai ketepatan latihan yang tinggi, ia dapat mempelajari corak dengan baik pada data latihan. Manakala ketepatan validasi menunjukkan seberapa baik model menggeneralisasi kepada data baharu yang tidak kelihatan, dan ia merupakan kunci utama untuk pemilihan model.



Rajah 5: Output Pemilihan Model

Rajah di atas menunjukkan output latihan ketiga-tiga model Pohon Keputusan, SVM dan CNN dalam pemilihan model. Berdasarkan output latihan untuk ketiga-tiga model, pohon keputusan mencapai 100% ketepatan latihan dan 86.97% ketepatan validasi. Keputusan ini menunjukkan bahawa model boleh menghafal dengan sangat cepat dan baik pada data latihan. Walau bagaimanapun, ketepatan validasi menurun dengan ketara pada data yang tidak kelihatan, menunjukkan bahawa ia menghadapi masalah *overfitting* dan mempunyai generalisasi yang lemah. Untuk Mesin Vektor Sokongan (SVM), ia mencapai 97.66% ketepatan latihan dan 94.96% ketepatan validasi. Keputusan ini secara relatifnya lebih baik daripada Pohon Keputusan kerana ia mempunyai ketepatan latihan dan validasi yang kukuh. Namun, jika berbanding dengan model, CNN yang telah mencapai 99.86% ketepatan latihan dan 98.11% ketepatan validasi, model CNN mempelajari corak visual dengan sangat baik dan mempunyai *overfitting* yang rendah disebabkan oleh konsistensi ketepatan latihan dan validasi. Oleh itu, CNN ialah model yang terbaik untuk dipilih kerana ia mencapai ketepatan validasi tertinggi (98.11%) dan kurang *overfit*.

Output model CNN ringan (*MobileNetV2*):



Rajah 6: Plot Ketepatan dan Kehilangan Latihan dan Validasi

Rajah di atas menunjukkan plot ketepatan dan kehilangan latihan dan validasi semasa proses latihan *MobileNetV2*. Pengelas *MobileNetV2* yang terlatih dan diperhalusi sepenuhnya disimpan dan sejarah latihan untuk kedua-dua pra-latihan dan penalaan halus diplot dalam satu plot. Berdasarkan plot ketepatan latihan dan validasi, ketepatan semakin meningkat sepanjang 10 *epoch* pertama dan mencapai 95.36% ketepatan latihan sehingga *epoch* 10. Ini menunjukkan bahawa kepala klasifikasi tersuai mempelajari ciri-ciri tersebut dengan pantas dan baik. Manakala ketepatan validasi mencapai 96.85%, menunjukkan bahawa fasa latihan awal adalah berkesan dan model digeneralisasikan dengan baik. Garis putus-putus merah adalah proses penalaan halus diambil, ia bermula dengan kedua-dua ketepatan latihan dan pengesanan mengalami penurunan yang jelas. Ini berlaku kerana lapisan yang tidak membeku dan kadar pembelajaran yang berubah. Kemudian kedua-dua latihan dan ketepatan validasi kembali normal dengan cepat dan terus meningkat. Pada akhirnya, ketepatan latihan akhir mencapai 97.9% dan ketepatan validasi mencapai 99.58% serta tiada *overfitting* yang ketara berlaku dalam model ini.

Manakala berdasarkan plot kehilangan latihan dan pengesanan, kedua-dua kehilangan latihan dan validasi menurun secara konsisten dari epoch 0-9, iaitu antara 0.2 - 0.15. Apabila penalaan halus bermula, kehilangan latihan meningkat secara tiba-tiba disebabkan lapisan yang tidak membeku. Kehilangan pengesanan juga sedikit meningkat selepas penalaan halus bermula. Kemudian, kedua-dua kehilangan latihan dan validasi berkurang dengan cepat. Pada akhirnya, kehilangan latihan ialah 0.07 dan kehilangan validasi ialah 0.03. Ini menunjukkan bahawa penalaan halus telah berjaya memperhalusi model tanpa *overfit*.

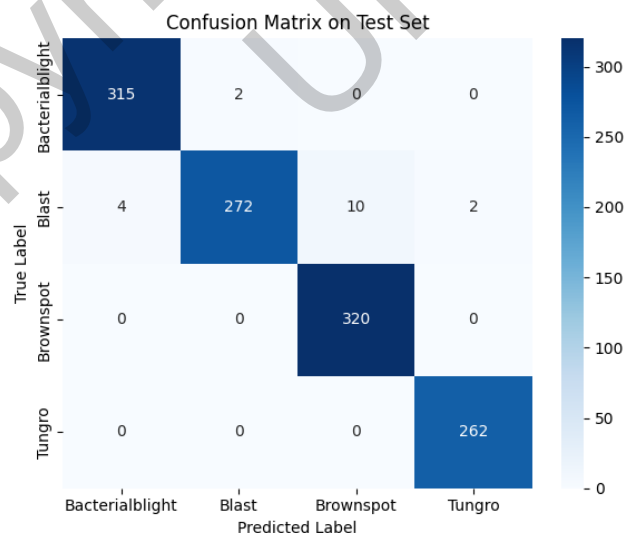
4.2 Keputusan Objektif 2

Ketepatan Ujian: 0.9848- Kehilangan Ujian: 0.0501

Classification Report:				
	precision	recall	f1-score	support
Bacterialblight	0.99	0.99	0.99	317
Blast	0.99	0.94	0.97	288
Brownspot	0.97	1.00	0.98	320
Tungro	0.99	1.00	1.00	262
accuracy			0.98	1187
macro avg	0.99	0.98	0.98	1187
weighted avg	0.99	0.98	0.98	1187

Rajah 7: Laporan Klasifikasi

Rajah di atas menunjukkan laporan klasifikasi yang dihasilkan semasa penilaian model. Berdasarkan keputusan daripada laporan klasifikasi, model ini berprestasi baik pada penyakit hawar bakteria, bintik coklat dan *tungro* dengan ketepatan dan *recall* yang hampir sempurna, yang menunjukkan sedikit salah klasifikasi. 6% daripada imej *blast* sebenar telah salah dikenal pasti sebagai penyakit lain kerana *blast* mempunyai *recall* yang lebih rendah iaitu 0.94. Purata makro mengira min tidak wajaran metrik untuk setiap kelas. Keputusan kedua-dua purata makro dan purata wajaran menunjukkan skor tinggi yang menunjukkan bahawa model mempunyai prestasi yang baik secara konsisten merentas semua kelas tanpa berat sebelah terhadap kelas yang lebih besar.



Rajah 8: Matriks Kekeliruan

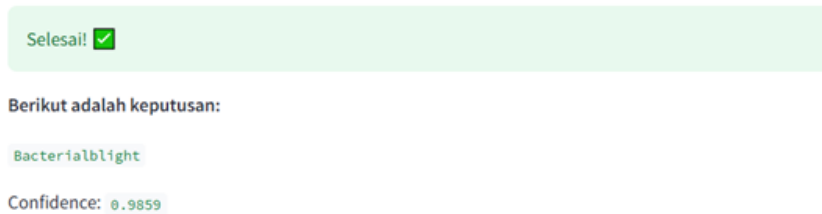
Rajah di atas menunjukkan matriks kekeliruan yang dihasilkan semasa penilaian model. Matriks kekeliruan secara visual mengesahkan dan menerangkan nombor dalam laporan

klasifikasi. Setiap baris mewakili kelas sebenar dan setiap lajur mewakili ramalan model. Nilai pepenjuru dalam matriks kekeliruan menunjukkan pengelasan yang betul dan nilai luar pepenjuru menunjukkan salah klasifikasi imej. Daripada matriks kekeliruan, 315 daripada 317 sampel hawar bakteria dalam set data ujian diramalkan dengan betul. Terdapat 2 imej hawar bakteria yang dikelaskan secara salah sebagai *blast*. Selain itu, untuk *blast*, 272 daripada 288 sampel dikenal pasti dengan betul sebagai *blast*. 4 imej salah dikelaskan sebagai hawar bakteria, 10 imej salah dikelaskan sebagai bintik coklat dan 2 imej salah dikelaskan sebagai *tungro*. Untuk bintik coklat dengan 320 sampel, kesemuanya diramalkan dengan betul. Oleh itu, tiada salah klasifikasi. Bagi *tungro* dengan 262 sampel, kesemuanya juga diramalkan dengan betul. Oleh itu, tiada juga salah klasifikasi. Oleh itu, kelas *blast* mempunyai kekeliruan yang paling banyak, dan biasanya ia disalahklasifikasikan sebagai bintik coklat dan model ini sangat baik dalam mengenal pasti bintik coklat dan *tungro*.

4.3 Pengujian Keperluan Fungsian



Rajah 9: Pilih dan Muat Naik Imej Tumbuhan



Rajah 10: Menunjukkan Keputusan Ramalan

Rajah di atas menunjukkan pengujian keperluan kefungsiian untuk muat naik imej dan paparan output ramalan. Imej berjaya dimuat naik dan keputusan ramalan dalam label kelas dan skor ketepatannya dipaparkan dalam format yang betul. Hanya jenis fail imej boleh dipilih semasa memuat naik imej. Keputusan ramalan dipaparkan dalam 2-3 saat.

5.0 KESIMPULAN

Terdapat beberapa kekuatan sistem yang dibangunkan. Model *MobileNetV2* mencapai ketepatan tinggi dengan ketepatan keseluruhan 98% pada set ujian, menunjukkan generalisasi yang kukuh. Ketepatan latihan dan pengesanan yang mempunyai jurang rendah di antaranya menunjukkan bahawa tidak berlaku *overfitting*. Selain itu, model yang dibina adalah model ringan yang sesuai untuk penggunaan pada mudah alih. Ia juga menunjukkan pembelajaran pemindahan yang berkesan dengan memanfaatkan pemberat *ImageNet* yang telah terlatih. Ini membantu mengurangkan masa latihan dan meningkatkan prestasi. Tambahan pula, antara muka yang ringkas membolehkan pengguna menggunakan sistem dengan mudah tanpa memerlukan pengetahuan teknikal. Walaupun terdapat banyak kekuatan pada sistem, sistem yang dibangunkan masih mempunyai beberapa kekangan. Antaranya, terdapat salah klasifikasi antara *blast* dan bintik coklat, yang mungkin disebabkan oleh persamaan visual. Selain itu, sistem ini hanya dapat mengesan penyakit empat kelas yang sudah dilatih, jadi ia sangat terhad apabila pelbagai jenis penyakit daun padi diperlukan untuk mengklasifikasikannya. Sistem ini bergelut dalam keadaan dunia sebenar disebabkan bunyi, variasi cahaya, sudut atau kekacauan latar belakang, oleh itu salah klasifikasi berlaku apabila diuji kepada imej dunia sebenar. Oleh itu, terdapat beberapa penambahbaikan kajian pada masa hadapan yang boleh dilakukan untuk meningkatkan lagi prestasi sistem pengesanan penyakit tumbuhan. Pertama, memperluas kelas penyakit dengan mengumpul lebih banyak set data penyakit padi tambahan untuk meningkatkan kebolehskalaan sistem. Selain itu, melatih model pada imej dengan pelbagai variasi, cahaya, sudut dan latar belakang untuk meningkatkan keteguhan dan mengurangkan salah klasifikasi. Di samping itu, tambahkan ciri maklum balas pengguna kepada sistem untuk membolehkan pembelajaran automatik dan berterusan daripada kes penggunaan dunia sebenar.

6.0 PENGHARGAAN

Penyelia projek ini, iaitu Ts. Dr. Afzan Adam, keluarga dan rakan-rakan yang membantu memperkayakan hasil projek ini.

7.0 RUJUKAN

- Chaisamrej, T., Wanichsan, N. & Olanviwichit, R. 2019. DrLADA: Diagnosing black pepper pests and diseases with decision tree. *International Journal on Advanced Science, Engineering and Information Technology* 9(5): 1659–1665. doi:10.18517/ijaseit.9.5.6818.
- Girmaw,D.W.&Taye,G.B.2025. Mobilenetv2modelfordetectingandgradingdiabetic foot ulcer. *Discover Applied Sciences* 7: 268. doi:10.1007/s42452-025-06745-4. URL<https://doi.org/10.1007/s42452-025-06745-4>.
- Hosna, A., Merry, E., Gyalmo, J., Das, B. & Roy, A. 2022. Transfer learning: A friendly introduction. *Journal of Big Data* 9(1): 102. doi:10.1186/s40537-022-00652-w. URL<https://doi.org/10.1186/s40537-022-00652-w>.
- Jadon, R. & Gupta, P. 2023. Plant disease detection using machine learning models. *Kilby* 100: 7th.
- Jung, M., Song, J.S., Shin, A.Y., Choi, B., Go, S., Kwon, S.Y., Park, J., Park, S.G. & Kim, Y.M. 2023. Construction of deep learning-based disease detection model in plants. *Scientific reports* 13(1): 7331.
- Kulkarni, P., Karwande, A., Kolhe, T., Kamble, S., Joshi, A. & Wyawahare, M. 2021. Plant disease detection using image processing and machine learning. *arXiv preprint arXiv:2106.10698* .
- Liu, J. & Wang, X. 2021. Plant diseases and pests detection based on deep learning: a review. *Plant Methods* 17: 1–18.
- Ma, N., Zhang, X., Zheng, H.T. & Sun, J. 2018. Shufflenet v2: Practical guidelines for efficient cnn architecture design. *Proceedings of the European Conference on Computer Vision (ECCV)*.
- Makam,S.,Komatineni,B.K.,Meena,S.S.&Meena,U.2024. Unmannedaerialvehicles (uavs): an adoptable technology for precise and smart farming. *Discover Internet of Things* 4(1): 12.
- O’shea, K. & Nash, R. 2015. An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458* .
- Perez, L.&Wang,J.2017. Theeffectivenessofdataaugmentationinimageclassification using deep learning. *arXiv preprint arXiv:1712.04621* .

Raymond, R. & Savarimuthu, S.M.A. 2022. Software design patterns and architecture patterns—a study explored. 2022 5th International Conference on Contemporary Computing and Informatics (IC3I), hlm. 1998–2006. doi:10.1109/IC3I56241. 2022.10073279.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. & Chen, L.C. 2018. Mobilenetv2: Inverted residuals and linear bottlenecks. Proceedings of the IEEE conference on computer vision and pattern recognition, hlm. 4510–4520.

Sankalana, N. n.d. Rice leaf disease image dataset. URL <https://www.kaggle.com/datasets/nimalsankalana/rice-leaf-disease-image?resource=download>. Accessed: 2025-01-21.

Sarkar, C., Gupta, D., Gupta, U. & Hazarika, B.B. 2023. Leaf disease detection using machine learning and deep learning: Review and challenges. Applied Soft Computing 145: 110534.

Shorten, C. & Khoshgoftaar, T.M. 2019. A survey on image data augmentation for deep learning. Journal of Big Data 6(1): 60. doi:10.1186/s40537-019-0197-0.

Tan, M. & Le, Q. 2019. EfficientNet: Rethinking model scaling for convolutional neural networks. K. Chaudhuri & R. Salakhutdinov (pnyt.) Proceedings of the 36th International Conference on Machine Learning, jil. 97 of Proceedings of Machine Learning Research, hlm. 6105–6114. PMLR. URL <https://proceedings.mlr.press/v97/tan19a.html>.

Thakur, M.G., Rajendran, R., Sharma, N. & Rana, D.S. 2023. Plantix app: A success story of artificial intelligence in plant protection. ResearchGate URL https://www.researchgate.net/publication/370650382_Plantix_app_A_Success_story_of_artificial_intelligence_in_plant_protection.

Thakur, P., Sheorey, T. & Ojha, A. 2022. Vgg-icnn: A lightweight cnn model for crop disease identification. Multimedia Tools and Applications 82. doi:10.1007/s11042-022-13144-z.

Wirth, R. & Hipp, J. 2000. Crisp-dm: Towards a standard process model for data mining. URL <https://api.semanticscholar.org/CorpusID:1211505>

Yong Zi Xian (A194982)

Ts. Dr. Afzan Adam

Fakulti Teknologi & Sains Maklumat

Universiti Kebangsaan Malaysia