

TESIS

**PENINGKATAN PERFORMA CONVOLUTIONAL NEURAL NETWORK
UNTUK IDENTIFIKASI PENYAKIT DAUN KENTANG
MENGUNAKAN AUGMENTASI DATA**



Disusun oleh:

Nama : Abdul Jalli Rozaqi
NIM : 19.51.1224
Konsentrasi : Informatics Technopreneurship

**PROGRAM STUDI S2 TEKNIK INFORMATIKA
PROGRAM PASCASARJANA UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2021

TESIS

**PENINGKATAN PERFORMA CONVOLUTIONAL NEURAL NETWORK
UNTUK IDENTIFIKASI PENYAKIT DAUN KENTANG
MENGUNAKAN AUGMENTASI DATA**

**PERFORMANCE IMPROVEMENT OF CONVOLUTIONAL NEURAL
NETWORK FOR IDENTIFICATION OF POTATO LEAF DISEASE
USING DATA AUGMENTATION**

Diajukan untuk memenuhi salah satu syarat memperoleh derajat Magister



Disusun oleh:

Nama : Abdul Jalli Rozaql
NIM : 19.51.1224
Konsentrasi : Informatics Technopreneurship

**PROGRAM STUDI S2 TEKNIK INFORMATIKA
PROGRAM PASCASARJANA UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2021

HALAMAN PENGESAHAN

PENINGKATAN PERFORMA CONVOLUTIONAL NEURAL NETWORK UNTUK IDENTIFIKASI PENYAKIT DAUN KENTANG MENGGUNAKAN AUGMENTASI DATA

PERFORMANCE IMPROVEMENT OF CONVOLUTIONAL NEURAL NETWORK FOR IDENTIFICATION OF POTATO LEAF DISEASE USING DATA AUGMENTATION

Diperiapkan dan Disusun oleh

Abdul Jalil Rozaqi

19.51.1224

Telah Ditujikan dan Dipertahankan dalam Sidang Ujian Tesis
Program Studi S2 Teknik Informatika
Program Pascasarjana Universitas AMIKOM Yogyakarta
pada hari NamaHari, tanggal ujian tesis

Tesis ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Magister Komputer

Yogyakarta, 05 Juli 2021

Rektor

Prof. Dr. M. Suyanto, M.M.
NIK. 190302001

HALAMAN PERSETUJUAN

PENINGKATAN PERFORMA CONVOLUTIONAL NEURAL NETWORK UNTUK IDENTIFIKASI PENYAKIT DAUN KENTANG MENGGUNAKAN AUGMENTASI DATA

PERFORMANCE IMPROVEMENT OF CONVOLUTIONAL NEURAL NETWORK FOR IDENTIFICATION OF POTATO LEAF DISEASE USING DATA AUGMENTATION

Dipersiapkan dan Disusun oleh

Abdul Jalil Rozaqi

19.51.1224

Telah Ditujikan dan Dipertahankan dalam Sidang Ujian Tesis
Program Studi S2 Teknik Informatika
Program Pascasarjana Universitas AMIKOM Yogyakarta
pada hari Senin, 05 Juli 2021

Pembimbing Utama

Dr. Andi Sunyoto, M.Kom.
NIK. 190302052

Anggota Tim Penguji

Prof. Dr. Ema Utami, S.Si., M.Kom.
NIK. 190302037

Pembimbing Pendamping

M. Rudyanto Arief, M.T.
NIK. 190302098

Prof. Dr. Kusriani, M.Kom.
NIK. 190302106

Dr. Andi Sunyoto, M.Kom
NIK. 190302052

Tesis ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Magister Komputer

Yogyakarta, 05 Juli 2021

Direktur Program Pascasarjana

Prof. Dr. Kusriani, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN TESIS

Yang bertandatangan di bawah ini,

Nama mahasiswa : Abdul Jalil Rozaqi
NIM : 19.51.1224
Konsentrasi : Informatics Technopreneurship

Menyatakan bahwa Tesis dengan judul berikut:
Peningkatan Performa Convolutional Neural Network Untuk Identifikasi Penyakit Daun Kentang Menggunakan Augmentasi Data

Dosen Pembimbing Utama : Dr. Andi Sanyoto, M.Kom.
Dosen Pembimbing Pendamping : M. Rudyanto Arief, M.T.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Tim Dosen Pembimbing
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi

Yogyakarta, 05 Juli 2021
Yang Menyatakan,



Abdul Jalil Rozaqi

HALAMAN PERSEMBAHAN

Segala puji bagi Allah SWT, tiada daya dan upaya melainkan dengan pertolongan dan bimbingan-Nya penulis dapat menyelesaikan penelitian tugas akhir ini. Penulisan tugas akhir ini dipersembahkan oleh penulis kepada:

1. Kedua orang tua saya, Bapak Darsono, B.A dan Ibu Umi Salamah, S.Pd.I yang telah memberikan didikan, arahan, kontrol dan kasih sayang terbaiknya kepada semua anak-anaknya termasuk penulis dari ketika ruh dihembuskan hingga sekarang. Tak lupa juga do'a dari setiap saat yang selalu terucap dari kedua orang tua lah yang membuat penulis bisa menyelesaikan penelitian ini.
2. Bapak Dr. Andi Sunyoto, M.Kom dan Bapak Muhammad Rudyanto, M.Kom telah memberikan bimbingan dan dorongan selama pelaksanaan penelitian tugas akhir ini.
3. Penelitian ini juga saya persembahkan kepada seluruh keluarga, teman-teman dan semua pihak yang tidak bisa saya sebutkan satu persatu.

HALAMAN MOTTO

“Hidup Sekali Hiduplah yang Berarti, Berani Hidup tak Takut Mati,
Takut Mati Jangan Hidup, Takut Hidup Mati Saja.”

(K.H Hasan Abdullah Sahal)

“Yang Menang Belum Tentu Benar dan Yang Terpilih

Belum Tentu yang Terbaik”

(K.H Hasan Abdullah Sahal)

“Tulislah Apa yang Kamu Kerjakan dan Kerjakanlah Apa yang Kamu Tulis”

(Dr. Hamid Fahmy Zarkasyi, M.A.)

“The Steps You Take Don't Need to be Big, They Just Need
to Take You in Right Direction”

(Jemma Simmons)

KATA PENGANTAR

Alhamdulillah, pujian tiada henti dan rasa syukur tiada batas kepada Allah SWT yang telah memberikan hambanya kemudahan dan kelancaran dalam penyelesaian penelitian tugas akhir hingga dapat dibaca oleh para pembaca.

Penelitian tugas akhir atau yang biasa disebut Tesis ini disusun dalam rangka memenuhi salah satu syarat kelulusan perguruan tinggi Program Studi Strata-2 Magister Teknik Informatika di Universitas AMIKOM Yogyakarta dan meraih gelar Magister Komputer (M.Kom). Selain itu penelitian ini juga bertujuan untuk memberikan kontribusi pengetahuan dalam bidang pengolahan citra digital yang berfokus pada peningkatan performa dari arsitektur *Convolutional Neural Network* dengan menggunakan *Data Augmentation* untuk mengidentifikasi penyakit yang menyerang pada daun tanaman kentang.

Pada kesempatan ini penulis ingin menyampaikan ucapan terimakasih yang sebesar-besarnya kepada:

1. Bapak Prof. Dr. M. Suyanto, MM selaku Rektor Universitas AMIKOM Yogyakarta.
2. Ibu Prof. Dr. Kusriani, M.Kom selaku Direktur Program Pascasarjana Universitas AMIKOM Yogyakarta.
3. Bapak Dr. Andi Sunyoto M.Kom dan Bapak Muhammad Rudyanto Arief M.Kom selaku pembimbing dalam penyelesaian Tesis ini.

4. Kedua orang tua Bapak Darsono, B.A dan Ibu Umi Salamah, S.Pd.I yang telah memberikan dukungan moril maupun materil dan juga yang tidak pernah lupa mendo'akan penulis di setiap waktu.
5. Sasvita Diana Sari, S.AB yang telah membantu dalam penulisan dan penyusunan naskah penelitian ini.
6. Teman-teman dekat saya semasa kuliah yang telah banyak berbagi pengetahuan dan pengalaman.
7. Teman-teman Magister Teknik Informatika (MTI) angkatan 22 kelas A.

Penulis juga mengucapkan rasa terimakasih kepada semua pihak yang telah membantu penyelesaian penelitian tugas akhir atau Tesis ini yang tidak bisa penulis sebutkan satu persatu. Semoga Tesis ini dapat menambah pengetahuan dan juga memberikan manfaat kepada para pembaca maupun manfaat untuk penulis sendiri, semoga Tesis ini dapat digunakan sebagai salah satu referensi dalam mengembangkan penelitian yang masih dalam satu tema ataupun penelitian yang lain.

Yogyakarta, 05 Juli 2021

Penulis

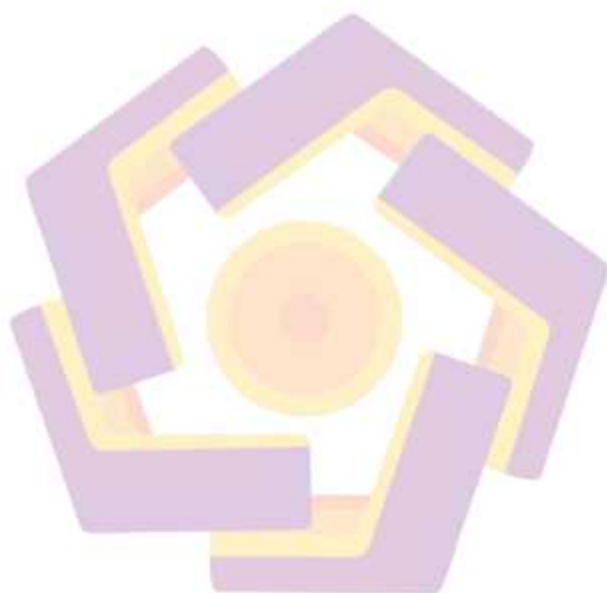
DAFTAR ISI

HALAMAN JUDUL.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERSETUJUAN.....	iv
HALAMAN PERNYATAAN KEASLIAN TESIS	v
HALAMAN PERSEMBAHAN	vi
HALAMAN MOTTO.....	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xiv
DAFTAR GAMBAR.....	xv
INTISARI.....	xviii
<i>ABSTRACT</i>	xix
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang Masalah	1
1.2. Rumusan Masalah.....	8
1.3. Batasan Masalah	8
1.4. Tujuan Penelitian	9
1.5. Manfaat Penelitian	10
BAB II TINJAUAN PUSTAKA.....	11
2.1. Tinjauan Pustaka.....	11
2.2. Keaslian Penelitian.....	15
2.3. Landasan Teori.....	19

2.3.1 Citra digital.....	19
2.3.2 Pengolahan citra digital.....	20
2.3.3 Klasifikasi.....	24
2.3.4 Convolutional Neural Network (CNN).....	25
2.3.5 Data augmentation.....	29
2.3.6 Confusion matrix.....	31
2.3.7 Confusion matrix multi class.....	34
BAB III METODE PENELITIAN.....	36
3.1. Jenis, Sifat, dan Pendekatan Penelitian.....	36
3.2. Metode Pengumpulan Data.....	36
3.3. Metode Analisis Data.....	37
3.4. Alur Penelitian.....	38
BAB IV HASIL PENELITIAN DAN PEMBAHASAN.....	44
4.1. Pengumpulan Data.....	44
4.2. Analisis Data.....	45
4.2.1 Persebaran data.....	45
4.2.2 Augmentasi data untuk menyeimbangkan data.....	46
4.2.3 Preprocessing data.....	49
4.2.4 Augmentasi data lanjutan.....	51
4.3. Model Klasifikasi Convolutional Neural Network.....	52
4.4. Skenario Percobaan.....	54

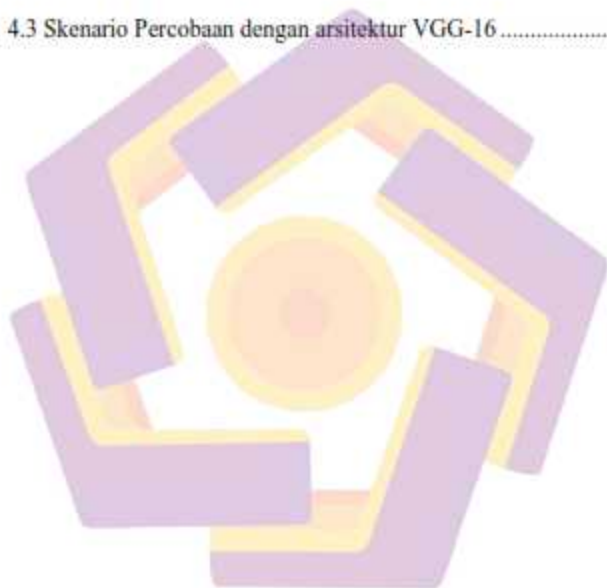
4.5. Hasil dan Evaluasi Model Klasifikasi.....	59
4.5.1 Hasil pelatihan model klasifikasi.....	60
4.5.2 Pengujian confusion matrix.....	61
4.5.2.1 Pengujian pada skenario 1.....	61
4.5.2.2 Pengujian pada skenario 2.....	64
4.5.3 Pengaruh data seimbang dan tidak seimbang.....	67
4.5.4 Analisis perbandingan augmentasi data.....	71
4.5.5 Rekomendasi teknik augmentasi data terbaik.....	75
4.5.6 Pengujian augmentasi data lanjutan.....	78
4.5.6.1 Pengujian pada skenario 9.....	78
4.5.6.2 Pengujian pada skenario 10.....	81
4.5.7 Perbandingan hasil seluruh skenario percobaan.....	84
4.6 Skenario Percobaan dengan Arsitektur VGG-16.....	86
4.6.1 Hasil pelatihan model klasifikasi arsitektur VGG-16.....	88
4.6.2 Perbandingan skenario 1 dan 2 pada arsitektur VGG-16.....	90
4.6.3 Perbandingan augmentasi data pada arsitektur VGG-16.....	93
4.6.4 Augmentasi data lanjutan pada arsitektur VGG-16.....	95
4.6.5 Perbandingan hasil seluruh skenario pada arsitektur VGG-16.....	96
BAB V PENUTUP.....	99
5.1. Kesimpulan.....	99

5.2. Saran	100
DAFTAR PUSTAKA	101



DAFTAR TABEL

Tabel 2.1. Matriks literatur review dan posisi penelitian.....	15
Tabel 2.2 <i>Confusion Matrix</i>	32
Tabel 2.3 <i>Confusion matrix multi class</i>	34
Tabel 4.1 Detail dataset daun kentang	44
Tabel 4.2 Skenario Percobaan.....	54
Tabel 4.3 Skenario Percobaan dengan arsitektur VGG-16	87



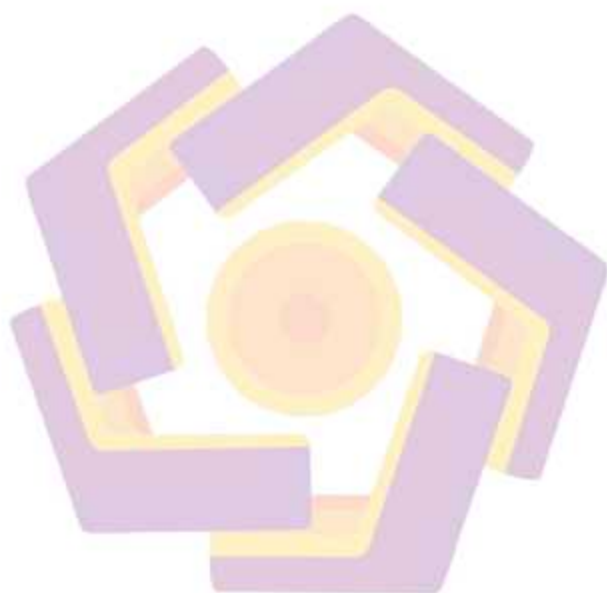
DAFTAR GAMBAR

Gambar 1.1. Penyakit daun pada tanaman kentang (a) <i>early blight</i> (b) <i>late blight</i>	2
Gambar 2.1. Representasi matriks citra digital	20
Gambar 2.2 Proses <i>convolution</i> dengan menggunakan <i>filter</i>	26
Gambar 2.3 Proses pada <i>Max Pooling</i>	27
Gambar 2.4 Arsitektur pada <i>Convolutional Neural Network</i>	28
Gambar 2.5 <i>Output</i> dari augmentasi data dengan <i>rotation</i>	29
Gambar 2.6 <i>Output</i> dari augmentasi data dengan <i>horizontal flip</i>	29
Gambar 2.7 <i>Output</i> dari augmentasi data dengan <i>vertical flip</i>	30
Gambar 2.8 <i>Output</i> dari augmentasi data dengan <i>zooming</i>	30
Gambar 2.9 <i>Output</i> dari augmentasi data dengan <i>shearing</i>	31
Gambar 2.10 <i>Output</i> dari augmentasi data dengan <i>brightness</i>	31
Gambar 3.1 Kerangka alur penelitian	38
Gambar 3.2 <i>Data augmentation</i> (a) citra awal (b) <i>rotation</i> (c) <i>horizontal flip</i> (d) <i>vertical flip</i>	41
Gambar 4.1 Persebaran jumlah data daun kentang	46
Gambar 4.2 Persebaran jumlah data setelah di augmentasi	49
Gambar 4.3 Pembagian data 80%:10%:10%	51
Gambar 4.4 Model arsitektur CNN	53
Gambar 4.5 Akurasi pada data <i>training</i> dan <i>validation</i>	60
Gambar 4.6 Confusion matrix skenario 1	62
Gambar 4.7 Confusion matrix skenario 2	65
Gambar 4.8 Grafik perbandingan hasil dari skenario 1 dan 2	67

Gambar 4.9 Perbandingan nilai recall tiap kelas pada skenario 1 dan 2.....	69
Gambar 4.10 Perbandingan nilai precision tiap kelas pada skenario 1 dan 2.....	70
Gambar 4.11 Perbandingan hasil dari pemanfaatan augmentasi data.....	73
Gambar 4.12 Hasil teknik augmentasi data zoom.....	75
Gambar 4.13 Hasil teknik augmentasi data brightness.....	75
Gambar 4.14 Hasil teknik augmentasi data rotation.....	76
Gambar 4.15 Hasil teknik augmentasi data horizontal flip.....	76
Gambar 4.16 Hasil teknik augmentasi data vertical flip.....	77
Gambar 4.17 Hasil teknik augmentasi data shear.....	77
Gambar 4.18 Confusion matrix pada skenario 9.....	79
Gambar 4.19 Confusion matrix pada skenario 10.....	82
Gambar 4.20 Perbandingan nilai akurasi dari data training dan data validasi.....	84
Gambar 4.21 Nilai Akurasi dari seluruh skenario percobaan.....	85
Gambar 4.22 Nilai Precision dari seluruh skenario percobaan.....	85
Gambar 4.23 Nilai Recall dari seluruh skenario percobaan.....	86
Gambar 4.24 Akurasi pada data training dan validation (VGG-16).....	89
Gambar 4.25 Grafik perbandingan skenario 1 dan 2 pada arsitektur VGG-16.....	90
Gambar 4.26 Nilai <i>recall</i> tiap kelas dari skenario 1 dan 2 pada VGG-16.....	91
Gambar 4.27 Nilai precision tiap kelas dari skenario 1 dan 2 pada VGG-16.....	92
Gambar 4.28 Perbandingan pemanfaatan augmentasi data pada VGG-16.....	94
Gambar 4.29 Perbandingan skenario 9 dan 10 pada VGG-16.....	95
Gambar 4.30 Perbandingan nilai akurasi dari data training dan data validation ..	96
Gambar 4.31 Nilai akurasi dari seluruh skenario dengan VGG-16.....	97

Gambar 4.32 Nilai Precision dari seluruh skenario dengan VGG-16..... 97

Gambar 4.33 Nilai Recall dari seluruh skenario dengan VGG-16..... 97



INTISARI

Kentang merupakan salah satu tumbuhan yang memiliki banyak manfaat bagi kehidupan umat manusia. Pengolahan pertanian kentang memiliki beberapa masalah diantaranya adalah penyakit yang menyerang pada daun. Perkembangan teknologi pada saat ini dapat membantu untuk menyelesaikan permasalahan para petani dengan menggunakan *image processing*, teknik ini dapat mengidentifikasi objek berdasarkan citra yang telah dikumpulkan terlebih dahulu. Metode Convolutional Neural Network (CNN) sangat cocok untuk klasifikasi citra apabila data yang digunakan memiliki jumlah yang besar, namun data yang bisa dikumpulkan tidak selalu dalam jumlah yang besar. Data yang tidak seimbang pada tiap kelasnya juga berpengaruh pada proses klasifikasi. Penelitian ini menggunakan teknik *data augmentation* untuk menyeimbangkan data, karena pada penelitian ini data yang digunakan memiliki jumlah yang tidak seimbang. Selanjutnya penelitian ini menggunakan teknik augmentasi data untuk memperbanyak variasi yang dimiliki citra dalam meningkatkan performa model klasifikasi.

Penelitian ini memiliki beberapa skenario percobaan untuk mendapatkan hasil yang paling baik. Skenario awal adalah membandingkan penggunaan data yang tidak seimbang dengan data yang telah diseimbangkan. Penelitian ini kemudian melakukan klasifikasi dengan menggunakan 6 teknik augmentasi data yaitu *zoom*, *shear*, *brightness*, *rotation*, *horizontal flip* dan *vertical flip*. Penggunaan teknik augmentasi data memerlukan pemilihan yang matang, karena hasil uji coba menunjukkan bahwa tidak semua teknik augmentasi data dapat meningkatkan performa model klasifikasi. Seluruh skenario percobaan yang dilakukan menggunakan arsitektur CNN yang diusulkan dan arsitektur VGG-16.

Hasil akhir dari penelitian ini adalah model klasifikasi dengan menggunakan data yang tidak seimbang membuat model klasifikasi sulit mengenali data pada kelas minoritas. Skenario percobaan terbaik adalah dengan penggunaan data yang telah diseimbangkan dan ditambah dengan tiga teknik augmentasi data terbaik yaitu *zoom*, *brightness* dan *rotation*. Nilai akurasi yang dihasilkan adalah 95% untuk penggunaan arsitektur CNN yang diusulkan dan 97% untuk arsitektur VGG-16.

Kata kunci: kentang, penyakit daun, image processing, CNN, VGG-16

ABSTRACT

Potato is a plant that has many benefits for human life. Potato farming has several problems, one of which is a disease that attacks the leaves. Technological developments at this time can help to solve the problems of farmers by using image processing. This technique can identify objects based on existing image data. The Convolutional Neural Network (CNN) method is suitable for image classification if the data used is extensive. However, the data that can be collected is not always in large amounts. Unbalanced data in each class also affects the classification process. This research uses data augmentation techniques to balance the data because the data in this study has an unbalanced amount. Furthermore, this research uses data augmentation techniques to increase the image variation in improving the classification model's performance.

This research has several experimental scenarios to get the best results. The first scenario is to compare the use of unbalanced data with balanced data. This research then performs a classification using 6 data augmentation techniques: zoom, shear, brightness, rotation, horizontal flip, and vertical flip. The use of data augmentation techniques requires careful selection because the trial results show that not all data augmentation techniques can improve the performance of the classification model. All experimental scenarios in this research use the proposed CNN architecture and the VGG-16 architecture.

The final result of this research is the classification model using unbalanced data makes it difficult for the classification model to recognize data in the minority class. The best experimental scenario is the use of data that has been balanced and added with the three best data augmentation techniques, namely zoom, brightness, and rotation. The resulting accuracy value is 95% for using the proposed CNN architecture and 97% for the VGG-16 architecture.

Keywords: potato, leaf disease, image processing, CNN, VGG-16

BAB I

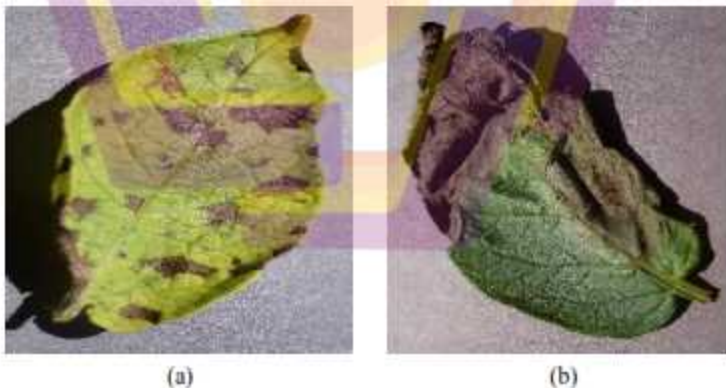
PENDAHULUAN

1.1. Latar Belakang Masalah

Seiring dengan berjalannya waktu banyak perkembangan yang dialami oleh umat manusia dalam memenuhi kebutuhan hidup, baik perkembangan dalam berpolitik, perkembangan perlindungan hak kemanusiaan, perkembangan kesehatan, perkembangan ekonomi juga perkembangan pertanian sebagai penunjang kehidupan untuk bertahan hidup. Perkembangan yang dirasakan manusia tidak semuanya memiliki dampak positif dikarenakan timbul juga perkembangan yang berdampak negatif terutama pada perkembangan pertanian, mulai banyak permasalahan yang timbul dalam perkembangan pertanian. Permasalahan pada perkembangan pertanian sangatlah meresahkan karena berhubungan dengan ketahanan pangan masyarakat. Beberapa permasalahan yang sering terjadi dalam perkembangan pertanian adalah rusaknya hasil pertanian dikarenakan penyakit yang menyerang pada daun tumbuhan, dimana fungsi daun bagi tumbuhan adalah untuk melakukan proses fotosintesis yang berarti proses pembuatan makanan untuk tumbuhan itu sendiri. Salah satu penyakit yang menyerang pada daun tumbuhan ini adalah penyakit busuk daun pada tanaman kentang yang dapat merusak daun dan mengakibatkan rusaknya hasil produksi kentang.

Tanaman kentang adalah tanaman yang memiliki banyak manfaat bagi keberlangsungan kehidupan manusia. Kentang merupakan salah satu makanan yang

mengandung karbohidrat yang merupakan bahan penting bagi manusia untuk menghasilkan energi sehingga kentang termasuk makanan pokok manusia (Furrer, Chegeni and Ferruzzi, 2018), selain itu kentang mengandung beberapa nutrisi penting yang dibutuhkan tubuh manusia yaitu vitamin C, kalium dan serat makanan (Beals, 2019). Pengembangan tanaman kentang menjadi sangat penting mengingat manfaat yang dihasilkan oleh tanaman ini sangatlah banyak bagi manusia. Namun, pertanian kentang memiliki beberapa masalah yang berakibat fatal, diantaranya adalah jenis penyakit yang menyerang pada daun kentang dimana penyakit ini menghambat produksi tanaman kentang (Tafesse *et al.*, 2018). Penyakit pada daun tanaman ini apabila dibiarkan akan menular dan dapat mengakibatkan penurunan produksi pangan tanaman kentang, bahkan apabila sudah sangat parah dapat menyebabkan gagal produksi.



Gambar 1.1. Penyakit daun pada tanaman kentang (a) *early blight* (b) *late blight*

Penyakit pada tanaman kentang yang paling banyak ditemui adalah busuk daun (*late blight*), dan juga ada penyakit bercak daun (*early blight*) penyakit ini memiliki sifat merusak daun yang berakibat merusak pertanian tanaman kentang

(Fitriana and Hakim, 2019). Penyebab terjadinya penyakit bercak daun (*early blight*) adalah jamur *Alternaria Solani*, penyakit *early blight* memiliki gejala awal ditandai dengan bercak-bercak melingkar di bagian tengah daun dan bisa juga pada bagian tepi daun seperti yang ditunjukkan pada Gambar 1.1 (a) penyakit ini dapat terus menyebar hingga merusak daun tanaman. Tempat dingin dan lembab adalah salah satu faktor penyakit busuk daun berkembang, selain itu penyebab penyakit busuk daun juga disebabkan oleh *Microbe Phytophthora infestans de Bary*. Gambar 1.1 (b) menunjukkan penyakit busuk daun pada tanaman kentang, penyakit ini memiliki gejala awal ditandai dengan munculnya lesi berwarna hitam atau kecoklatan pada daun di bagian tepi atau bagian tengah daun yang kemudian akan terus merambat hingga merusak tanaman (Kumbar *et al.*, 2019). Gejala awal dari penyakit tanaman kentang ini dapat dijadikan acuan oleh para petani untuk melakukan pencegahan penyakit ini menular, namun hal ini hanya dapat dilakukan oleh para petani yang memiliki pengetahuan mendalam tentang penyakit ini (Patil, Yaligar and Meena, 2017).

Penanganan masalah pada penyakit daun tanaman kentang dalam mengurangi dampak kerusakan telah banyak dilakukan dalam beberapa penelitian, tidak hanya penelitian didalam bidang pertanian saja tetapi pada bidang teknologi pun turut ikut andil dalam permasalahan ini, salah satunya adalah pemanfaatan bidang informatika dalam mengidentifikasi penyakit yang ada pada daun tanaman kentang dengan memanfaatkan data citra atau gambar daun. Identifikasi dengan menggunakan citra ini dapat kita sebut *image processing* atau pengolahan citra digital (Suttapakti and Bunpeng, 2019). *Image processing* dapat membantu para

pengelola pertanian dengan mengidentifikasi tanda-tanda awal yang dimiliki oleh penyakit, setelah mengetahui daun tanaman yang diduga berpenyakit maka para pengembang pertanian tanaman kentang dapat memberikan penanganan secara efektif dan efisien untuk menghindari penyebaran penyakit pada daun.

Perkembangan penelitian dalam bidang teknologi mengenai pengolahan citra digital saat ini telah banyak dilakukan, penelitian pengolahan citra digital yang dikembangkan dalam bidang pertanian baik untuk mengidentifikasi penyakit ataupun mengidentifikasi hasil produksi pertanian yang bagus. Beberapa penelitian yang sudah melakukan penelitian dengan pemanfaatan citra adalah penelitian yang dilakukan oleh Iqbal (Iqbal and Talukder, 2020) penelitian ini bertujuan untuk melakukan deteksi penyakit pada daun kentang menggunakan citra daun menggunakan algoritma *Random Forest*. Pada penelitian ini menggunakan teknik segmentasi untuk mendapatkan bagian-bagian yang terpenting dalam proses klasifikasi. Dalam melakukan klasifikasi pada tanaman yang mana daun memiliki tekstur karakteristik masing-masing maka penentuan bagian-bagian terpenting pada objek merupakan hal yang penting. Adapun penelitian yang menggunakan citra *grayscale* (abu-abu) dalam mengenali tekstur yang dimiliki oleh citra daun seperti yang dijelaskan pada penelitian milik Ni'mah (Ni'mah, Sutojo and Setiadi, 2018) dalam menentukan tanaman obat berdasarkan karakteristik pada citra daun. Penelitian ini mengekstraksi ciri untuk mendapatkan karakteristik penting dari citra daun dengan menggunakan *Gray Level Co-Occurrence Level Matrix (GLCM)*.

Penelitian yang dilakukan Rakhmawati (Rakhmawati, Pranoto and Setyati, 2018) menegaskan bahwa jumlah dari data yang digunakan mempengaruhi tingkat

akurasi. Penelitian ini menghasilkan tingkat akurasi yang cukup baik sebesar 80% dengan tahapan merubah ukuran citra ke 300x300 piksel yang kemudian dilakukan segmentasi dengan menggunakan *K-Means clustering* untuk memisahkan bagian-bagian yang dianggap penting, kemudian dilakukan ekstraksi dengan menggunakan Grey Level Co-Occurrence Level Matrix untuk mendapatkan karakteristik atau ciri khas pada tiap citra. Penelitian yang dilakukan oleh Islam. M (Islam *et al.*, 2017) dalam penelitiannya juga menggunakan metode *Support Vector Machine* (SVM) untuk mendeteksi penyakit pada daun kentang dengan jumlah data gambar daun yang digunakan 100 pada *early blight*, 100 pada *late blight* data daun yang sakit dan 100 daun yang normal atau sehat. Proses pada penelitian ini melakukan segmentasi citra dan ekstraksi tekstur citra daun menggunakan *Gray Level Co-occurrence Matrix* (GLCM). Hasil dari penelitian ini memiliki akurasi sebesar 95%. Penggunaan data yang seimbang juga berpengaruh pada ketepatan klasifikasi.

Arrofiqoh pada penelitiannya melakukan klasifikasi tanaman pada citra menggunakan *Convolutional Neural Network* (CNN) memiliki hasil akurasi 100% pada data training, 93% pada data validasi dan 83% pada data *testing*. Peneliti menerangkan bahwa metode CNN merupakan metode *deep learning* yang mampu mengenali objek secara mandiri, lapisan convolutional pada algoritma CNN memiliki fungsi untuk mengekstraksi fitur yang dimiliki oleh citra, peneliti juga menerangkan bahwa jumlah data training pada klasifikasi dengan menggunakan CNN mempengaruhi performa model klasifikasi (Arrofiqoh and Harintaka, 2018). Penelitian lain menggunakan CNN juga dilakukan untuk identifikasi penyakit daun tomat pada 5 jenis citra daun dengan total 5000 data yang digunakan, hasil dari

klasifikasi penyakit yang dilakukan menghasilkan akurasi yang sangat baik sebesar 96% (Widiyanto, Fitrianto and Wardani, 2019). Penelitian selanjutnya adalah deteksi penyakit daun tomat pada 9 kelas jenis penyakit dan 1 kelas daun sehat, penelitian ini membandingkan model CNN yang diusulkan dengan beberapa model *pre-trained* atau sudah dilatih terhadap dataset lain (Agarwal *et al.*, 2020). Model yang diusulkan mendapatkan akurasi yang lebih baik dibanding model *pre-trained* dengan tingkat akurasi sebesar 91%, peneliti merasa hasil akurasi yang didapatkan masih dapat ditingkatkan dengan beberapa percobaan atau *trial and error* pada lapisan konvolusi pada arsitektur CNN.

Klasifikasi citra dengan menggunakan *Convolutional Neural Network* (CNN) memiliki hasil yang baik pada data yang besar dan jumlah data yang digunakan pada tiap kelasnya merata atau seimbang. Permasalahan yang ditemui adalah data yang didapatkan untuk melakukan klasifikasi oleh para peneliti tidak selalu memiliki jumlah yang besar ataupun memiliki jumlah yang seimbang tiap kelasnya. *Overfitting* adalah ketika model klasifikasi terlalu berfokus ke data *training* sehingga akurasi pada data *training* akan lebih tinggi dibandingkan data *testing*, *overfitting* sering terjadi ketika jumlah data yang kecil atau data tidak seimbang digunakan pada model klasifikasi (Wu *et al.*, 2017). Ada beberapa teknik yang digunakan untuk mengatasi data yang kecil salah satunya adalah *data augmentation* (Buslaev *et al.*, 2018). Fungsi yang dimiliki dari *data augmentation* adalah dapat memperbanyak jumlah data gambar yang digunakan untuk klasifikasi, dengan teknik ini maka para peneliti dapat mengatasi permasalahan ketika data yang digunakan sedikit. *Data augmentation* juga memiliki fungsi untuk menambah

keberagaman data gambar, hal ini membuat model klasifikasi dapat bekerja lebih baik lagi dikarenakan informasi dari data yang digunakan menjadi lebih banyak dan lebih bervariasi (Tumewu, Setiabud and Sugiarto, 2020).

Berdasarkan uraian yang telah disampaikan maka peneliti tertarik untuk melakukan penelitian yang bertujuan untuk membuat sistem yang dapat membantu para petani atau pengelola pertanian dalam mengidentifikasi penyakit pada daun kentang dengan memanfaatkan data gambar daun kentang. Identifikasi daun pada tanaman kentang ini memiliki tiga kelas yaitu kelas daun sehat, kelas daun berpenyakit *late blight* dan yang terakhir adalah kelas daun berpenyakit *early blight*. Rincian jumlah data yang dimiliki dari setiap kelas adalah daun sehat berjumlah 150 citra, *late blight* 1000 citra dan *early blight* 1000 citra. Data yang digunakan memiliki jumlah yang tidak seimbang, pada data daun sehat memiliki jumlah yang lebih sedikit dibandingkan dengan data daun *early blight* dan *late blight*. Peneliti akan melakukan augmentasi data untuk menghindari terjadinya *overfitting* ataupun performa yang jelek pada model klasifikasi. Maka, penelitian ini akan melakukan identifikasi menggunakan algoritma *Convolutional Neural Network* (CNN) yang merupakan salah satu dari metode *Deep Learning* yang memiliki kemampuan untuk mengekstraksi fitur pada lapisan *convolution* dan memanfaatkan *data augmentation* untuk menyeimbangkan data dan juga digunakan untuk peningkatan kinerja model CNN yang diusulkan.

1.2. Rumusan Masalah

Dari pemaparan latar belakang yang telah disampaikan, maka dapat diambil perumusan permasalahan sebagai berikut:

- a. Apa pengaruh dari penerapan *data augmentation* untuk menyeimbangkan data yang digunakan pada klasifikasi penyakit daun tanaman kentang dengan algoritma *Convolutional Neural Network (CNN)*?
- b. Arsitektur *Convolutional Neural Network (CNN)* seperti apa yang cocok untuk menangani masalah identifikasi penyakit daun kentang dengan membandingkan beberapa skenario percobaan agar mendapatkan akurasi tinggi?

1.3. Batasan Masalah

Pada suatu penelitian dibutuhkan pembatasan masalah sehingga penelitian dapat lebih terarah dan tidak menyimpang jauh. Maka batasan masalah pada penelitian ini adalah:

- a. Metode yang digunakan untuk klasifikasi citra adalah *Convolutional Neural Network (CNN)*.
- b. Arsitektur CNN yang digunakan adalah arsitektur yang diusulkan dengan menggunakan 4 lapisan convolution dan 4 lapisan max pooling.
- c. Data yang digunakan pada model klasifikasi yang akan diteliti terbagi kedalam 3 kelas berupa kelas daun sehat, daun berpenyakit *early blight* dan daun berpenyakit *late blight*.
- d. Data yang digunakan untuk penelitian ini bersifat data publik yang didapatkan dari website www.kaggle.com.

- e. Data yang digunakan adalah data daun kentang dengan perincian 150 data daun sehat, 1000 data *early blight* dan 1000 data *late blight* yang mana data ini termasuk *imbalance data*.
- f. Augmentasi data yang digunakan untuk menyeimbangkan data adalah *horizontal flip*, *vertical flip* dan *rotation*.
- g. Augmentasi data tambahan yang digunakan untuk pengujian dalam meningkatkan kinerja model klasifikasi adalah *zooming*, *shearing* dan *brightness*.
- h. Pengujian kinerja model klasifikasi CNN yang diusulkan menggunakan *Confusion Matrix Multi Class*.
- i. Arsitektur CNN tambahan yang digunakan untuk menguji efektifitas dari susunan skenario percobaan terbaik adalah VGG-16.

1.4. Tujuan Penelitian

Tujuan dari pelaksanaan penelitian ini adalah sebagai berikut:

- a. Mengetahui apakah penerapan *data augmentation* pada data yang tidak seimbang memiliki pengaruh terhadap model klasifikasi penyakit daun tanaman kentang.
- b. Mencari rancangan arsitektur model klasifikasi untuk mendapatkan akurasi yang terbaik pada identifikasi penyakit daun tanaman kentang menggunakan arsitektur Convolutional Neural Network (CNN).

1.5. Manfaat Penelitian

Manfaat dari penelitian yang dilakukan dibagi kedalam beberapa kategori adalah sebagai berikut:

- a. Manfaat bagi masyarakat
 - i. Hasil rancangan model klasifikasi dapat digunakan untuk mempercepat penanganan dalam mengidentifikasi penyakit daun tanaman kentang.
- b. Manfaat bagi peneliti
 - i. Rancangan model yang diterapkan dapat dijadikan pembelajaran sebagai tambahan ilmu pengetahuan baru.
 - ii. Memberikan kontribusi penelitian pada penerapan *data augmentation* untuk menangani data yang tidak seimbang.
 - iii. Memberikan kontribusi penelitian pada penerapan arsitektur *Convolutional Neural Network (CNN)* yang cocok digunakan dan menghasilkan tingkat akurasi yang tinggi.
 - iv. Hasil dari penelitian sebagai salah satu syarat penyelesaian studi.

BAB II

TINJAUAN PUSTAKA

2.1. Tinjauan Pustaka

Sebelum melaksanakan penelitian terkait identifikasi penyakit pada daun tanaman kentang, peneliti terlebih dahulu melakukan peninjauan pustaka terkait penelitian-penelitian yang sudah dilakukan oleh para peneliti sebelumnya sebagai acuan dalam pengembangan penelitian yang akan dilakukan. Dari peninjauan pustaka yang dilakukan peneliti didapat beberapa kajian penelitian terdahulu yang berkaitan dan relevan pada penelitian yang akan dilaksanakan.

Penelitian dengan tujuan untuk melakukan deteksi penyakit pada daun kentang menggunakan citra daun menggunakan algoritma *Random Forest*. Citra daun yang digunakan pada penelitian ini didapatkan dari dataset publik dengan jumlah total 450 buah citra pada daun kentang yang sehat dan daun kentang yang berpenyakit. Tahapan penelitian yang dilakukan adalah dengan melakukan segmentasi gambar kedalam beberapa bagian dengan menggunakan *Thresholding* pada citra HSV, tujuan dari segmentasi ini adalah untuk mendapatkan bagian-bagian yang terpenting dalam proses klasifikasi, karena yang ingin di deteksi adalah daun yang memiliki warna hijau dan kemudian penyakit pada daun yang berwarna coklat maka pada penelitian ini segmentasi dilakukan berdasarkan warna citra hijau dan coklat. Setelah proses segmentasi maka dilakukan pembagian data kedalam data latih sebesar 80% dan kedalam data uji sebesar 20%, hasil dari klasifikasi ini menghasilkan nilai akurasi sebesar 97 % dengan pembagian label data pada daun

sehat sejumlah 150 citra, daun early blight 150 citra dan daun late blight sejumlah 150 citra (Iqbal and Talukder, 2020).

Penelitian yang dilakukan oleh Islam. M (Islam *et al.*, 2017) dalam pada penelitiannya "*Detection of Potato Diseases Using Image Segmentation and Multiclass Support Vector Machine*" menggunakan metode Support Vector Machine (SVM) untuk mendeteksi penyakit pada daun kentang dengan jumlah data gambar daun yang digunakan 200 data daun yang sakit dan 100 daun yang normal atau sehat. Proses pada penelitian ini adalah dengan melakukan segmentasi citra untuk menampilkan penyakit pada daun saja tanpa menampilkan background dan daun yang normal dalam artian warna hijau daun, kemudian gambar daun akan di ekstrak teksturnya menggunakan *Gray Level Co-occurrence Matrix* (GLCM). Hasil dari penelitian ini memiliki akurasi sebesar 95%.

Arrofiqoh pada penelitiannya dalam melakukan klasifikasi tanaman pada citra menggunakan *Convolutional Neural Network* (CNN) memiliki hasil akurasi 100% pada data training, 93% pada data validasi dan 83% pada data testing. Peneliti menerangkan bahwa metode CNN merupakan metode *deep learning* yang mampu mengenali objek dengan proses pembelajaran secara mandiri, lapisan *convolutional* milik CNN memiliki fungsi untuk mengekstraksi fitur yang dimiliki oleh citra. Penelitiannya bertujuan untuk melakukan identifikasi jenis tanaman dengan menggunakan foto dari lima jenis tanaman berupa tanaman padi, bawang merah, kelapa, pisang, dan cabai yang mana dari masing-masing jenis terdapat 100 foto. Peneliti juga menerangkan bahwa jumlah data training yang digunakan dalam

pengklasifikasian tanaman dengan menggunakan *Convolutional Neural Network* (CNN) mempengaruhi hasil akurasi (Arrofiqoh and Harintaka, 2018).

Penelitian mengenai Klasifikasi penyakit daun kentang berdasarkan fitur tekstur dan fitur warna menggunakan Support Vector Machine. Pada penelitian ini dijelaskan bahwa jumlah dari data yang digunakan mempengaruhi tingkat akurasi yang dihasilkan. Penelitian ini menggunakan klasifikasi *Support Vector Machine* untuk mendeteksi penyakit pada tanaman kentang, pada penelitian ini menggunakan jumlah data 100 data pada citra *Early blight*, 100 data pada citra *Late Blight* dan 100 citra pada daun non-disease. Sebelum dilakukan klasifikasi ukuran citra dirubah ke 300x300 piksel yang kemudian dilakukan segmentasi dengan menggunakan *K-Means clustering* untuk memisahkan bagian-bagian yang dianggap penting, kemudian dilakukan ekstraksi dengan menggunakan Grey Level Co-Occurrence Level Matrix untuk mendapatkan karakteristik atau ciri khas pada tiap citra. Penelitian ini menghasilkan tingkat akurasi sebesar 87% (Rakhmawati, Pranoto and Setyati, 2018).

Penelitian selanjutnya adalah penelitian tentang *image processing* dengan judul *Tomato Leaf Disease Detection Using Convolutional Neural Network* yang dilakukan oleh Mohit Agarwal (Agarwal *et al.*, 2020). Pendeteksian penyakit pada daun tomat dengan memanfaatkan lapisan konvolusi untuk ekstraksi ciri pada citra daun, penelitiannya membandingkan model CNN yang diusulkan dengan beberapa model *pre-trained* atau sudah dilatih terhadap dataset lain. Model yang diusulkan mendapatkan akurasi yang lebih baik dibanding model *pre-trained* yang digunakan.

Model CNN yang diusulkan mendapatkan tingkat akurasi sebesar 91% dan untuk model *pre-trained VGG16* 77%, *Mobilenet* 63% dan *InceptionV3* 63%.

Penelitian selanjutnya adalah tentang penerapan data augmentation untuk menangani *imbalance* data yang dilakukan oleh (Nisa, Puspaningrum and Maulana, 2020). Tujuan dari penelitian ini adalah membuat model klasifikasi untuk identifikasi penyakit pada daun apel, algoritma yang digunakan adalah Convolutional Neural Network (CNN) dengan arsitektur Inception-V3. Jumlah data yang digunakan adalah 1.821 citra daun apel yang dibagi kedalam 4 kelas berupa 1 kelas untuk daun sehat dan 3 kelas untuk daun yang memiliki penyakit. Permasalahan yang dihadapi pada penelitian ini adalah berupa data yang dimiliki untuk klasifikasi memiliki persebaran jumlah data yang tidak ideal yaitu jumlah yang tidak seimbang di antara kelas-kelas yang digunakan. Penanganan masalah data yang tidak seimbang ini berupa penggunaan augmentasi data untuk menambah jumlah dari data citra yang digunakan. Hasil dari penelitian ini mendapatkan akurasi untuk tiap kelas adalah 90%, 62%, 94%, dan 92%, akurasi terkecil adalah milik dari kelas yang memiliki jumlah data paling sedikit.

2.2. Keaslian Penelitian

Tabel 2.1. Matriks literatur review dan posisi penelitian

Peningkatan Performa Convolutional Neural Network Untuk Identifikasi Penyakit Daun Kentang Menggunakan Augmentasi Data

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
1	Detection of Potato Disease Using Image Segmentation and Machine Learning	M. Iqbal and K. Talukder. <i>International Conference on Wireless Communications, Signal Processing and Networking, WISPNET</i> . 2020	Penelitian ini bertujuan untuk melakukan klasifikasi penyakit pada daun dengan melakukan thresholding HSV pada citra dan klasifikasi dengan algoritma Random Forest	Hasil dari penelitian ini adalah mendapatkan nilai akurasi 97% dengan menggunakan algoritma Random Forest, dimana sebelum dilakukan klasifikasi terlebih dahulu dilakukan ekstraksi warna dari citra daun yang digunakan, warna hijau untuk daun yang dianggap sehat dan warna coklat untuk daun yang dianggap memiliki penyakit.	Kelemahan dari penelitian ini adalah sedikitnya data yang digunakan, dan kurangnya pada fitur ekstraksi ciri yang hanya menggunakan warna hijau untuk daun sehat dan coklat untuk daun sakit.	Pada penelitian sebelumnya ekstraksi ciri dengan cara threshold pada warna hijau dan coklat. Pada penelitian yang akan dilakukan menggunakan Ekstraksi fitur lapisan convolutional yang mana dapat mengekstraksi karakteristik citra secara mendalam.
2	Detection of Potato Diseases Using Image Segmentation and Multiclass Support Vector Machine	M. Islam, A. Dinih, K. Wahid. <i>Canadian Conference on Electrical and Computer</i>	Tujuan penelitian ini adalah untuk melakukan klasifikasi penyakit pada daun kentang menggunakan	Pada penelitian ini melakukan segmentasi pemisahan object dengan background menggunakan fitur threshold warna	Meskipun penelitian ini mendapatkan akurasi yang bagus tetapi dalam penggunaan datanya tidak maksimal, karena jumlah data yang	Pada penelitian sebelumnya data yang digunakan berjumlah 300 data. Pada penelitian ini akan dilakukan dengan menggunakan data yang lebih banyak (2150

Tabel 2.1 Lanjutan

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
		Engineering, 2017	Algoritma Support Vector Machine dan GLCM untuk ekstraksi fitur pada citra yang sudah dilakukan segmentasi antara daun dan background.	kemudian ekstraksi fitur menggunakan GLCM. Kemudian Klasifikasi penyakit daun kentang menggunakan Algoritma Support Vector Machine menghasilkan akurasi sebesar 95%.	digunakan pada penelitian ini adalah 100 data pada Early Blight, 100 pada Late Blight dan 100 pada daun sehat.	citra) kemudian dilakukan augmentasi data untuk mencimbangkan data dari tiap kelas.
3	Klasifikasi Penyakit Daun Kentang Berdasarkan Fitur Tekstur Dan Fitur Warna Menggunakan Support Vector Machine	P. Rakhmawati, Y. Pranoto, E. Setyati. Seminar Nasional Teknologi dan Rekayasa (SENTRA). 2018	Penelitian ini bertujuan untuk melakukan klasifikasi penyakit pada daun kentang dengan 300 jumlah data pada 3 kelas menggunakan Grey Level Co-Occurrence Level Matrix untuk ekstraksi ciri, color moment untuk ekstraksi warna, dan Support Vector Machine untuk klasifikasinya	Hasil dari penelitian ini adalah tingkat akurasi sebesar 87%, kemudian dari masing-masing jenis ini memberikan nilai akurasi untuk citra early 90%, citra late 90% dan non disease 83.33 %.	Kelemahan penelitian ini adalah akurasi pada pendeteksian daun non-disease menurun sekitar 83%. Penyebabnya adalah pengaruh pola bercak pada daun yang diekstraksi sebagai acuan klasifikasi masih kurang dikarenakan data yang digunakan sedikit, sehingga tidak kaya akan informasi.	Pada penelitian sebelumnya data yang digunakan masih kurang, sehingga menghasilkan penurunan akurasi pada daun non-disease. Pada penelitian ini data yang digunakan lebih banyak dan akan dilakukan Augmentasi data guna menambah data sehingga kaya akan informasi untuk dilakukan klasifikasi.

Tabel 2.1 Lanjutan

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
4	Implementasi Metode Convolutional Neural Network Untuk Klasifikasi Tanaman Pada Citra Resolusi Tinggi	E. Arrofiqoh dan Harintaka. Jurnal Ilmiah Geomatika-JIG. 2018	Tujuan penelitian ini adalah melakukan klasifikasi jenis tanaman menggunakan 5 kelas jenis yaitu padi, bawang merah, kelapa, pisang, dan cabai dengan masing-masing data berjumlah 100	Penelitian ini menggunakan 70% data training dan 30% untuk data validasi dan testing. Hasil dari penelitian ini adalah akurasi yang didapat dengan klasifikasi convolutional neural network 100% pada data training, kemudian 93% pada data validasi. Kemudian diujikan dengan data testing dan mendapat akurasi sebesar 82%.	Kelemahan pada penelitian ini adalah data yang sedikit dan ditambah tidak menggunakan pre-processing data, hasil akurasi pada data testing yang menurun jauh dalam artian klasifikasi data terlalu berfokus pada data training.	Pada penelitian yang sebelumnya tidak dilakukan proses pre-processing dan penggunaan dataset yang masih kurang, sedangkan penelitian yang akan dilakukan menggunakan data yang lebih banyak dan juga menggunakan pre-processing berupa augmentasi data seperti zoom-range, rotation range, shear range, horizontal flip.
5	ToLeD: Tomato Leaf Disease Detection using Convolutional Neural Network	M. Agarwal, A. Singhh, S. Arjariac. Procedia Computer Science 167. 2020	Penelitian ini bertujuan untuk klasifikasi penyakit pada daun tomat dengan membandingkan model CNN yang diusulkan dengan model pre-trained VGG16.	Hasil dari penelitian ini adalah akurasi yang didapatkan dari VGG16 sebesar 63%, InceptionV3 64%, Mobilenet 77% dan Model CNN yang diusulkan 91%.	Saran dari penelitian ini adalah untuk memodifikasi model yang dibuat karena dirasa akurasi yang didapat masih kurang. Pada penelitian ini masih kurang membandingkan skenario dari perancangan arsitektur	Pada penelitian sebelumnya membandingkan model pre-trained VGG16, InceptionV3, Mobilenet dengan model CNN yang diusulkan. Pada penelitian yang akan dilakukan membandingkan beberapa skenario model klasifikasi CNN dengan penerapan augmentasi data yang

Tabel 2.1 Lanjutan

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
			InceptionV3 dan Mobilenet.		CNN supaya mendapatkan akurasi terbaik.	berbeda-beda untuk dianalisis supaya mendapatkan akurasi terbaik.
6	Penerapan Metode Convolutional Neural Network untuk Klasifikasi Penyakit Daun Apel pada Imbalanced Data	C. Nisa, E. Puspaningrum, H. Maulana. Prosiding Seminar Nasional Informatika Bela Negara (SANTIKA). 2020	Tujuan penelitian ini melakukan klasifikasi untuk identifikasi daun apel yang dibagi kedalam 4 kelas. Algoritma CNN dengan arsitektur Inception-V3 dan juga penerapan Augmentasi data untuk menangani masalah imbalanced data.	Hasil dari penelitian ini adalah tingkat akurasi yang cukup baik pada model klasifikasi CNN dengan arsitektur Inception-V3 dan pemanfaatan Augmentasi data untuk menyeimbangkan data yang digunakan. Akurasi pada masing-masing kelas 90%, 62%, 94%, dan 92%.	Kelemahan dari penelitian ini adalah tingkat akurasi dari salah satu kelas yang memiliki data paling sedikit mendapatkan akurasi terkecil sebesar 62%. Hal ini disebabkan augmentasi data dilakukan pada semua kelas sehingga data yang paling sedikit tetap akan memiliki perbedaan yang jauh dengan kelas lainnya.	Perbedaan penelitian yang akan dilakukan adalah melakukan beberapa percobaan skenario untuk mendapatkan akurasi terbaik. Salah satu skenario melakukan penyeimbangan data dengan menerapkan augmentasi data rotation, horizontal flip dan vertical flip pada data yang memiliki jumlah paling sedikit, sehingga jumlah data dari masing-masing kelas menjadi seimbang.

Pada Tabel 1.1 Matrik literatur review merupakan pemaparan perbedaan penelitian yang dilakukan sekarang dengan beberapa penelitian terdahulu, secara garis besar penelitian yang akan dilakukan adalah pemanfaatan lapisan konvolusi dari arsitektur *Convolutional Neural Network (CNN)* untuk mengekstraksi ciri atau karakteristik dari citra daun kentang yang digunakan dan juga pemanfaatan *data augmentation* untuk menyeimbangkan data yang tidak seimbang dan juga meningkatkan variasi dari data yang dimiliki sehingga kaya akan informasi untuk dapat dijadikan bahan dalam mendapatkan akurasi yang baik.

2.3. Landasan Teori

2.3.1 Citra digital

Citra merupakan representasi visual dari suatu objek, seseorang atau suatu lokasi peristiwa yang dihasilkan oleh sebuah diavis optis seperti kaca lensa, atau kamera. Citra digital merupakan representasi atas citra dua dimensi menggunakan sejumlah berhingga dari titik-titik yang biasanya dikenal dengan elemen-elemen gambar, pel, atau piksel. Setiap piksel direpresentasikan oleh satu atau lebih nilai numeris. Citra monokrom (keabuan) setiap nilai merepresentasikan intensitas dan piksel (dalam rentang $[0, 255]$). Pada citra berwarna direpresentasikan oleh 3 nilai warna (intensitas komponen merah (R), hijau (G), dan biru (B)) (Siahaan and Sianipar, 2020).

$$f(x) = \begin{bmatrix} f(0,0) & f(0,1) & \dots & f(0,N-1) \\ f(1,0) & f(1,1) & \dots & f(1,N-1) \\ \vdots & \vdots & \vdots & \vdots \\ f(M-1,0) & f(M-1,1) & \dots & f(M-1,N-1) \end{bmatrix}$$

Gambar 2.1. Representasi matriks citra digital

Pada Gambar 2.1. merupakan contoh penulisan citra digital kedalam suatu matriks, Citra digital didefinisikan sebagai fungsi dua dimensi $f(x, y)$, dimana x dan y yang dimaksudkan disini adalah koordinat spasial, dan amplitudo f pada sembarang pasangan koordinat (x, y) . Koordinat (x, y) menunjukkan letak/posisi piksel dalam suatu citra, sedangkan amplitudo $f(x, y)$ menunjukkan nilai intensitas warna citra (Nafi'yah and Mujilahwati, 2018).

2.3.2 Pengolahan citra digital

Pengolahan citra atau bisa disebut juga *image processing* adalah proses pengolahan pada suatu citra dengan hasil keluaran berupa citra juga (Nafi'yah and Mujilahwati, 2018). Awal mula dari pengolahan citra adalah perbaikan untuk citra, dengan perkembangan zaman dibarengi dengan perkembangan dunia komputer maka *image processing* semakin berkembang dan banyak digunakan untuk berbagai kebutuhan. Seiring perkembangannya, pengolahan citra memiliki tujuan utama sebagai berikut:

- a. Perbaikan kualitas yang dimiliki citra, di mana keluaran dari hasil pengolahan citra ini adalah dapat menampilkan informasi pada citra secara lebih jelas.
- b. Ekstraksi ciri dari suatu citra, hasil keluaran dari proses pengolahan ini adalah berupa informasi citra di mana manusia mendapatkan informasi ciri dari citra secara numerik atau dengan kata lain komputer melakukan interpretasi terhadap informasi yang ada pada citra melalui besar-besaran (berupa besaran numerik) data yang dapat dibedakan secara jelas.

Pengolahan citra digital dari tujuan utamanya telah banyak dikembangkan pada saat ini untuk diterapkan ke dalam berbagai bidang, penerapan pada berbagai bidang ini tak lain untuk membantu manusia dalam melakukan analisis dan pekerjaan (Hidayatullah, 2017). Bidang-bidang pengembangan pengolahan citra digital pada kehidupan sehari-hari manusia adalah sebagai berikut:

1. Bidang Militer

Pengolahan citra digital pada bidang militer dapat dimanfaatkan seperti untuk pelacakan target sehingga sebuah peluru kendali dapat mengenali sasarannya.

2. Bidang Industri

Bidang industri sangat membutuhkan kecepatan dalam proses industrinya, terutama pada bagian pemrosesan yang apabila ditangani oleh manusia akan membutuhkan waktu yang lama atau bahkan suatu pemrosesan yang tidak bisa ditangani manusia. Contoh untuk sortir suatu barang yang didasarkan

pada perbedaan warna, ukuran ataupun kode barang yang tertera pada barang tersebut.

3. Bidang Medis

Peran pengolahan citra digital pada bidang medis adalah terletak pada pekerjaan yang membutuhkan suatu ketelitian dalam menghitung dan menganalisis. Contoh pada diagnosis darah manusia, dengan memanfaatkan pengolahan citra digital seorang dokter dapat dipermudah pekerjaannya dalam mendiagnosis penyakit yang diderita pasien.

4. Bidang Transportasi

Bidang transportasi bisa sangat terbantu dengan menggunakan kamera CCTV yang terpasang di jalanan dalam memantau lalu lintas, dengan pemanfaatan pengolahan citra digital fungsi dari kamera CCTV bisa bertambah seperti untuk menghitung jumlah kendaraan yang melintas. Pengolahan citra juga dapat membantu menganalisa celah-celah jalan pada saat terjadi kemacetan, sehingga bisa mengarahkan kendaraan tertentu untuk menghindari kemacetan.

5. Bidang Hukum dan Keamanan

Bidang hukum dan keamanan juga dapat terbantu dengan pengolahan citra digital. Contoh ketika terjadi kasus pencurian, dengan memanfaatkan pengolahan citra digital maka dapat melacak pelaku pencurian dengan jejak sidik jari yang ditinggalkan di TKP (Tempat Kejadian Perkara).

6. Bidang Pemetaan Lahan Daerah

Pemetaan daerah menjadi bisa dilakukan dengan pemanfaatan citra digital, citra yang diambil oleh satelit dapat diolah untuk mendapatkan informasi. Contoh dengan mengolah citra dari satelit dapat digunakan untuk menghitung luas suatu hutan pada daerah tertentu.

7. Bidang Robotika

Bidang robotika sangat bergantung pada pengolahan citra digital, karena robot yang dirancang manusia biasanya membutuhkan kamera untuk mengenali daerah sekitarnya, dan untuk mengenalinya maka dibutuhkannya pengolahan citra.

8. Bidang Fotografi

Fotografi adalah bidang yang banyak menggunakan pengolahan citra digital, contoh suatu gambar iklan yang ditampilkan begitu menarik dan enak dipandang mata merupakan hasil dari editing menggunakan aplikasi seperti Photoshop atau Corel Draw.

9. Bidang Film

Bidang industri film saat ini sudah banyak memanfaatkan pengolahan citra untuk menghasilkan gambar video yang dapat dinikmati oleh para penonton. Contoh video manusia dapat bernafas di ruang angkasa, dibalik

itu semua adalah hasil pengolahan citra yang sebenarnya adalah perpaduan video dengan latar belakang green screen dan video ruang angkasa.

10. Pencarian Gambar Berdasarkan Kandungan Citra

Penerapan pencarian gambar berdasarkan kandungan citra dapat kita temukan pada *search engine* milik google.com, dengan mengunggah suatu citra maka google dapat menampilkan citra yang sejenis.

11. Pemahaman Kandungan Citra

Pengembangan pengolahan citra digital juga dapat digunakan untuk sebuah sistem dapat memahami kandungan dari sebuah citra. Contoh citra seorang anak sedang bermain sepeda di taman, maka sistem dapat mendefinisikan isi dari citra tersebut dan menjelaskan bahwa pada citra tersebut terdapat citra anak kemudian juga terdapat citra sepeda dan juga citra pohon, rumput dan lain sebagainya yang merupakan komponen dari suatu taman.

2.3.3 Klasifikasi

Klasifikasi banyak digunakan pada berbagai hal diantaranya adalah untuk deteksi kecurangan, pengolahan pelanggan, deteksi kematangan buah, identifikasi penyakit dan masih banyak yang lainnya. Klasifikasi sendiri memiliki pengertian suatu teknik yang digunakan untuk mempelajari sekumpulan data yang mana menghasilkan suatu aturan dan ketentuan tertentu yang bisa digunakan untuk mengenali data-data yang baru dan belum pernah ditemui sebelumnya (Suyanto, 2017).

Klasifikasi dalam *machine learning* dikategorikan ke dalam supervised learning, atau pembelajaran yang diawasi. Supervised learning memiliki artian bahwa data yang digunakan sudah dikelompokkan pada kelas-kelas yang sudah diberi label.

2.3.4 Convolutional Neural Network (CNN)

Convolutional Neural Network merupakan salah satu metode *Deep Learning* yang dapat digunakan untuk klasifikasi citra dalam artian dapat menerima input berupa citra, menentukan ciri atau aspek objek apa saja dalam sebuah citra yang bisa digunakan mesin untuk belajar mengenali gambar, dan membedakan antara satu gambar dengan yang lainnya (Shanmugamani, 2018). Pada arsitektur *Convolutional Neural Network* (CNN) dapat menangkap informasi kontekstual yang terkandung di dalam data seperti pixel yang saling berdekatan pada sebuah citra, yang berarti model *Convolutional Neural Network* dapat melakukan ekstraksi ciri pada sebuah citra secara mandiri (Heryadi and Irwansyah, 2020). Sebuah model *Convolutional Neural Network* pada dasarnya terdiri dari beberapa layer untuk proses penyelesaian klasifikasi pada citra sebagai berikut:

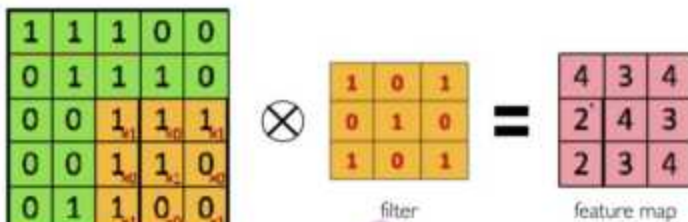
a. *Layer input*

Layer/lapisan ini adalah dimana data citra yang akan digunakan sebagai dataset untuk proses klasifikasi dimasukkan ke dalam model klasifikasi.

b. *Layer convolutional*

Layer convolutional atau lapisan konvolusi ini berfungsi untuk operasi *convolution* terhadap sejumlah node pada citra menggunakan beberapa *filter*,

pada operasi ini membantu *neural network* (jaringan syaraf) untuk mengenali objek seperti citra daun berdasarkan atribut-atributnya.



Gambar 2.2 Proses *convolution* dengan menggunakan *filter*

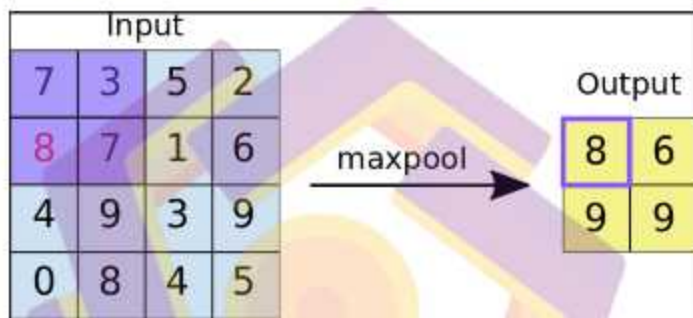
Lapisan konvolusi terdiri dari susunan sedemikian rupa neuron yang membentuk sebuah filter. Pada Gambar 2,2 adalah contoh proses *convolution* dengan menggunakan filter 3x3, pada proses konvolusi ini piksel pada citra dikalikan dengan piksel *filter* dan menghasilkan keluaran yang biasa disebut *feature map*.

c. Layer ReLu

Layer ini merupakan fungsi aktivasi dari output layer sebelumnya. Pada arsitektur *Convolutional Neural Network*, fungsi aktivasi terletak pada perhitungan akhir keluaran *feature map* atau sesudah proses perhitungan konvolusi atau *pooling* untuk menghasilkan suatu pola fitur. Pada fungsi ReLU atau Rectified Linear Unit melakukan treshold dari 0 hingga infinity, dalam artian fungsi ini masukan dari neuron-neuron berupa bilangan negatif, maka fungsi ini akan menerjemahkan nilai tersebut kedalam nilai 0, dan jika masukan bernilai positif maka output dari neuron adalah nilai aktivasi itu sendiri.

d. *Layer Pooling*

Pooling layer merupakan lapisan yang mengurangi dimensi dari feature map, pada proses *pooling* ini digunakan untuk mengurangi resolusi gambar dengan tetap mempertahankan informasi pada gambar, sehingga akan memberikan kecepatan yang lebih dalam melakukan komputasi untuk klasifikasi. Lapisan ini biasa diterapkan setelah lapisan konvolusi.

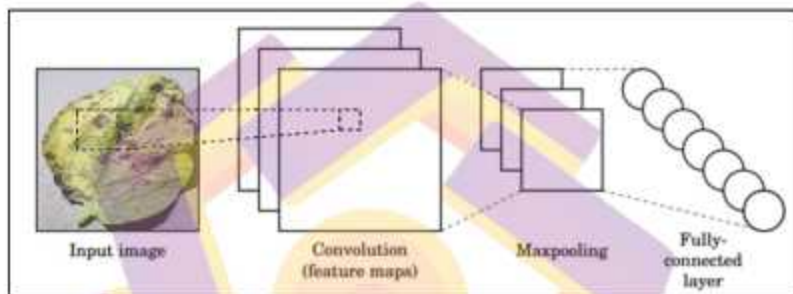


Gambar 2.3 Proses pada *Max Pooling*

Lapisan pooling memiliki beberapa jenis diantaranya adalah max pooling dan average pooling. Pada penelitian ini pooling yang digunakan adalah *max pooling* yang mana dalam proses ini pada setiap area dengan luas piksel tertentu, akan diambil satu buah piksel dengan nilai tertinggi. Lapisan pooling yang digunakan terdiri dari sebuah filter yang memiliki ukuran tertentu dan juga stride tertentu, stride adalah jumlah pergeseran dari filter pada saat proses pooling. Gambar 2.3 merupakan contoh proses *max pooling* pada citra dengan ukuran 2x2 dan 2 stride, dari beberapa nilai diambil nilai yang tertinggi dan menyimpannya di output, kemudian filter akan bergeser ke kanan sebanyak 2 stride atau sebanyak 2 kotak.

e. *Layer fully connected*

Lapisan ini berfungsi untuk menghitung hasil output dari layer convolution dan pooling. Pada lapisan konvolusi keluaran yang dihasilkan masih dalam bentuk multidimensional array, maka diperlukan proses yang bernama *flatten*, maksud *flatten* disini adalah mengubah keluaran hasil konvolusi (*feature map*) menjadi bentuk vektor.



Gambar 2.4 Arsitektur pada *Convolutional Neural Network*

Gambar 2.4 merupakan gambaran dari arsitektur *Convolutional Neural Network* (CNN). Proses pada arsitektur CNN dimulai dari *image input* (citra masukan) yaitu memasukkan citra yang digunakan sebagai dataset kedalam model. Proses selanjutnya adalah citra yang sudah dimasukkan akan dilakukan *convolution* untuk mendapatkan *feature maps* yang dimiliki oleh citra. Hasil dari proses konvolusi akan dilakukan *pooling* dengan *max pooling* untuk mengurangi dimensi citra menjadi lebih kecil sehingga proses klasifikasi data akan lebih cepat. Tahap akhir adalah hasil dari *pooling* akan dimasukkan pada proses *fully connected layer* untuk klasifikasi citra.

2.3.5 Data augmentation

Data augmentation merupakan teknik atau strategi yang dapat digunakan untuk menambahkan jumlah data citra yang dimiliki. Teknik *data augmentation* dapat membuat data citra yang digunakan untuk proses klasifikasi menjadi lebih bervariasi dan kaya akan informasi, teknik ini sangat berguna ketika data yang dimiliki kecil (Shanmugamani, 2018). Citra yang dihasilkan dari proses augmentasi akan memiliki dimensi yang sama dengan citra sebelum dilakukan augmentasi. Ada beberapa teknik *data augmentation* untuk menambah gambar sebagai berikut:

A. Rotation



Gambar 2.5 Output dari augmentasi data dengan *rotation*

Pada teknik ini citra akan diputar untuk menangani tingkat perubahan pada objek. Cara kerja teknik ini adalah dengan menentukan rentang rotasi kemudian data yang dihasilkan diputar secara acak. Gambar 2.5 adalah hasil keluaran dari teknik augmentasi dengan *rotation*.

B. Horizontal flip



Gambar 2.6 Output dari augmentasi data dengan *horizontal flip*

Pada teknik ini citra masukkan dicerminkan atau dibalik dalam arah horizontal. Pada Gambar 2.6 adalah hasil dari teknik *augmentation* dengan *horizontal flip*.

C. *Vertical flip*



Gambar 2.7 *Output* dari augmentasi data dengan *vertical flip*.

Vertical flip adalah kebalikan dari teknik *horizontal flip*, pada teknik ini citra masukkan dicerminkan atau dibalik dalam arah *vertical*. Pada Gambar 2.7 ditunjukkan hasil dari teknik *augmentation* dengan *vertical flip*.

D. *Zooming*



Gambar 2.8 *Output* dari augmentasi data dengan *zooming*.

Pada teknik ini citra diperbesar atau diperkecil secara acak dengan rentang *zoom* yang ditentukan, bagian citra yang diperbesar dilatih untuk menangani berbagai skala gambar. Gambar 2.8 adalah keluaran citra dari teknik *augmentation* dengan *zoom*.

E. *Shearing*



Gambar 2.9 Output dari augmentasi data dengan *shearing*

Shearing adalah teknik menggeser dan memiringkan bentuk citra. Teknik ini berbeda dari teknik *rotation* yang mana pada teknik ini ditetapkan satu sumbu kemudian meregangkan citra pada sudut lain yang disebut sumbu geser. Gambar 2.9 adalah contoh citra hasil dari teknik *shearing*.

F. *Brightness*



Gambar 2.10 Output dari augmentasi data dengan *brightness*

Pada Gambar 2.10 citra daun memiliki tingkat kecerahan yang berbeda-beda ada yang terang dan ada yang lebih gelap. Pada teknik *augmentation* dengan *brightness* citra masukan akan digeser atau dirubah nilai kecerahannya secara acak berdasarkan rentang nilai kecerahan yang ditentukan.

2.3.6 Confusion matrix

Metode *confusion matrix* adalah salah satu metode pengujian untuk kinerja suatu model klasifikasi. Metode ini berupa matriks yang berisi hasil prediksi model

klasifikasi dan data aktual atau data yang sesungguhnya yang dilakukan oleh sistem klasifikasi (Andono, T.Sutojo and Muljono, 2017). Terdapat empat istilah sebagai representasi hasil proses klasifikasi. Keempat istilah tersebut adalah *True Positive* (TP), *True Negative* (TN), *False Positive* (FP) dan *False Negative* (FN).

Tabel 2.2 *Confusion Matrix*

		Nilai Sebenarnya	
		<i>True</i>	<i>False</i>
Nilai Prediksi	<i>True</i>	TP (<i>True Positive</i>) <i>Correct result</i>	FP (<i>False Positive</i>) <i>Unexpected result</i>
	<i>False</i>	FN (<i>False Negative</i>) <i>Missing result</i>	TN (<i>True Negative</i>) <i>Correct absence of result</i>

Pada Tabel 2.2 dapat dijelaskan bahwa nilai *True Positive* (TP) merupakan data positif yang terdeteksi benar. *False Negative* (FN) merupakan kebalikan dari *True Positive*, sehingga data positif, namun terdeteksi sebagai data negatif. Sementara itu, *True Negative* (TN) merupakan jumlah data negatif yang terdeteksi dengan benar, sedangkan *False Positive* (FP) merupakan data negatif namun terdeteksi sebagai data positif. Dari empat istilah representasi hasil proses klasifikasi pada confusion matrix dapat diperoleh nilai *accuracy*, *precision*, dan *recall* (Harani and Hasanah, 2020).

A. *Accuracy*

Akurasi adalah tingkat kedekatan nilai dari hasil prediksi terhadap nilai yang sebenarnya, dengan kata lain nilai akurasi merupakan perbandingan antara data yang terklasifikasi benar dengan keseluruhan data.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.1)$$

Nilai akurasi didapatkan berdasarkan rumus persamaan 2.1. Dalam kasus pada penelitian ini berarti berapa persen daun yang diprediksi berpenyakit dan daun yang diprediksi sehat dari keseluruhan dataset.

B. Precision

Presi adalah tingkat ketepatan dari nilai yang diminta pengguna dengan jawaban yang diberikan oleh sistem, dalam artian rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif.

$$Precision = \frac{TP}{TP + FP} \quad (2.2)$$

Nilai presisi didapatkan berdasarkan rumus persamaan 2.2. Dalam kasus pada penelitian ini berarti berapa persen daun yang benar berpenyakit dari keseluruhan daun yang diprediksi berpenyakit.

C. Recall

Recall (sensitivitas) adalah tingkat keberhasilan sebuah sistem dalam menemukan kembali sebuah informasi. dengan kata lain rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif.

$$Recall = \frac{TP}{TP + FN} \quad (2.3)$$

Nilai recall didapatkan berdasarkan rumus persamaan 2.3. Dalam kasus pada penelitian ini berarti berapa persen daun yang diprediksi berpenyakit dibandingkan keseluruhan daun yang sebenarnya berpenyakit.

2.3.7 Confusion matrix multi class

Confusion matrix multi class memiliki pengertian yang sama dengan *confusion matrix*, hanya saja *confusion matrix* digunakan untuk menguji model klasifikasi yang memiliki dua kelas saja sedangkan *confusion matrix multi class* digunakan untuk menguji kinerja dari suatu model klasifikasi yang memiliki lebih dari dua kelas.

Tabel 2.3 *Confusion matrix multi class*

		Nilai Prediksi			
		Kelas 1	Kelas 2	...	Kelas n
Nilai Sebenarnya	Kelas 1	X_{11}	X_{12}	...	X_{1n}
	Kelas 2	X_{21}	X_{22}	...	X_{2n}

	Kelas n	X_{n1}	X_{n2}	...	X_{nn}

Tabel 2.3 adalah confusion matrix untuk pengujian klasifikasi pada 3 kelas atau lebih. Representasi hasil pengujian model klasifikasi dari confusion matrix multi class juga memiliki istilah yang sama seperti confusion matrix 2 kelas yaitu TP, FP, TN dan FN.

$$TFN_i = \sum_{\substack{j=1 \\ j \neq i}}^n x_{ij} \quad (2.4)$$

$$TFP_i = \sum_{\substack{j=1 \\ j \neq i}}^n x_{ji} \quad (2.5)$$

$$TTN_i = \sum_{\substack{j=1 \\ j \neq i}}^n \sum_{\substack{k=1 \\ k \neq i}}^n x_{jk} \quad (2.6)$$

$$TTP_{all} = \sum_{j=1}^n x_{jj} \quad (2.7)$$

Namun, pada *confusion matrix multi class* istilah representasi hasilnya adalah *Total False Negatif* (TFN), *Total False Positif* (TFP) dan *Total True Negatif* (TTN) yang mana dihitung untuk tiap kelas i berdasarkan rumus persamaan 2.4, 2.5 dan 2.6. *Total True Positif* (TTP) berarti total dari seluruh nilai *True Positif* (TP) dari tiap kelas dan dapat dihitung berdasarkan rumus persamaan 2.7 (Manliguez, 2016).



BAB III

METODE PENELITIAN

3.1. Jenis, Sifat, dan Pendekatan Penelitian

Penelitian ini termasuk kedalam jenis eksperimen atau didasarkan pada percobaan dan dengan melakukan pengujian. Percobaan yang dilakukan pada penelitian ini adalah pemanfaatan teknik *data augmentation* untuk mengatasi jumlah data yang tidak seimbang pada data citra daun kentang yang digunakan untuk bahan percobaan. Penelitian ini juga memanfaatkan *data augmentation* untuk mendapatkan variasi data yang lebih banyak dari data yang dimiliki. Kemudian di tahap selanjutnya adalah tahap *trial and error* pada arsitektur *Convolutional Neural Network (CNN)* untuk mendapatkan tingkat akurasi yang terbaik. Penelitian ini menggunakan pendekatan kuantitatif karena hasil dari penelitian ini berupa presentase angka. Sifat dari penelitian yang dilakukan adalah deskriptif, penelitian bersifat deskriptif merupakan penelitian yang membuat gambaran atau deskripsi dari percobaan yang dilakukan secara sistematis (Rukajat, 2018).

3.2. Metode Pengumpulan Data

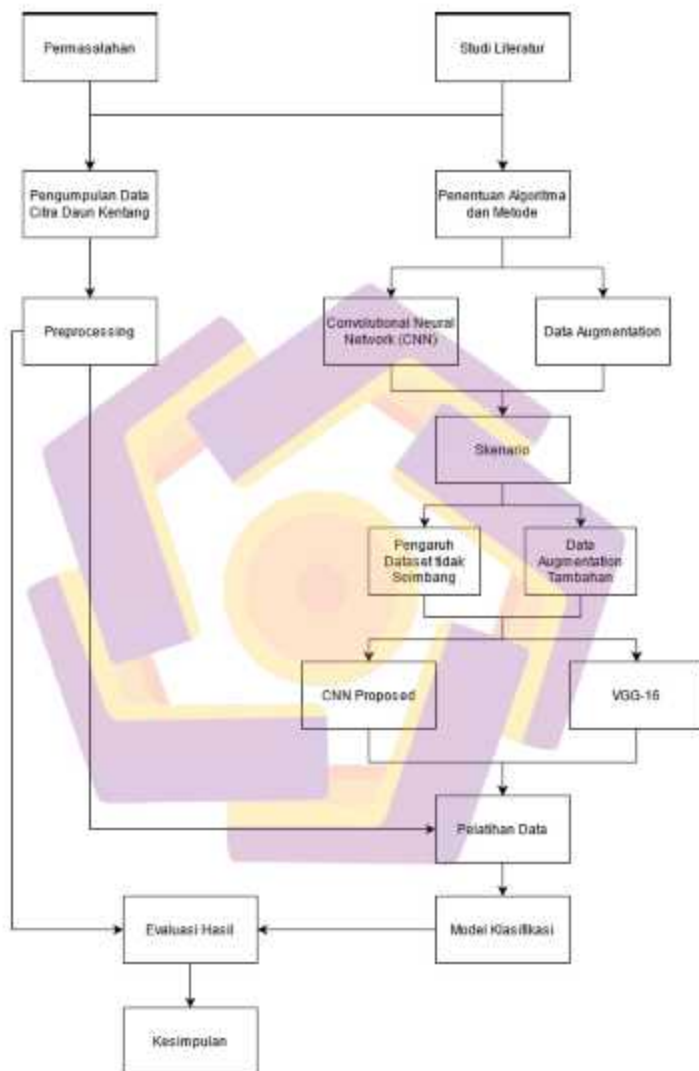
Data yang digunakan pada penelitian ini adalah dengan memanfaatkan data publik yang dapat diakses oleh semua orang. Penyedia data publik yang digunakan pada penelitian ini adalah sebuah website bernama www.kaggle.com, website ini memfasilitasi para peneliti untuk mendapatkan data yang bisa digunakan. Kaggle juga menjadi tempat bagi para peneliti ataupun orang lain yang ingin berkontribusi

dalam menyediakan data demi memajukan penelitian di dunia, website ini juga dapat digunakan para peneliti untuk membagikan hasil percobaan seperti uji coba dengan metode menggunakan bahasa pemrograman python. Penelitian ini menggunakan data yang didapatkan dari website kaggle dengan nama Plant Village. Data yang digunakan adalah data jenis gambar atau citra, peneliti menggunakan data citra daun kentang. Citra daun kentang yang didapatkan terbagi menjadi 3 kelas dengan total jumlah 2150 data, rincian dari data ini adalah 150 untuk data daun sehat, 1000 data untuk daun berpenyakit early blight dan 1000 data untuk daun berpenyakit late blight.

3.3. Metode Analisis Data

Metode analisis data merupakan tahapan pengelolaan data yang digunakan pada penelitian ini untuk mendapatkan hasil yang diharapkan dari permasalahan yang telah dirumuskan. Pada penelitian ini analisis yang digunakan berupa analisis kuantitatif dengan mengolah data yang telah dikumpulkan menggunakan metode *Convolutional Neural Network (CNN)*. Data daun tanaman kentang yang telah dikumpulkan akan dilakukan beberapa proses tertentu dengan bertahap sampai dapat digunakan proses klasifikasi antara daun yang sehat dan daun yang berpenyakit, penjelasan lebih detail nya dalam melakukan klasifikasi daun kentang ini akan dipaparkan pada alur penelitian.

3.4. Alur Penelitian



Gambar 3.1 Kerangka alur penelitian

Pada penyelesaian penelitian dalam mengidentifikasi penyakit pada daun tanaman kentang ada beberapa kegiatan yang dilakukan secara bertahap.

Kerangka metode penelitian yang akan dilaksanakan ditunjukkan pada Gambar 3.1, ada beberapa tahapan untuk menyelesaikan penelitian tentang identifikasi penyakit pada daun tanaman kentang. Tahapan pada penelitian ini dimulai dari pengumpulan dan identifikasi permasalahan, kemudian studi literatur untuk mengetahui beberapa permasalahan dan penyelesaian dari penelitian terdahulu yang relevan dengan penelitian saat ini. Setelah permasalahan dan studi literatur beranjak ke tahap selanjutnya yaitu penentuan metode dan algoritma yang digunakan untuk menyelesaikan permasalahan, selain menentukan algoritma maka perlu untuk menentukan dan mengumpulkan dataset yang akan digunakan. Tahapan-tahapan penelitian ini dapat dijabarkan lebih lanjut sebagai berikut:

a. Identifikasi Permasalahan

Tahap ini adalah menentukan permasalahan apa yang sedang terjadi pada suatu objek tertentu yang layak untuk diteliti, permasalahan ini bisa didapatkan dari pengamatan ataupun studi literatur dari penelitian-penelitian yang telah dilakukan. Permasalahan yang didapat pada penelitian ini berupa identifikasi penyakit pada daun tanaman kentang, juga permasalahan dari data yang tidak seimbang.

b. Studi Literatur

Studi literatur merupakan kegiatan menelaah hasil dari penelitian terdahulu, pada tahap ini peneliti melakukan studi literatur dari jurnal penelitian dan juga beberapa sumber buku yang relevan pada permasalahan yang ditemukan pada tahap identifikasi permasalahan. Fungsi dari studi literatur pada

penelitian terdahulu juga untuk mendapatkan referensi tentang metode dan algoritma apa yang cocok dan dapat dikembangkan pada penelitian ini.

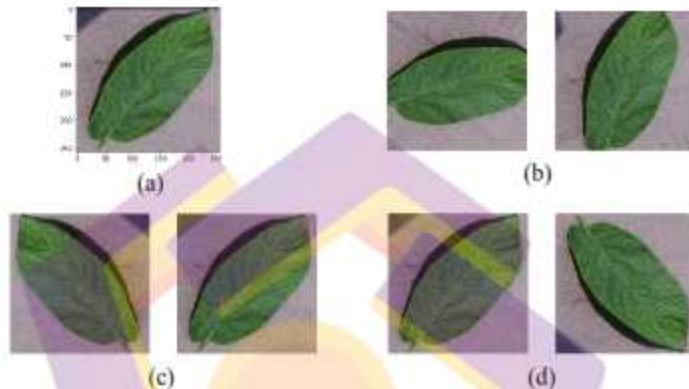
c. Penentuan Algoritma dan Pengumpulan data

Tahap ini merupakan penentuan algoritma yang akan digunakan, penentuan algoritma ini didasarkan pada tahap sebelumnya yaitu studi literatur. Pada penelitian ini algoritma yang digunakan untuk melakukan identifikasi penyakit pada daun kentang adalah Convolutional Neural Network. Sembari menentukan algoritma yang digunakan peneliti juga mengumpulkan dataset atau kumpulan data yang bisa digunakan untuk penelitian, pada penelitian ini mendapatkan data dari website penyedia dataset publik yaitu www.kaggle.com. Data yang didapatkan ternyata memiliki persebaran jumlah data yang tidak seimbang, maka peneliti menentukan metode augmentasi data guna menyeimbangkan persebaran jumlah data pada tiap kelasnya.

d. Pre-processing

Preprocessing merupakan proses untuk menghindari data yang digunakan menghasilkan performa model klasifikasi yang buruk, karena dengan banyaknya data yang digunakan data menjadi memiliki kemungkinan inkonsisten (Arhami and Nasir, 2020). Pada tahap ini peneliti melakukan persiapan untuk data yang sudah diseimbangkan dengan teknik augmentasi, persiapan yang dilakukan adalah dengan merubah ukuran citra yang sebelumnya 250x250 piksel diubah ke ukuran 150x150 piksel untuk mempercepat proses pelatihan. Pada tahap ini data yang dimiliki akan dibagi kedalam *data training*, *data validation*, dan *data testing*. Selanjutnya adalah

proses augmentasi data untuk menyeimbangkan persebaran jumlah data pada tiap kelas. Pada tahap ini augmentasi data akan diterapkan untuk daun sehat supaya jumlah data nya seimbang dengan data daun penyakit *early blight* dan *late blight* yang memiliki total data lebih banyak.



Gambar 3.2 *Data augmentation* (a) citra awal (b) *rotation* (c) *horizontal flip* (d) *vertical flip*

Augmentasi yang dilakukan pada data daun sehat ini adalah *rotation*, *horizontal flip*, dan *vertical flip*. 3 metode augmentasi data ini dipilih untuk menyeimbangkan data daun sehat karena metode ini hanya merubah sudut pandang dari citra. Gambar 3.2 adalah contoh hasil dari augmentasi pada citra daun kentang, Gambar 3.2 (a) adalah citra awal atau citra asli sebelum di augmentasi, Gambar 3.2 (b) adalah hasil dari augmentasi dengan metode *rotation*, Gambar 3.2 (c) adalah hasil dari augmentasi dengan metode *horizontal flip*, dan yang terakhir adalah augmentasi *vertical flip* yang ditunjukkan pada Gambar 3.2 (d).

e. Penentuan Skenario Percobaan

Penentuan skenario percobaan pada penelitian ini berguna untuk menganalisis perbandingan dari hasil percobaan. Pada penelitian ini skenario yang digunakan adalah skenario percobaan klasifikasi penyakit daun kentang dengan menggunakan dataset yang tidak seimbang, kemudian skenario selanjutnya adalah melakukan klasifikasi dengan dataset yang telah diseimbangkan. Skenario ini untuk menganalisis dampak dari penggunaan dataset yang memiliki persebaran jumlah data pada tiap kelasnya tidak seimbang. Skenario selanjutnya adalah percobaan dengan menggunakan beberapa metode augmentasi data untuk meningkatkan nilai akurasi dari model klasifikasi. Augmentasi data yang digunakan pada penelitian ini adalah random zoom, random shear, random brightness, random rotation, horizontal flip dan vertical flip. Dari beberapa augmentasi data yang digunakan akan diterapkan pada skenario yang berbeda untuk mendapatkan metode augmentasi data yang cocok digunakan pada klasifikasi penyakit daun kentang. Maka total dari skenario berjumlah 8 skenario percobaan. Semua skenario percobaan yang akan dilakukan masing-masing diujikan dengan 2 kali percobaan menggunakan arsitektur CNN yang berbeda yaitu arsitektur CNN yang diusulkan dan arsitektur VGG-16.

f. Pelatihan data

Tahap selanjutnya adalah tahap pelatihan data yang telah disiapkan pada penelitian ini. Pelatihan data yang dilakukan adalah menggunakan beberapa skenario yang telah ditentukan dan untuk menentukan skenario yang sesuai

dengan hasil terbaik maka dilakukan beberapa percobaan atau *trial and error*. Pada kerangka alur penelitian yang ditunjukkan Gambar 3.1 skenario yang akan dilakukan berjumlah 8 skenario yang berarti pada penelitian ini akan melakukan pelatihan data sebanyak 8 model klasifikasi. Pelati

g. Evaluasi Hasil

Model rancangan *Convolutional Neural Network (CNN)* yang diusulkan untuk identifikasi penyakit pada daun kentang ini akan dilakukan pengujian untuk menentukan model pada skenario yang mana yang memiliki hasil terbaik. Metode pengujian yang digunakan pada penelitian ini adalah metode *confusion matrix* untuk mendapatkan nilai akurasi, *recall* dan *precision* yang terbaik. Penggunaan nilai *precision* dan *recall* pada model klasifikasi adalah untuk melihat apakah pengaruh yang dihasilkan dari penggunaan data tidak seimbang.

h. Kesimpulan

Tahap akhir dari alur penelitian ini adalah pengambilan kesimpulan setelah dilakukan beberapa tahapan dimulai dari identifikasi permasalahan hingga evaluasi hasil dari beberapa skenario percobaan yang ditentukan. Tahap ini akan memaparkan temuan-temuan yang dihasilkan dari skenario percobaan, seperti dampak dari penggunaan data tidak seimbang, kemudian metode augmentasi data seperti apa yang cocok untuk diterapkan pada identifikasi penyakit daun kentang sehingga dapat meningkatkan nilai akurasi, *precision* dan *recall*.

BAB IV

HASIL PENELITIAN DAN PEMBAHASAN

Bab ini merupakan penjelasan tentang hasil dari penelitian yang dilakukan sesuai dengan kerangka alur penelitian pada Gambar 3.1. Penelitian ini diselesaikan dengan beberapa tahapan dimulai dari identifikasi permasalahan kemudian pengumpulan data yang akan digunakan untuk model klasifikasi, kemudian analisis data, membuat beberapa skenario percobaan, dan yang terakhir melakukan evaluasi dengan menguji skenario percobaan dari model klasifikasi untuk mendapatkan hasil yang terbaik.

4.1. Pengumpulan Data

Metode pengumpulan data pada penelitian ini adalah dengan melakukan pencarian data pada website penyedia data publik yang bisa diakses oleh semua orang. Data yang digunakan pada penelitian ini didapat pada website www.kaggle.com, lebih tepatnya kumpulan data yang diberi nama Plant Village.

Tabel 4.1 Detail dataset daun kentang

Sampel Data	Jumlah Data	Sumber Data
Daun Sehat	150	Kaggle (Plant Village)
Daun Penyakit <i>Early Blight</i>	1000	
Daun Penyakit <i>Late Blight</i>	1000	
Total	2150	

Data plant village ini berfokus untuk menyediakan data citra penyakit daun dari beberapa tanaman yang sering dikonsumsi manusia seperti daun kentang, daun tomat, daun ubi dan lain sebagainya. Data yang dipilih untuk penelitian ini adalah data penyakit pada daun kentang karena penelitian ini bertujuan untuk menganalisis fungsi model klasifikasi pada daun kentang. Detail dari data yang digunakan pada penelitian ini ditampilkan pada Tabel 4.1, total dari keseluruhan data yang digunakan adalah 2150 data citra. Persebaran data citra daun kentang ini dibagi kedalam 3 kelas dengan perincian data daun sehat berjumlah 150 buah data, data dengan penyakit *early blight* berjumlah 1000 buah data dan yang terakhir adalah data daun dengan penyakit *late blight* berjumlah 1000 data.

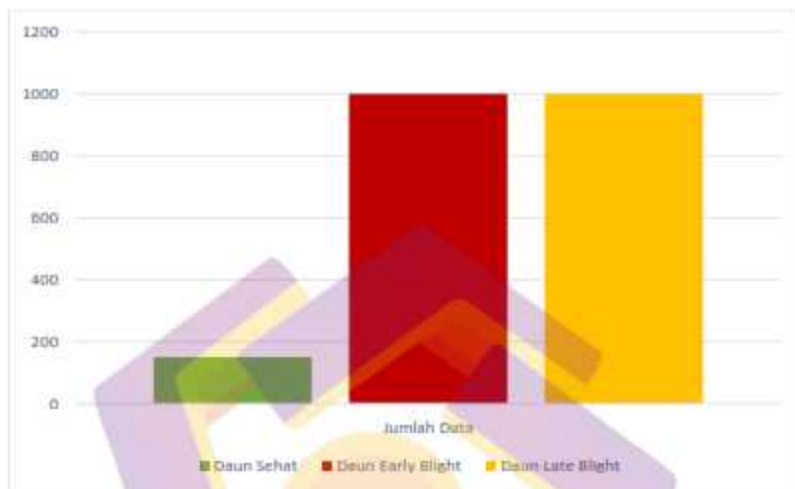
4.2. Analisis Data

Data yang telah didapatkan akan dilakukan analisis datanya untuk siap digunakan pada model klasifikasi. Tabel 4.1 menunjukkan persebaran data daun kentang yang digunakan pada penelitian ini sebagai data yang tidak seimbang pada persebaran jumlah antar kelasnya. Tahap ini akan melakukan beberapa perlakuan untuk data guna membuat data yang digunakan menjadi cukup ideal untuk sebuah model klasifikasi.

4.2.1 Persebaran data

Data yang digunakan pada penelitian ini memiliki persebaran data yang tidak seimbang, dapat dilihat pada Gambar 4.1 adalah grafik persebaran data daun kentang yang digunakan. Efek dari data yang tidak seimbang ini dapat

menimbulkan hasil model klasifikasi yang kurang baik karena model persebaran datanya tidak ideal.



Gambar 4.1 Persebaran jumlah data daun kentang

Persebaran data pada daun sehat memiliki jumlah data yang paling sedikit diantara kedua data lainnya, jumlah data daun sehat tidak sampai 20% dari jumlah data kelas *early blight* ataupun data *late blight*. Permasalahan data tidak seimbang ini dapat diatasi dengan menggunakan teknik augmentasi data pada data daun sehat untuk diseimbangkan dengan data yang lainnya.

4.2.2 Augmentasi data untuk menyeimbangkan data

Augmentasi data adalah teknik untuk menambahkan jumlah citra yang digunakan dengan cara merubah atau memodifikasi citra. Data daun sehat yang digunakan pada penelitian ini memiliki jumlah data yang paling sedikit, maka peneliti melakukan augmentasi data untuk menambahkan jumlah datanya. Teknik

augmentasi data yang digunakan untuk menambah jumlah data daun sehat ini adalah *random rotation*, *horizontal flip* dan *vertical flip*. Proses augmentasi data ini dilakukan pada google colab dengan menggunakan library keras, berikut adalah potongan dari kode augmentasi data pada daun kentang yang sehat untuk menambahkan datanya.

```

from keras.preprocessing.image import ImageDataGenerator
import zipfile, os

local_zip = 'DaunSehat.zip'
zip_ref = zipfile.ZipFile(local_zip, 'r')
zip_ref.extractall()
zip_ref.close()

data = './DaunSehat/'

save_to = './DaunSehat/'
target_size = (256,256)
batch_size = 32

train_datagen = ImageDataGenerator(
    rescale = 1./255,
    rotation_range = 20,
    horizontal_flip = True,
    vertical_flip=True
)

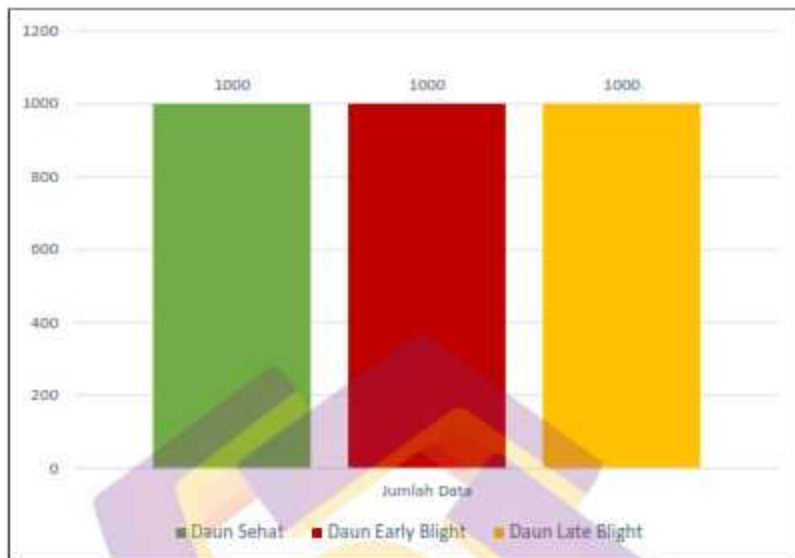
train_generator = train_datagen.flow_from_directory(
    data,
    target_size = target_size,
    batch_size = batch_size,
    save_to_dir = save_to,
)

for i in range(0,27):
    x,y = next(train_generator)

#untuk merubah folder ke bentuk ZIP
!zip -r /content/DaunSehatAugmented.zip /content/DaunSehat/

```

Pada potongan kode dapat kita lihat pada bagian awal melakukan import library preprocessing image berupa image data generator, kemudian baris kedua adalah import library zipfile untuk dapat mengolah file dalam bentuk zip pada google colab. Selanjutnya data daun sehat yang memiliki jumlah 150 data diupload ke dalam google colab dalam bentuk file zip, yang nantinya setelah di upload dilakukan ekstraksi dari bentuk zip supaya menjadi bentuk directory untuk dapat diakses file citra yang ada didalamnya. Kemudian akan dilakukan augmentasi data berupa *horizontal flip*, *vertical flip* dan *random rotation* dengan sudut 20° pada data daun sehat yang telah diupload. Target size adalah ukuran dari citra yang digunakan, pada tahap ini peneliti masih menggunakan ukuran asli dari citra yang didapatkan yaitu 256×256 piksel. Batch size merupakan pembagian jumlah data supaya ketika dilakukan perulangan dengan augmentasi data tidak memakan waktu lama, pada penelitian ini menggunakan 32 batch size sehingga pada proses augmentasi datanya akan dikelompokkan setiap kloter nya berjumlah 32 citra. Selanjutnya untuk perulangan dalam melakukan augmentasi nya dilakukan sebanyak 27 kali, karena dengan nilai 27 dapat menghasilkan jumlah citra yang mendekati angka 1000. Maka perhitungannya adalah 27 perulangan dikalikan dengan 32 per batch nya, $27 \times 32 = 864$. Jumlah ini adalah citra yang di augmentasi, sedangkan jumlah citra yang asli adalah 150. Jadi total keseluruhan citra daun sehat sekarang adalah $864 + 150 = 1014$ data, supaya nilainya sama dengan kelas daun yang lainnya maka yang diambil adalah 1000 data.



Gambar 4.2 Persebaran jumlah data setelah di augmentasi

Data daun sehat yang telah diberikan perlakuan augmentasi menjadi bisa seimbang dengan data yang lainnya. Gambar 4.2 merupakan grafik dari data daun kentang setelah dilakukan augmentasi pada daun sehat, jumlah dari persebaran data yang akan digunakan sekarang menjadi lebih ideal daripada sebelum dilakukan augmentasi.

4.2.3 Preprocessing data

Preprocessing data merupakan proses untuk menghindari data yang digunakan menjadi tidak ideal untuk digunakan pada model klasifikasi. Pada tahap augmentasi data terhadap daun sehat yang memiliki jumlah data paling sedikit juga termasuk dari preprocessing data, karena pada tahap augmentasi data daun sehat

yang telah digunakan berfungsi untuk menyeimbangkan persebaran jumlah data dari semua kelas yang digunakan untuk model klasifikasi.

Tahap setelah data yang digunakan memiliki persebaran data yang seimbang atau bisa dibilang telah ideal, maka akan dilakukan pembagian data untuk digunakan pada model klasifikasi. Pembagian data ini adalah melakukan pembagian data menjadi data training, data validation dan data testing. Data training adalah data yang digunakan untuk melatih model klasifikasi, kemudian data validation merupakan data yang digunakan untuk melihat apakah hasil pelatihan model klasifikasi menggunakan data training memiliki kendala overfitting atau model klasifikasinya sudah baik. Terakhir adalah data testing yang berarti adalah data uji, data ini digunakan untuk menguji model klasifikasi yang ditraining dan nantinya dapat digunakan untuk evaluasi model klasifikasi yang digunakan. Pembagian data ke dalam data training, data validation dan data. Pembagiannya adalah data training memiliki jumlah 80% data, kemudian data validation 10% dan yang terakhir data testing 10%.



Gambar 4.3 Pembagian data 80%:10%:10%

Gambar 4.3 adalah pembagian data dengan proporsi perbandingannya adalah 80%:10%:10%. Data training dengan 80% berjumlah 800 data, kemudian data validation dengan 10% berjumlah 100 data dan selanjutnya data testing 10% berjumlah 100 data.

Tahap preprocessing sebelum data digunakan untuk model klasifikasi adalah merubah ukuran citra data yang digunakan menjadi ukuran 150x150. Fungsi dari merubah ukuran citra ini adalah supaya seluruh citra data yang digunakan untuk model klasifikasi memiliki ukuran yang seragam, selain itu juga berfungsi untuk mempercepat proses klasifikasi apabila sebelumnya citra data yang digunakan memiliki ukuran yang besar.

4.2.4 Augmentasi data lanjutan

Augmentasi data merupakan teknik untuk menambah jumlah data seperti yang dijelaskan pada sub bab sebelumnya, tahap sebelumnya augmentasi data

digunakan untuk menyeimbangkan data pada daun kentang supaya memiliki persebaran jumlah data yang seimbang. Augmentasi data juga memiliki fungsi untuk menambah keberagaman data sehingga data yang digunakan untuk model klasifikasi menjadi lebih kaya akan informasi. tahap ini augmentasi data tidak hanya digunakan pada data daun sehat saja, akan tetapi digunakan pada data di seluruh kelas.

Augmentasi data tambahan yang dilakukan akan menggunakan tiga teknik yang berbeda yaitu augmentasi data *random zoom*, *random shear* dan *random brightness*. Penggunaan tiga teknik augmentasi data ini karena dirasa hasil dari model klasifikasi yang digunakan masih dapat ditingkatkan untuk menjadi lebih baik lagi. Analisis dari hasil model klasifikasi dengan menggunakan augmentasi data tambahan ini nantinya akan dilakukan perbandingan untuk dapat dilihat pengaruh dari augmentasi data yang diterapkan.

4.3. Model Klasifikasi Convolutional Neural Network

Penelitian ini menggunakan metode Convolutional Neural Network (CNN) dalam melakukan klasifikasi penyakit daun kentang. CNN memiliki beberapa lapisan arsitektur yang dapat disesuaikan untuk melakukan klasifikasi. Kelebihan yang dimiliki oleh CNN adalah lapisan convolutional nya yang dapat digunakan untuk ekstraksi ciri atau ekstraksi tekstur dari citra secara mendalam. Lapisan yang tidak kalah berguna adalah lapisan *pooling* yang digunakan untuk mengurangi dimensi yang dimiliki oleh citra dengan tetap mempertahankan informasi penting yang dimiliki citra.

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 150, 150, 16)	448
max_pooling2d (MaxPooling2D)	(None, 75, 75, 16)	0
conv2d_1 (Conv2D)	(None, 75, 75, 32)	4640
max_pooling2d_1 (MaxPooling2D)	(None, 37, 37, 32)	0
conv2d_2 (Conv2D)	(None, 37, 37, 164)	47296
max_pooling2d_2 (MaxPooling2D)	(None, 18, 18, 164)	0
conv2d_3 (Conv2D)	(None, 18, 18, 128)	189056
max_pooling2d_3 (MaxPooling2D)	(None, 9, 9, 128)	0
flatten (Flatten)	(None, 10368)	0
dense (Dense)	(None, 5)	31107
Total params: 272,647		
Trainable params: 272,647		
Non-trainable params: 0		

Gambar 4.4 Model arsitektur CNN

Gambar 4.4 adalah model arsitektur convolutional neural network yang digunakan pada penelitian ini. Arsitektur ini memiliki 4 lapisan konvolusi untuk ekstraksi citra. Lapisan convolutional yang pertama dapat kita lihat pada gambar 4.5 Conv2D (150, 150, 16) yang berarti pada lapisan ini citra memiliki ukuran 150 x 150 piksel dan memiliki 16 jumlah *channel* atau filter yang digunakan untuk ekstraksi ciri atau tekstur dari citra. Ukuran citra 150x150 dikarenakan pada *preprocessing* dilakukan perubahan ukuran semua citra yang digunakan menjadi 150 x 150 piksel. Setelah proses konvolusi akan dilanjut proses *pooling*, pada Gambar 4.5 pada baris setelah Conv2D adalah max_pooling. Max pooling adalah salah satu dari pooling yang digunakan pada penelitian ini, seperti yang dijelaskan sebelumnya pooling merupakan proses untuk mereduksi atau mengurangi resolusi yang dimiliki oleh citra. Max pooling merupakan filter untuk mereduksi resolusi

citra dengan mengambil nilai piksel tertinggi dari filter yang melewati citra. Penelitian ini menggunakan 4 lapisan *max pooling*. Hasil dari tahap konvolusi dan *max pooling* dinamakan *feature map* dan ini adalah hasil ekstraksi ciri yang bisa digunakan untuk klasifikasi. Tahap selanjutnya adalah *flatten* yang merupakan proses merubah *feature map* hasil dari pooling menjadi vektor, sehingga dapat dilakukan klasifikasi dengan neural network. Pada baris terakhir yang merupakan lapisan dense terdapat angka 3 yang menunjukkan jumlah kelas yang digunakan yaitu kelas daun sehat, daun berpenyakit *early blight* dan daun berpenyakit *late blight*.

4.4. Skenario Percobaan

Percobaan pada penelitian ini menggunakan beberapa skenario, fungsi dari penggunaan beberapa skenario ini adalah untuk dapat dilakukan analisis hasil dari percobaan yang dilakukan. Hasil analisis dari beberapa skenario percobaan yang dilakukan dapat dijadikan acuan untuk menentukan model klasifikasi yang terbaik. Beberapa skenario percobaan yang digunakan pada penelitian ini ditampilkan pada tabel 4.2.

Tabel 4.2 Skenario Percobaan

No	Skenario	Arsitektur	Augmentasi Data	
			Penyeimbangan Data	Augmentasi Tambahan
1	S1	CNN 16,32,64,128	No	No
2	S2	CNN 16,32,64,128	Yes	No

Tabel 4.2 Lanjutan

No	Skenario	Arsitektur	Augmentasi Data	
			Penyeimbangan Data	Augmentasi Tambahan
3	S3	CNN 16,32,64,128	Yes	Random Zoom
4	S4	CNN 16,32,64,128	Yes	Random Shear
5	S5	CNN 16,32,64,128	Yes	Random Brightness
6	S6	CNN 16,32,64,128	Yes	Rotation
7	S7	CNN 16,32,64,128	Yes	Horizontal flip
8	S8	CNN 16,32,64,128	Yes	Vertical flip

Tabel 4.2 adalah beberapa skenario yang digunakan pada penelitian ini untuk mencari model klasifikasi yang terbaik dalam melakukan identifikasi penyakit pada daun kentang. Skenario percobaan ini akan diujikan dengan pembagian data 80% data training, 10% data validation dan 10% data testing. Penjelasan detail dari beberapa skenario adalah sebagai berikut:

1. Skenario percobaan 1 (S1)

Pada skenario percobaan yang pertama dilakukan adalah menggunakan data daun kentang yang belum diseimbangkan, data ini berjumlah 2150 data dengan rincian data 150 data daun sehat, 1000 data daun berpenyakit *early blight* dan 1000 data daun berpenyakit *late blight*. Percobaan ini menggunakan model arsitektur CNN yang ditunjukkan pada Gambar 4.4, arsitektur CNN yang digunakan berisi 4 lapisan konvolusi dan 4 lapisan max pooling. Percobaan ini dilakukan untuk menganalisis apa dampak dari

penggunaan dataset dengan persebaran jumlah yang tidak seimbang pada masing-masing kelasnya.

2. Skenario percobaan 2

Skenario kedua adalah melakukan percobaan dengan data yang diseimbangkan, skenario ini nantinya dapat dibandingkan dengan skenario pertama untuk dianalisis perbedaan dari data yang tidak seimbang dan data yang seimbang ketika digunakan untuk model klasifikasi. Data daun kentang yang digunakan memiliki rincian 1000 data daun sehat, 1000 data daun berpenyakit *early blight* dan 1000 data daun berpenyakit *late blight*. Arsitektur CNN yang digunakan pada skenario ini menggunakan arsitektur yang sama seperti skenario 1 yaitu 4 lapisan konvolusi dan 4 lapisan *max pooling*.

3. Skenario percobaan 3

Skenario ketiga pada penelitian ini adalah untuk mengetahui apakah model klasifikasi dengan data yang sudah diseimbangkan masih dapat ditingkatkan lagi hasilnya, bisa pada tingkat akurasi, presisi ataupun recall nya. Skenario ketiga ini menggunakan data yang sudah diseimbangkan dengan rincian 1000 data pada daun sehat, 1000 data pada daun berpenyakit *early blight* dan 1000 data daun berpenyakit *late blight*. Arsitektur CNN yang digunakan pada skenario ini menggunakan arsitektur yang sama seperti skenario 1 yaitu 4 lapisan konvolusi dan 4 lapisan *max pooling*. Data yang digunakan akan di augmentasi dengan memanfaatkan teknik augmentasi data *random zoom* pada semua datanya baik dari data training, validation dan testing.

4. Skenario percobaan 4

Skenario keempat pada penelitian ini memiliki fungsi percobaan yang sama dengan skenario ketiga yaitu untuk mengetahui apakah data yang sudah diseimbangkan masih dapat ditingkatkan lagi hasilnya, baik pada tingkat akurasi, presisi ataupun recall nya. Skenario keempat ini menggunakan data yang sudah diseimbangkan dengan rincian 1000 data pada daun sehat, 1000 data pada daun berpenyakit *early blight* dan 1000 data daun berpenyakit *late blight*. Arsitektur CNN yang digunakan pada skenario ini menggunakan arsitektur yang sama seperti skenario 1 yaitu 4 lapisan konvolusi dan 4 lapisan *max pooling*. Data yang digunakan akan di augmentasi dengan memanfaatkan teknik augmentasi data *random shear* pada semua datanya baik dari data *training*, *validation* dan *testing*.

5. Skenario percobaan 5

Skenario kelima memiliki fungsi yang sama pada percobaan skenario ketiga dan keempat yaitu untuk mengetahui apakah data yang sudah diseimbangkan masih dapat ditingkatkan lagi hasilnya, baik pada tingkat akurasi, presisi ataupun recall nya. Hanya saja jenis augmentasi datanya saja yang berbeda, hal ini untuk mengetahui apakah ada perbedaan yang signifikan dari setiap teknik augmentasi yang digunakan. Data augmentasi yang digunakan pada skenario percobaan kelima ini adalah *random brightness* yang diterapkan pada data *training*, data *validation* dan data *testing*. Data yang digunakan pada skenario kelima ini adalah data yang sudah diseimbangkan pada tiap kelasnya yaitu 1000 data daun sehat, 1000

data daun berpenyakit *early blight* dan 1000 data daun berpenyakit *late blight*.

6. Skenario Percobaan 6

Skenario percobaan keenam melakukan uji coba penggunaan augmentasi data yang sebelumnya sudah digunakan untuk menyeimbangkan data, karena pada proses penyeimbangan data teknik augmentasi hanya diberikan kepada kelas daun sehat saja, maka akan dilakukan uji coba dengan menggunakan augmentasi terhadap seluruh kelas data yaitu data daun sehat, data daun berpenyakit *early blight* dan data daun berpenyakit *late blight*. Augmentasi yang digunakan pada skenario ini adalah *random rotation*.

7. Skenario Percobaan 7

Skenario ketujuh juga melakukan uji coba penggunaan augmentasi data yang sebelumnya sudah digunakan untuk menyeimbangkan data. Uji coba pada skenario ini menggunakan augmentasi data *horizontal flip* yang diterapkan terhadap seluruh kelas data yaitu data daun sehat, data daun berpenyakit *early blight* dan data daun berpenyakit *late blight*.

8. Skenario Percobaan 8

Skenario kedelapan adalah percobaan terakhir dari pemanfaatan augmentasi data pada arsitektur Convolutional Neural Network (CNN). Percobaan ini menggunakan teknik augmentasi data *vertical flip* yang diterapkan terhadap seluruh kelas data daun tanaman kentang yang digunakan.

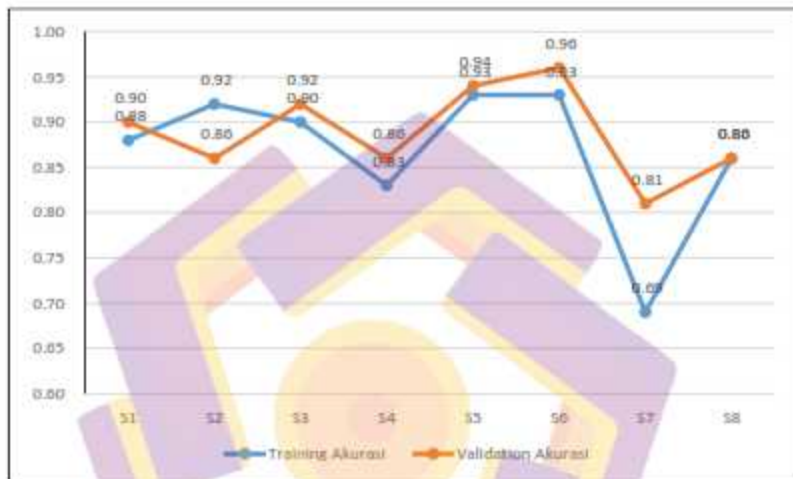
Setiap skenario yang akan digunakan yaitu skenario 1 sampai dengan skenario 8 akan diujikan dalam pembelajaran data dengan menggunakan 30 *epochs*. Maksud *epoch* disini adalah ketika seluruh data telah melalui proses pembelajaran data pada *neural network*. Setiap skenario yang digunakan akan diujikan menggunakan rasio pembagian 80% data *training*, 10% data *validation* dan 10% data *testing*. Penggunaan augmentasi data dari skenario ketiga sampai dengan kedelapan akan dilakukan analisis pengaruh yang dihasilkan pada model klasifikasi.

4.5. Hasil dan Evaluasi Model Klasifikasi

Proses klasifikasi pada daun tanaman kentang dilakukan dengan menggunakan skenario-skenario percobaan yang telah ditentukan pada penjelasan sebelumnya. Hasil dari percobaan yang telah dilakukan akan dijelaskan secara rinci dan dianalisis apa pengaruh dari penggunaan data yang persebaran jumlahnya tidak seimbang dengan yang seimbang. Metode pengujian yang digunakan pada penelitian ini adalah *Confusion Matrix*, dengan metode ini tidak hanya menunjukkan akurasi dari model klasifikasi saja akan tetapi memberikan nilai *precision* dan *recall* yang bisa digunakan untuk menganalisa efek dari keseimbangan persebaran data pada tiap kelasnya.

4.5.1 Hasil pelatihan model klasifikasi

Model klasifikasi untuk identifikasi penyakit pada daun tanaman kentang telah dilatih pada 8 skenario, dalam penelitian ini hasil dari pelatihan pada data *training* dan data validasi pada masing-masing skenario dianalisis.



Gambar 4.5 Akurasi pada data *training* dan *validation*

Grafik perbandingan dari hasil percobaan pada skenario 1 sampai dengan 8 ditampilkan pada Gambar 4.5. Sejah ini hasil akurasi pada data training dan data validation dari tiap skenario tidak ada yang memiliki perbedaan cukup tinggi kecuali pada skenario 7 dengan hasil yang paling rendah, hal ini menandakan bahwa model klasifikasi dengan arsitektur Convolutional Neural Network (CNN) yang digunakan cukup bagus untuk identifikasi penyakit pada daun tanaman kentang.

Hasil akurasi pada skenario 1 dengan menggunakan data yang belum diseimbangkan terlihat normal dan mendapatkan akurasi yang bagus sebesar 88% pada data training dan 90% pada data validation. Sedangkan pada skenario 7 tingkat

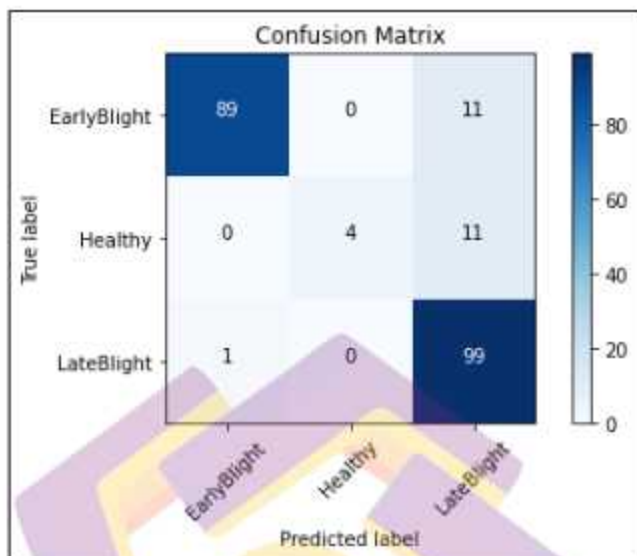
akurasi yang didapatkan menurun dan merupakan akurasi yang terkecil dibandingkan skenario yang lainnya, walaupun akurasi yang didapatkan masih bisa dikategorikan akurasi yang baik. Selanjutnya akan dilakukan pengujian model klasifikasi dari skenario 1 hingga delapan dengan menggunakan data *testing*, kemudian metode analisis untuk menghitung hasilnya menggunakan Confusion Matrix.

4.5.2 Pengujian confusion matrix

Confusion matrix adalah metode pengujian untuk mengukur kinerja dari sebuah model klasifikasi. Metode pengujian ini tersusun dari matrix hasil prediksi model klasifikasi dengan data aktual atau data yang sesungguhnya.

4.5.2.1 Pengujian pada skenario 1

Skenario pertama adalah percobaan klasifikasi penyakit daun tanaman kentang dengan menggunakan data asli yaitu data yang tidak seimbang dengan jumlah 2150 data dengan rincian 150 data daun sehat, 1000 data daun berpenyakit *early blight* dan 1000 data daun berpenyakit *late blight*.



Gambar 4.6 Confusion matrix skenario 1

Gambar 4.6 adalah representasi hasil pengujian confusion matrix dari skenario 1, pengujian ini menggunakan aplikasi web yaitu google colab. Pada *confusion matrix* terdapat empat istilah yang digunakan untuk menghitung nilai akurasi, precision dan recall yaitu True Positive (TP), False Positive (FP), True Negative (TN) dan False Negative (FN). Klasifikasi penyakit pada daun kentang memiliki 3 kelas, maka pengujian yang digunakan adalah confusion matrix multi class dimana untuk mendapatkan nilai precision dan recall dihitung pada masing-masing kelas.

1. Akurasi

Nilai akurasi didapatkan dari keseluruhan jumlah True Positive dari masing-masing kelas. Pada Gambar 4.6 jumlah nilai TP adalah $89 + 4 + 99 = 192$, kemudian jumlah data *testing* secara keseluruhan adalah 215 data. Nilai

akurasi adalah total TP dibagi total data kemudian dikalikan 100%, maka
 $(192 / 215) * 100 = 89\%$

2. Precision

Precision didapatkan dengan menghitung nilai TP dibagi dengan TP + FP, namun untuk kasus pada confusion matrix multi class dalam menentukan nilai precision dihitung pada masing-masing kelas.

a. Nilai precision kelas *early blight*

$$\text{Precision} = 89 / (89 + 0 + 1) = 89 / 90 = 0.99 = 99\%$$

b. Nilai Precision kelas daun sehat

$$\text{Precision} = 4 / (4 + 0 + 0) = 4 / 4 = 1 = 100\%$$

c. Nilai Precision kelas *late blight*

$$\text{Precision} = 99 / (99 + 11 + 11) = 99 / 121 = 0.82 = 82\%$$

Hasil precision secara keseluruhan didapatkan dengan menjumlahkan semua nilai precision dari masing-masing kelas yang kemudian dibagi dengan jumlah kelas. Sehingga precision dari skenario 1 adalah $(0.99 + 1 + 0.82) / 3 = 2.81 / 3 = 0.94 = 94\%$

3. Recall

Recall didapatkan dengan menghitung nilai TP dibagi dengan TP + FN, namun untuk kasus pada confusion matrix multi class dalam menentukan nilai recall dihitung pada masing-masing kelas.

a. Nilai recall kelas *early blight*

$$\text{Recall} = 89 / (89 + 0 + 11) = 89 / 100 = 0.89 = 89\%$$

- b. Nilai recall kelas daun sehat

$$\text{Recall} = 4 / (4 + 0 + 11) = 4 / 15 = 0.27 = 27\%$$

- c. Nilai recall kelas *late blight*

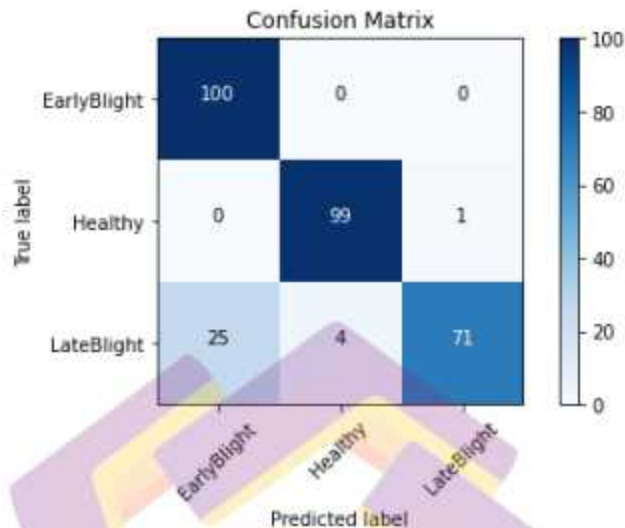
$$\text{Recall} = 99 / (99 + 0 + 1) = 99 / 100 = 0.99 = 99\%$$

Hasil recall secara keseluruhan didapatkan dengan menjumlahkan semua nilai recall dari masing-masing kelas yang kemudian dibagi dengan jumlah kelas. Sehingga recall dari skenario 1 adalah $(0.89 + 0.27 + 0.99) / 3 = 2.15 / 3 = 0.72 = 72\%$

4.5.2.2 Pengujian pada skenario 2

Skenario kedua adalah percobaan klasifikasi penyakit daun tanaman kentang dengan menggunakan data yang sudah di seimbangkan persebaran jumlah data pada masing-masing kelasnya. Pada hasil percobaan ini akan dianalisis perbedaan yang terjadi ketika menggunakan data yang seimbang dan data yang tidak seimbang.

Data yang sudah di seimbangkan yang digunakan pada skenario ini berjumlah 300 data pada 3 kelas, rincian datanya adalah 1000 data daun sehat, 1000 data daun berpenyakit *early blight* dan 1000 data daun berpenyakit *late blight*.



Gambar 4.7 Confusion matrix skenario 2

Gambar 4.7 adalah representasi hasil dari pengujian confusion matrix pada skenario 2. Data yang digunakan untuk proses uji ini sejumlah 10% dari data yang digunakan. Dapat dilihat perbedaan dari skenario 1 dan 2 ini adalah jumlah dari data daun sehat yang digunakan. Nilai akurasi, precision dan recall pada skenario 2 akan dihitung guna menjadi perbandingan terhadap percobaan pada skenario 1.

1. Akurasi

Nilai akurasi didapatkan dari keseluruhan jumlah True Positive dari masing-masing kelas. Jumlah nilai TP adalah $100 + 99 + 71 = 270$, kemudian jumlah data *testing* secara keseluruhan adalah 300 data. Nilai akurasi adalah total TP dibagi total data kemudian dikalikan 100%, maka $(270 / 300) * 100 = 0.9 * 100 = 90\%$

2. Precision

Precision didapatkan dengan menghitung nilai TP dibagi dengan TP + FP, seperti pada skenario 1 untuk kasus pada confusion matrix multi class dalam menentukan nilai precision dihitung pada masing-masing kelas.

a. Nilai precision kelas *early blight*

$$\text{Precision} = 100 / (100 + 0 + 25) = 100 / 125 = 0.80 = 88\%$$

b. Nilai Precision kelas daun sehat

$$\text{Precision} = 99 / (0 + 99 + 4) = 99 / 103 = 0.96 = 96\%$$

c. Nilai Precision kelas *late blight*

$$\text{Precision} = 71 / (0 + 1 + 71) = 71 / 72 = 0.99 = 99\%$$

Hasil precision secara keseluruhan didapatkan dengan menjumlahkan semua nilai precision dari masing-masing kelas yang kemudian dibagi dengan jumlah kelas. Sehingga precision dari skenario 2 adalah $(0.80 + 0.96 + 0.99) / 3 = 2.75 / 3 = 0.92 = 92\%$

3. Recall

Recall didapatkan dengan menghitung nilai TP dibagi dengan TP + FN, namun untuk kasus pada confusion matrix multi class dalam menentukan nilai recall dihitung pada masing-masing kelas.

a. Nilai recall kelas *early blight*

$$\text{Recall} = 100 / (100 + 0 + 0) = 100 / 100 = 1 = 100\%$$

b. Nilai recall kelas daun sehat

$$\text{Recall} = 99 / (0 + 99 + 1) = 99 / 100 = 0.99 = 99\%$$

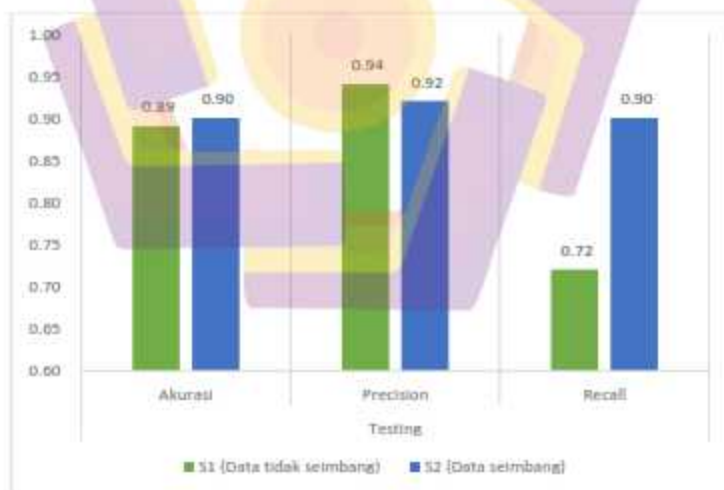
c. Nilai recall kelas *late blight*

$$\text{Recall} = 71 / (25 + 4 + 71) = 71 / 100 = 0.71 = 71\%$$

Hasil recall secara keseluruhan didapatkan dengan menjumlahkan semua nilai recall dari masing-masing kelas yang kemudian dibagi dengan jumlah kelas. Sehingga recall dari skenario 2 adalah $(1 + 0.99 + 0.71) / 3 = 2.7 / 3 = 0.90 = 90\%$

4.5.3 Pengaruh data seimbang dan tidak seimbang

Pengujian menggunakan *confusion matrix* diujikan pada model klasifikasi menggunakan data seimbang (skenario 2) dan data tidak seimbang (skenario 1). Pada tahap ini akan dibahas pengaruh yang dihasilkan oleh kedua skenario ini.

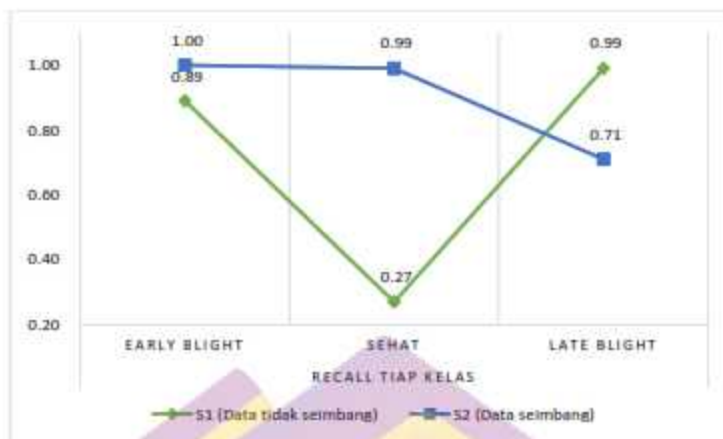


Gambar 4.8 Grafik perbandingan hasil dari skenario 1 dan 2

Gambar 4.8 adalah grafik perbandingan antara model klasifikasi yang menggunakan data tidak seimbang dan data yang seimbang. Grafik batang dengan

warna hijau adalah untuk skenario 1 atau percobaan dengan menggunakan data yang tidak seimbang, dan untuk grafik batang berwarna biru adalah skenario 2 atau percobaan dengan menggunakan data yang telah diseimbangkan. Grafik yang ditunjukkan jika diperhatikan menghasilkan perbandingan untuk nilai akurasi, precision dan recall. Akurasi adalah nilai tingkat kedekatan dari hasil prediksi terhadap keseluruhan dataset yang digunakan, akurasi dari skenario 1 dan 2 tidak memiliki perbedaan yang jauh karena selisih akurasi antara penggunaan data yang tidak seimbang dan data yang seimbang sebesar 1% dengan rincian 89% tingkat akurasi data tidak seimbang (skenario 1) dan akurasi sebesar 90% untuk penggunaan data seimbang (skenario 2).

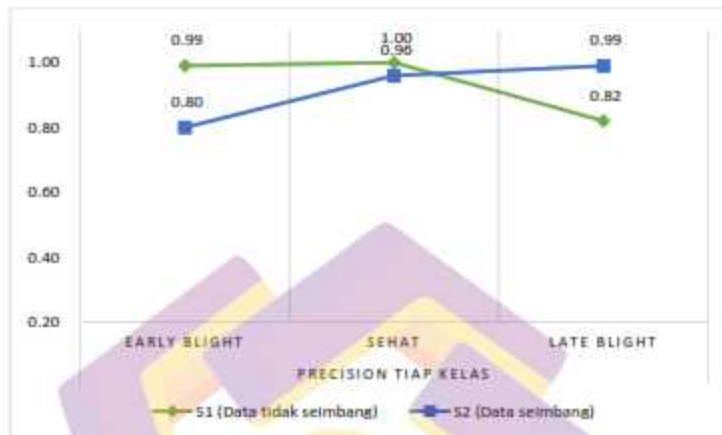
Hasil pengujian pada data testing untuk nilai precision menunjukkan bahwa skenario 1 menghasilkan nilai 94% yang mana lebih tinggi ketimbang nilai precision dari skenario 2 yaitu 92%. Namun pada nilai recall antara skenario 1 dan skenario 2 memiliki perbedaan yang cukup jauh yaitu 72% untuk skenario 1 dan 90% untuk skenario 2. Selanjutnya akan dijabarkan lebih lanjut nilai precision dan recall dari kedua skenario dan juga penyebab dari penurunan nilai recall pada skenario 1.



Gambar 4.9 Perbandingan nilai recall tiap kelas pada skenario 1 dan 2

Gambar 4.9 merupakan grafik perbandingan nilai recall dari penggunaan data yang tidak seimbang (skenario 1) dan data yang seimbang (skenario 2). Grafik garis berwarna hijau untuk skenario 1 yaitu penggunaan data tidak seimbang, kemudian grafik garis warna biru untuk skenario 2 yaitu penggunaan data yang sudah diseimbangkan. Skenario 1 atau penggunaan data tidak seimbang yang ditunjukkan pada Gambar 4.9 menghasilkan recall yang sangat kecil pada kelas daun sehat, recall sendiri adalah kemampuan suatu model klasifikasi untuk menemukan kembali data yang sesungguhnya. Jika kita kembali melihat Gambar 4.6 pada bagian kelas daun sehat jumlah data uji yang digunakan adalah 15 data, sedangkan data yang diprediksi benar (TP) adalah 4 data yang berarti 11 data lainnya diprediksi salah. Tingkat recall pada data daun sehat (skenario 1) adalah 27% dan ini sangat buruk. Jadi kerugian yang dialami jika menggunakan data tidak seimbang adalah kelas yang memiliki jumlah data paling sedikit atau kelas minoritas memiliki kemungkinan besar akan sulit ditemukan kembali data yang sebenarnya oleh

sistem, hal ini bisa disebabkan karena model klasifikasi akan cenderung kepada kelas yang memiliki data lebih banyak atau kelas mayoritas.



Gambar 4.10 Perbandingan nilai precision tiap kelas pada skenario 1 dan 2

Grafik yang ditampilkan pada Gambar 4.10 adalah perbandingan nilai precision untuk tiap kelas pada skenario 1 dan skenario 2. Pada grafik ini menunjukkan bahwa nilai precision untuk kelas daun sehat pada skenario 1 memiliki nilai yang lebih baik dari skenario 2, tetapi jika kita lihat perbedaan yang dimiliki tidak terlalu besar yaitu berjumlah 4%. Precision merupakan nilai dari ketepatan model klasifikasi dari hasil prediksi yang diberikan oleh sistem, sedangkan recall adalah nilai dari tingkat keberhasilan sebuah model klasifikasi menemukan data yang sesungguhnya. Dalam kasus pada skenario 1 pada kelas daun sehat meskipun nilai precision mendapatkan nilai 100% yang berarti ketepatan prediksi kelas daun sehat dari data yang diprediksi daun sehat sangat tinggi, akan tetapi nilai recall yang dimiliki sangatlah rendah yaitu 27% yang berarti

kemampuan model klasifikasi dalam memperoleh data daun sehat dari seluruh data testing yang digunakan sangatlah buruk.

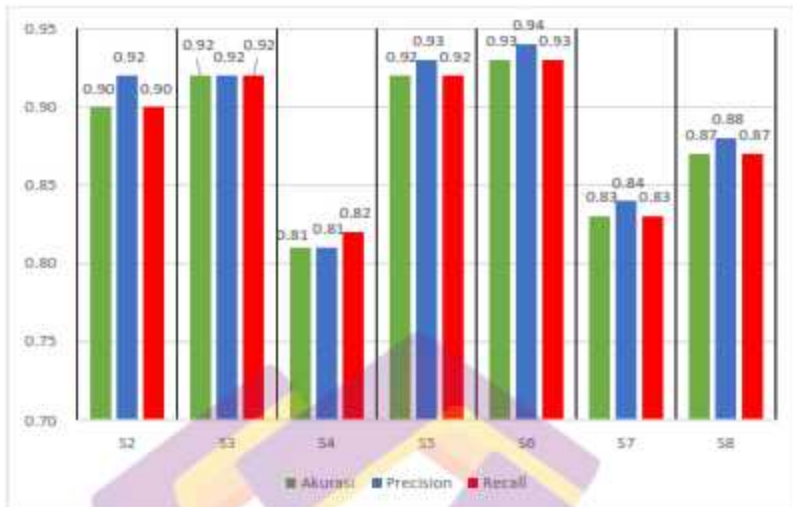
Nilai akurasi adalah nilai yang paling sering digunakan untuk mengukur performa model klasifikasi. Namun nilai akurasi tidak bisa dijadikan acuan utama jika data yang digunakan memiliki ketidak seimbangan antar kelasnya, karena nilai akurasi bisa saja tinggi hanya dengan didapat dari ketepatan memprediksi pada kelas yang memiliki data lebih banyak (Juba and Le, 2019). Nilai precision dan recall adalah nilai yang saling berhubungan kuat, meskipun nilai precision atau ketepatannya tinggi tetapi nilai recall atau kemampuan memperoleh datanya rendah akan berdampak buruk, begitu pula sebaliknya jika kemampuan model klasifikasi dalam memperoleh datanya tinggi tetapi nilai precision atau ketepatan dari seluruh data yang diperoleh oleh sistem rendah maka akan berakibat buruk juga. Selanjutnya jika kita perhatikan pada nilai presisi ataupun recall pada skenario 2 tidak memiliki penurunan nilai yang drastis seperti pada skenario 1 yaitu percobaan dengan menggunakan data yang tidak seimbang.

4.5.4 Analisis perbandingan augmentasi data

Pemanfaatan teknik augmentasi data untuk menyeimbangkan data pada tahap sebelumnya memberikan hasil yang baik yaitu untuk menghindari suatu model klasifikasi cenderung ke kelas yang memiliki data lebih banyak (mayoritas). Tahap ini akan melakukan uji coba pemanfaatan teknik augmentasi data untuk meningkatkan nilai akurasi, precision dan recall pada model klasifikasi. Rincian dari skenario percobaan pemanfaatan teknik augmentasi data untuk meningkatkan

nilai akurasi, precision dan recall telah ditunjukkan pada Tabel 4.2 pada bagian skenario 3 sampai dengan skenario 8. Uji coba yang dilakukan adalah untuk mengetahui teknik augmentasi data apa saja yang cocok untuk klasifikasi penyakit pada daun tanaman kentang, karena tidak semua teknik augmentasi cocok pada karakteristik dari citra yang diklasifikasi.

Percobaan penggunaan teknik augmentasi data diterapkan pada skenario 3 sampai dengan skenario 8, sedangkan skenario 2 adalah model klasifikasi tanpa menggunakan teknik augmentasi data pada seluruh kelas data yang digunakan. Skenario 3 adalah percobaan menggunakan teknik augmentasi data random zoom, selanjutnya skenario 4 adalah percobaan menggunakan teknik augmentasi data random shear, kemudian dilanjut pada percobaan skenario 5 dengan memanfaatkan teknik augmentasi data random brightness. Skenario 6 sampai dengan skenario 8 menggunakan teknik augmentasi data yang sebelumnya digunakan untuk menyeimbangkan data, akan tetapi pada penyeimbangan data teknik augmentasi hanya diterapkan pada kelas yang memiliki data paling sedikit yaitu pada kelas daun sehat. Skenario 6 menggunakan teknik augmentasi data random rotation, kemudian skenario 7 menggunakan teknik augmentasi data horizontal flip, dan yang terakhir skenario 8 menggunakan teknik augmentasi data vertical flip.



Gambar 4.11 Perbandingan hasil dari pemanfaatan augmentasi data

Perbandingan hasil dari beberapa penggunaan teknik augmentasi data ditunjukkan pada gambar 4.10. Grafik berwarna hijau menunjukkan tingkat akurasi dari model klasifikasi pada masing-masing skenario, grafik warna biru mewakili nilai dari precision tiap skenario, dan yang terakhir grafik warna merah adalah nilai dari recall. Skenario 2 (S2) yang ditunjukkan pada grafik Gambar 4.10 adalah hasil klasifikasi dengan menggunakan data yang telah diseimbangkan yaitu 1000 data daun sehat, 1000 data daun berpenyakit *early blight* dan 1000 data daun berpenyakit *early blight*. Skenario 2 menghasilkan tingkat akurasi 90% precision 92% dan recall 90%, dari hasil ini peneliti melakukan uji coba dengan menggunakan teknik augmentasi data pada data yang digunakan untuk dianalisis teknik augmentasi apa saja yang cocok pada citra daun kentang.

Percobaan yang telah dilakukan menghasilkan peningkatan pada nilai akurasi ataupun precision dan recall pada skenario 3, skenario 5 dan skenario 6.

Skenario 3 adalah model klasifikasi yang menggunakan teknik augmentasi data random zoom, hasil yang didapatkan dari skenario ini adalah nilai akurasi sebesar 92%, nilai precision sebesar 92% dan nilai recall sebesar 92%. Skenario 5 menghasilkan nilai akurasi sebesar 92%, kemudian nilai precision sebesar 93% dan nilai recall sebesar 92%. Skenario 6 adalah percobaan yang memiliki hasil terbaik dengan tingkat nilai akurasinya sebesar 93%, kemudian nilai precision sebesar 94% dan yang terakhir nilai recall sebesar 93%. Percobaan penggunaan augmentasi data pada skenario 4, skenario 7 dan 8 tidak menghasilkan peningkatan nilai akurasi dan juga tidak menghasilkan peningkatan nilai precision dan recall. Hasil dari teknik augmentasi random shear, horizontal flip dan vertical flip pada Gambar 4.10 grafik nilai akurasi, precision dan recall mengalami penurunan dari percobaan penggunaan model klasifikasi tanpa augmentasi data (skenario 2).

Hasil percobaan yang dilakukan menunjukkan bahwa semua teknik augmentasi data memberikan hasil yang baik karena menghasilkan nilai diatas 80% baik untuk akurasi, presisi dan recall. Namun dari beberapa teknik augmentasi yang digunakan ada yang menurunkan hasil akurasi dari percobaan tanpa menggunakan teknik augmentasi, padahal secara teori klasifikasi dengan arsitektur Convolutional Neural Network (CNN) akan menjadi lebih baik jika data yang digunakan semakin banyak. Hal ini bisa terjadi karena tidak semua teknik augmentasi data cocok dengan citra dataset yang digunakan untuk melakukan klasifikasi, dengan ketidakcocokan teknik augmentasi data akan menurunkan nilai dari model klasifikasi itu sendiri, oleh karena itu perlunya untuk memilih augmentasi data yang sesuai dengan citra yang digunakan (Hana, Liu and Fan, 2018).

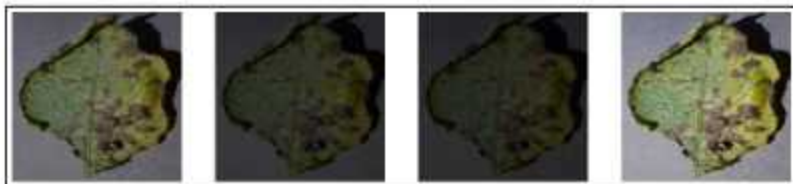
4.5.5 Rekomendasi teknik augmentasi data terbaik

Berdasarkan dari skenario 3 sampai 8 yaitu pengujian untuk mengetahui apakah nilai model klasifikasi masih dapat ditingkatkan menghasilkan bahwa penggunaan teknik augmentasi data zoom, brightness dan rotation memiliki hasil yang baik dibandingkan penggunaan teknik augmentasi data shear, horizontal flip dan vertical flip. Jika dilihat dari jenis teknik augmentasi yang digunakan dapat diambil kesimpulan bahwa teknik augmentasi data yang cocok untuk citra penyakit daun kentang adalah jenis teknik augmentasi yang dapat menghasilkan tingkat variasi datanya menjadi lebih beragam.



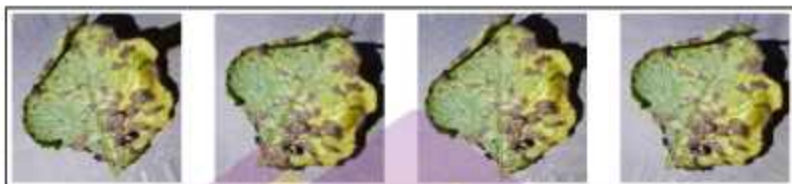
Gambar 4.12 Hasil teknik augmentasi data zoom

Gambar 4.12 menunjukkan augmentasi data zoom menghasilkan variasi citra yang diperkecil atau diperbesar dengan secara acak, hasil dari augmentasi ini nantinya dapat menjadi data tambahan supaya model klasifikasi dapat mengatasi apabila citra yang akan diidentifikasi memiliki ukuran yang kecil ataupun memiliki ukuran yang besar.



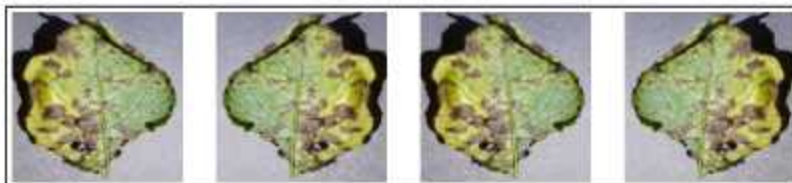
Gambar 4.13 Hasil teknik augmentasi data brightness

Gambar 4.13 menunjukkan augmentasi data brightness menghasilkan citra dengan tingkat kecerahan yang berbeda-beda secara acak. Hasil citra dari teknik ini dapat mewakili apabila citra yang akan diidentifikasi memiliki kecerahan yang berbeda-beda.

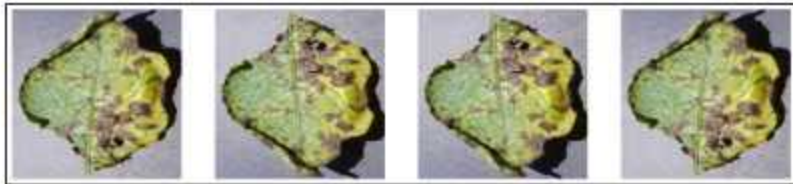


Gambar 4.14 Hasil teknik augmentasi data rotation

Gambar 4.14 menunjukkan augmentasi data rotation menghasilkan citra yang diputar secara acak. Hasil citra dari teknik ini dapat mewakili apabila citra yang akan diidentifikasi diambil dari sudut pandang yang tidak tegak lurus secara vertikal, karena saat melakukan identifikasi penyakit daun pada tanaman kentang posisi dari daun-daun nya akan acak. Teknik augmentasi data zoom, brightness dan rotation bisa cocok dalam meningkatkan hasil dari klasifikasi penyakit daun kentang dikarenakan dapat menjadikan citra lebih bervariasi, berbeda dengan teknik augmentasi horizontal dan vertical flip yang mana dua metode ini hanya dapat menghasilkan satu buah variasi data tambahan yaitu membalikkan citra secara horizontal dan vertical.



Gambar 4.15 Hasil teknik augmentasi data horizontal flip



Gambar 4.16 Hasil teknik augmentasi data *vertical flip*

Pada hasil uji penggunaan augmentasi data horizontal flip ditunjukkan Gambar 4.15, sedangkan untuk hasil penggunaan augmentasi data vertical flip ditunjukkan pada Gambar 4.16. Jika kita perhatikan pada citra yang dihasilkan setelah menggunakan teknik augmentasi horizontal dan vertical flip hanya dapat menambah variasi citra sekali sedangkan selanjutnya jika citra dibalikkan lagi maka akan menghasilkan seperti citra aslinya, jadi teknik ini tidak bisa menambah keberagaman variasi citra sebanyak teknik augmentasi zoom, brightness dan rotation.



Gambar 4.17 Hasil teknik augmentasi data shear

Gambar 4.17 merupakan citra yang dihasilkan dengan menggunakan teknik augmentasi data shear. Pengujian dengan menggunakan teknik ini menghasilkan nilai yang paling kecil diantara skenario percobaan yang lainnya. Citra dari penyakit daun early blight memiliki ciri-ciri bercak-bercak berwarna coklat yang ada pada daun, sedangkan untuk late blight memiliki ciri lesi hitam pada bagian daun. Teknik augmentasi data shear dirasa kurang cocok untuk identifikasi penyakit daun

tanaman kentang dikarenakan ciri dari citra penyakit early blight apabila diterapkan teknik augmentasi shear akan menghasilkan ciri penyakit yang menyerupai penyakit daun late blight.

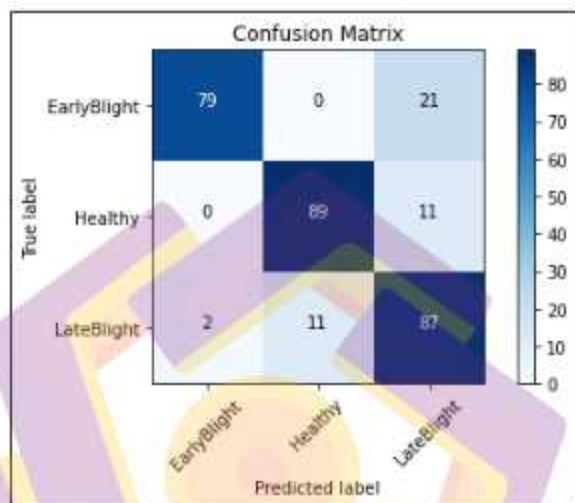
4.5.6 Pengujian augmentasi data lanjutan

Hasil dari perbandingan penggunaan augmentasi data yang digunakan menunjukkan bahwa teknik augmentasi data zoom, brightness dan rotation memiliki hasil lebih baik dibandingkan teknik augmentasi yang lainnya. Tahap selanjutnya adalah melakukan percobaan untuk menganalisa nilai akurasi, presisi dan recall dari klasifikasi penyakit daun kentang dengan menggabungkan teknik augmentasi data secara bersamaan. Tahap ini akan melakukan dua percobaan yaitu pertama dengan melakukan klasifikasi penyakit pada daun tanaman kentang dengan cara menggabungkan semua teknik augmentasi data yang digunakan pada skenario percobaan sebelumnya, kedua melakukan percobaan dengan menggunakan 3 augmentasi data yang terbaik. Percobaan menggunakan seluruh augmentasi data akan ditandai sebagai skenario 9 dan percobaan dengan menggunakan 3 teknik augmentasi data terbaik ditandai sebagai skenario 10.

4.5.6.1 Pengujian pada skenario 9

Skenario 9 adalah percobaan klasifikasi penyakit daun tanaman kentang dengan menggunakan seluruh augmentasi data yang digunakan pada percobaan sebelumnya. Teknik augmentasi data yang digunakan adalah random zoom, random shear, random brightness, random rotation, horizontal flip, dan vertical flip.

Percobaan skenario 9 ini menggunakan data yang sudah diseimbangkan seperti halnya pada skenario 2 sampai dengan 8, dan pembagian data juga sama yaitu 80% data training, 10% data validation dan 10% data testing.



Gambar 4.18 Confusion matrix pada skenario 9

Gambar 4.18 adalah hasil pengujian confusion matrix pada skenario 9. Nilai akurasi, precision dan recall dari hasil pengujian skenario 9 dapat dirincikan sebagai berikut:

1. Akurasi

Nilai akurasi didapatkan dari keseluruhan jumlah True Positive dari masing-masing kelas. Nilai *True Positive* dari seluruh kelas adalah $79 + 89 + 87 = 255$, kemudian jumlah data *testing* secara keseluruhan adalah 300 data. Nilai akurasi adalah total TP dibagi total data, maka $255 / 300 = 0.85$. Jadi nilai akurasi dari skenario 9 adalah sebesar 85%.

2. Precision

Precision didapatkan dengan menghitung nilai TP dibagi dengan TP + FP, namun untuk kasus pada confusion matrix multi class dalam menentukan nilai precision dihitung pada masing-masing kelas.

a. Nilai precision kelas *early blight*

$$\text{Precision} = 79 / (79 + 0 + 2) = 79 / 81 = 0.98 = 98\%$$

b. Nilai Precision kelas daun sehat

$$\text{Precision} = 89 / (0 + 89 + 11) = 89 / 100 = 0.89 = 89\%$$

c. Nilai Precision kelas *late blight*

$$\text{Precision} = 87 / (21 + 11 + 87) = 87 / 119 = 0.73 = 73\%$$

Hasil precision secara keseluruhan didapatkan dengan menjumlahkan semua nilai precision dari masing-masing kelas yang kemudian dibagi dengan jumlah kelas. Sehingga precision dari skenario 9 adalah $(0.98 + 0.89 + 0.73) / 3 = 2.6 / 3 = 0.87 = 87\%$

3. Recall

Recall didapatkan dengan menghitung nilai TP dibagi dengan TP + FN, namun untuk kasus pada confusion matrix multi class dalam menentukan nilai recall dihitung pada masing-masing kelas.

a. Nilai recall kelas *early blight*

$$\text{Recall} = 79 / (79 + 0 + 21) = 79 / 100 = 0.79 = 79\%$$

b. Nilai recall kelas daun sehat

$$\text{Recall} = 89 / (0 + 89 + 11) = 89 / 100 = 0.89 = 89\%$$

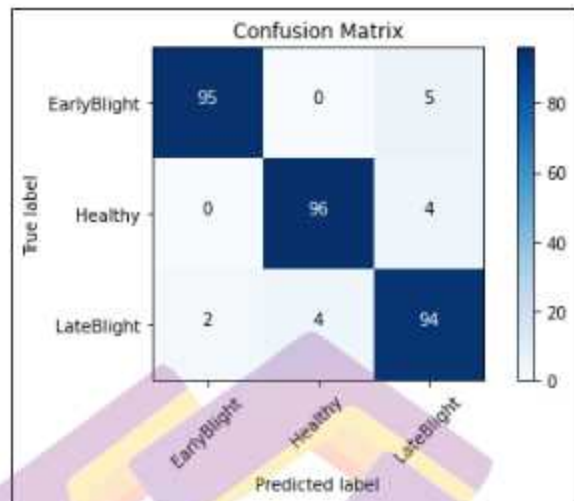
c. Nilai recall kelas *late blight*

$$\text{Recall} = 87 / (2 + 11 + 87) = 87 / 100 = 0.87 = 87\%$$

Hasil recall secara keseluruhan didapatkan dengan menjumlahkan semua nilai recall dari masing-masing kelas yang kemudian dibagi dengan jumlah kelas. Sehingga recall dari skenario 9 adalah $(0.79 + 0.89 + 0.87) / 3 = 2.55 / 3 = 0.85 = 85\%$

4.5.6.2 Pengujian pada skenario 10

Skenario 10 adalah percobaan klasifikasi penyakit daun tanaman kentang dengan menggunakan tiga buah teknik augmentasi data terbaik yang digunakan pada percobaan sebelumnya. Teknik augmentasi data yang digunakan adalah random zoom, random brightness, dan random rotation. Percobaan skenario 10 ini menggunakan data yang sudah diseimbangkan seperti halnya pada skenario 2 sampai dengan 9, dan pembagian data juga sama yaitu 80% data training, 10% data validation dan 10% data testing.



Gambar 4.19 Confusion matrix pada skenario 10

Gambar 4.19 adalah hasil pengujian confusion matrix pada skenario 10. Nilai akurasi, precision dan recall dari hasil pengujian skenario 10 dapat dirincikan sebagai berikut:

1. Akurasi

Nilai akurasi didapatkan dari keseluruhan jumlah True Positive dari masing-masing kelas. Pada Gambar 4.19 nilai True Positive (TP) berada pada matriks yang berwarna biru gelap yaitu $95 + 96 + 94 = 192$, kemudian jumlah data *testing* secara keseluruhan adalah 300 data. Nilai akurasi adalah total TP dibagi total data, maka $192 / 300 = 0.64$. Jadi nilai akurasi dari skenario 9 adalah sebesar 64%.

2. Precision

Precision didapatkan dengan menghitung nilai TP dibagi dengan TP + FP, namun untuk kasus pada confusion matrix multi class dalam menentukan nilai precision dihitung pada masing-masing kelas.

a. Nilai precision kelas *early blight*

$$\text{Precision} = 95 / (95 + 0 + 2) = 95 / 97 = 0.98 = 98\%$$

b. Nilai Precision kelas daun sehat

$$\text{Precision} = 96 / (0 + 96 + 4) = 96 / 100 = 0.96 = 96\%$$

c. Nilai Precision kelas *late blight*

$$\text{Precision} = 94 / (4 + 5 + 94) = 94 / 103 = 0.91 = 91\%$$

Hasil precision secara keseluruhan didapatkan dengan menjumlahkan semua nilai precision dari masing-masing kelas yang kemudian dibagi dengan jumlah kelas. Sehingga precision dari skenario 9 adalah $(0.98 + 0.96 + 0.91) / 3 = 2.85 / 3 = 0.95 = 95\%$

3. Recall

Recall didapatkan dengan menghitung nilai TP dibagi dengan TP + FN, namun untuk kasus pada confusion matrix multi class dalam menentukan nilai recall dihitung pada masing-masing kelas.

a. Nilai recall kelas *early blight*

$$\text{Recall} = 95 / (95 + 0 + 5) = 95 / 100 = 0.95 = 95\%$$

b. Nilai recall kelas daun sehat

$$\text{Recall} = 96 / (0 + 96 + 4) = 96 / 100 = 0.96 = 96\%$$

c. Nilai recall kelas *late blight*

$$\text{Recall} = 94 / (4 + 2 + 94) = 94 / 100 = 0.94 = 94\%$$

Hasil recall secara keseluruhan didapatkan dengan menjumlahkan semua nilai recall dari masing-masing kelas yang kemudian dibagi dengan jumlah kelas. Sehingga recall dari skenario 9 adalah $(0.95 + 0.96 + 0.94) / 3 = 2.85 / 3 = 0.95 = 95\%$.

4.5.7 Perbandingan hasil seluruh skenario percobaan

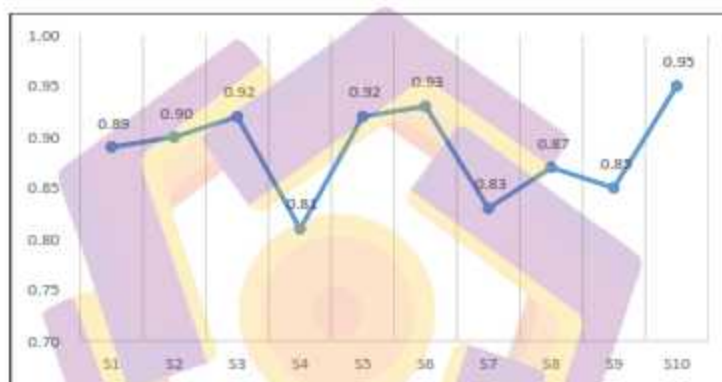
Tahap akhir dari penelitian ini adalah melakukan analisis perbandingan dari keseluruhan skenario yang telah diujikan. Pada penelitian ini terdapat 10 skenario percobaan secara keseluruhan.



Gambar 4.20 Perbandingan nilai akurasi dari data training dan data validasi

Gambar 4.20 adalah grafik hasil perbandingan dari pelatihan model klasifikasi dalam mengidentifikasi penyakit pada daun tanaman kentang dari

skenario 1 (S1) sampai dengan skenario 10 (S10). Hasil pada grafik menunjukkan bahwa Skenario 10 memiliki hasil yang paling baik dengan nilai akurasi pada data training 96% dan pada data validation 94%, kemudian diikuti oleh skenario 6 dengan nilai akurasi pada data training 93% dan pada data validation 96% dan kemudian skenario 5 dengan nilai akurasi pada data training 93% dan pada data validation 94%.



Gambar 4.21 Nilai Akurasi dari seluruh skenario percobaan



Gambar 4.22 Nilai Precision dari seluruh skenario percobaan



Gambar 4.23 Nilai Recall dari seluruh skenario percobaan.

Gambar 4.21 merupakan grafik perbandingan dari seluruh skenario percobaan pada nilai akurasi, Gambar 4.22 adalah grafik perbandingan untuk nilai precision dan Gambar 4.23 untuk nilai recall. Pada legenda grafik maksud dari S1 adalah skenario 1, kemudian S2 adalah skenario 2, S3 adalah skenario 3 dan selanjutnya hingga S10 adalah skenario 10. Nilai terbaik didapatkan oleh skenario 10. Skenario ini menggunakan arsitektur CNN seperti pada skenario yang lainnya, kemudian data yang digunakan adalah yang telah diseimbangkan, kemudian menggunakan teknik augmentasi data random zoom, random brightness dan random rotation yang mana adalah 3 teknik augmentasi terbaik dari keseluruhan teknik augmentasi yang digunakan. Hasil yang didapatkan pada skenario 9 ini adalah 95% pada nilai akurasi, 95% pada precision dan 95% pada recall.

4.6 Skenario Percobaan dengan Arsitektur VGG-16

Percobaan dalam melakukan klasifikasi penyakit daun tanaman kentang pada penelitian ini mendapatkan hasil yang baik. Percobaan yang dilakukan adalah

dengan menggunakan arsitektur CNN yang diusulkan dan dibantu dengan pemanfaatan teknik data augmentation dalam menyeimbangkan jumlah dari tiap-tiap kelas dari dataset yang digunakan, kemudian teknik data augmentation juga digunakan untuk meningkatkan hasil model klasifikasi. Dalam menghasilkan penelitian yang lebih baik atas saran dari dosen penguji peneliti melakukan percobaan dari seluruh skenario yang digunakan pada model klasifikasi dengan menggunakan arsitektur CNN yang berbeda. Arsitektur yang akan digunakan pada percobaan ini adalah VGG-16, penentuan arsitektur ini berdasarkan dari beberapa penelitian terdahulu seperti penelitian terkait klasifikasi penyakit daun padi menggunakan VGG-16 dan menghasilkan hasil yang sangat baik (Ghosal and Sarkar, 2020), kemudian penelitian mengenai klasifikasi penyakit hama daun tembakau menggunakan arsitektur VGG-16 menghasilkan akurasi yang sangat tinggi (Swasono, Tjandrasa and Fathicah, 2019), terakhir penelitian yang membandingkan beberapa algoritma untuk identifikasi penyakit daun dengan hasil akhir bahwa VGG-16 memiliki hasil yang paling baik dalam melakukan klasifikasi (Sujatha *et al.*, 2021).

Tabel 4.3 Skenario Percobaan dengan arsitektur VGG-16

No	Skenario	Arsitektur	Augmentasi Data	
			Penyeimbangan Data	Augmentasi Tambahan
1	S1	VGG-16	No	No
2	S2	VGG-16	Yes	No
3	S3	VGG-16	Yes	Random Zoom

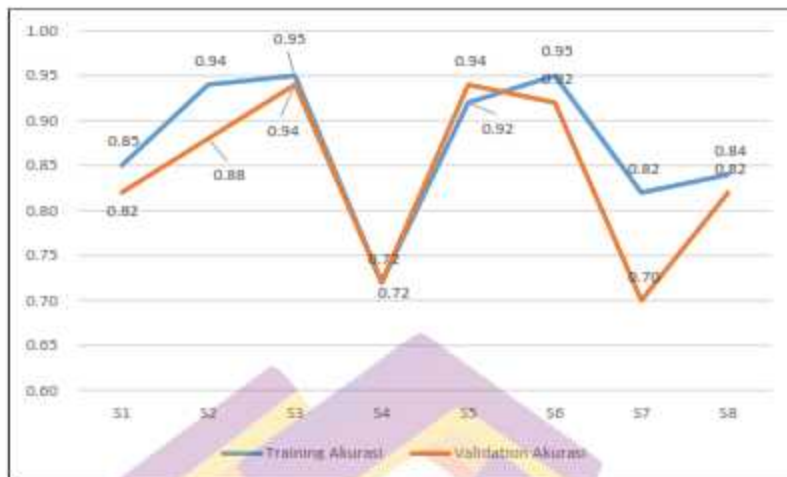
Tabel 4.3 Lanjutan

No	Skenario	Arsitektur	Augmentasi Data	
			Penyeimbangan Data	Augmentasi Tambahan
4	S4	VGG-16	Yes	Random Shear
5	S5	VGG-16	Yes	Random Brightness
6	S6	VGG-16	Yes	Rotation
7	S7	VGG-16	Yes	Horizontal flip
8	S8	VGG-16	Yes	Vertical flip

Tabel 4.3 adalah seluruh skenario percobaan untuk melakukan klasifikasi penyakit daun tanaman kentang dengan arsitektur VGG-16, seluruh skenario percobaan yang digunakan sama dengan skenario yang digunakan pada percobaan sebelumnya dengan menggunakan arsitektur CNN sederhana yang diusulkan.

4.6.1 Hasil pelatihan model klasifikasi arsitektur VGG-16

Percobaan dengan menggunakan arsitektur VGG-16 dilakukan pada 8 skenario seperti yang ditampilkan pada Tabel 4.3 dan seluruh skenario ini sama dengan skenario yang sudah dilakukan pada percobaan sebelumnya dengan menggunakan arsitektur CNN sederhana yang diusulkan. Skenario 1 adalah model klasifikasi dengan menggunakan data yang tidak seimbang, skenario 2 adalah model klasifikasi dengan data yang telah diseimbangkan, sedangkan untuk skenario 3 sampai skenario 8 adalah model klasifikasi dengan menggunakan data yang telah diseimbangkan dengan tambahan teknik augmentasi data.

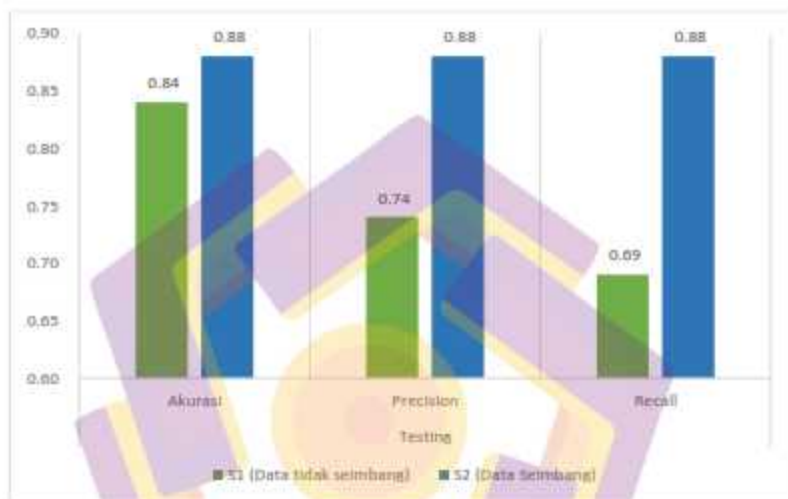


Gambar 4.24 Akurasi pada data training dan validation (VGG-16)

Hasil pelatihan untuk model klasifikasi dengan arsitektur VGG-16 ditampilkan pada Gambar 4.24. Hasil pelatihan dari nilai akurasi pada data training dan data validation dengan arsitektur VGG-16 skenario 3 mendapatkan hasil terbaik dengan nilai akurasi pada data training 95% dan nilai akurasi pada data validasi sebesar 94%. Namun ini belum dapat menentukan skenario terbaik dari model klasifikasi karena selanjutnya akan dilakukan klasifikasi dengan menggunakan data testing yang kemudian diujikan dengan confusion matrix untuk dianalisa perbandingan penggunaan data yang tidak seimbang (skenario 1) dengan data yang seimbang (skenario 2), selanjutnya juga dilakukan analisa perbandingan penggunaan dari augmentasi data (skenario 3 sampai 8).

4.6.2 Perbandingan skenario 1 dan 2 pada arsitektur VGG-16

Pengujian confusion matrix telah diujikan pada skenario 1 atau skenario percobaan dengan menggunakan data yang tidak seimbang dan skenario 2 atau skenario percobaan dengan menggunakan data yang telah diseimbangkan.

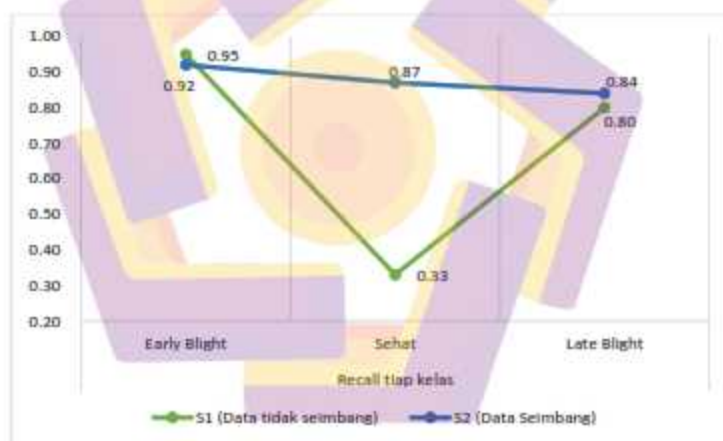


Gambar 4.25 Grafik perbandingan skenario 1 dan 2 pada arsitektur VGG-16

Gambar.4.25 adalah grafik perbandingan dari hasil percobaan menggunakan data *testing* pada skenario 1 dan skenario 2, dari grafik yang ditunjukkan dapat diperhatikan nilai yang dihasilkan adalah akurasi, precision dan recall. Grafik warna hijau untuk skenario 1 dan grafik berwarna biru untuk skenario 2. Nilai Akurasi yang dihasilkan pada skenario 1 adalah 84% dan skenario 2 88%, sama halnya pada perbandingan skenario 1 dan 2 pada arsitektur CNN sederhana untuk nilai akurasi memiliki hasil yang tidak jauh berbeda.

Hasil pengujian pada data testing untuk nilai precision menunjukkan bahwa nilai pada skenario 1 menghasilkan nilai 74% sedangkan skenario 2 menghasilkan

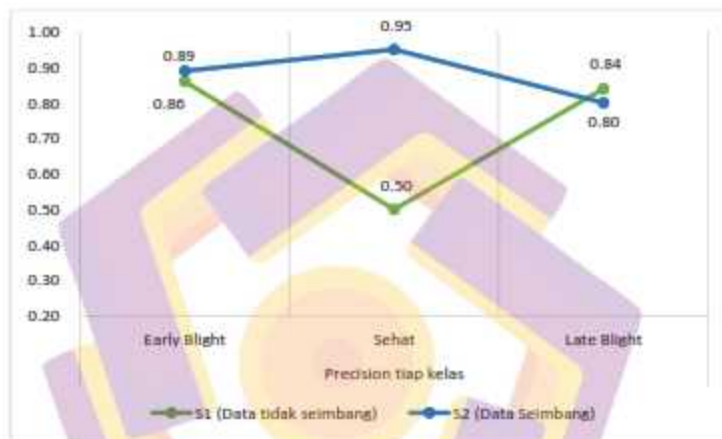
nilai 88%, berbeda dengan hasil perbandingan precision pada skenario percobaan dengan CNN sederhana dimana nilai precision milik skenario 1 lebih tinggi dibanding nilai dari skenario 2. Hasil pengujian data testing pada nilai recall dari skenario 1 dan 2 dari arsitektur VGG-16 menghasilkan nilai recall skenario 1 lebih rendah dibandingkan nilai recall dari skenario 2 dan ini sama dengan pengujian pada arsitektur CNN sederhana yang sebelumnya dilakukan. Nilai recall dari skenario 1 adalah 69% dan nilai recall dari skenario 2 adalah 88%. Selanjutnya akan dijabarkan lebih detail untuk mengetahui penyebab rendahnya nilai precision dan recall pada skenario 1.



Gambar 4.26 Nilai *recall* tiap kelas dari skenario 1 dan 2 pada VGG-16

Gambar 4.26 merupakan grafik dari perbandingan nilai recall dari skenario 1 (data tidak seimbang) dan skenario 2 (data seimbang). Grafik berwarna hijau untuk skenario 1 dan grafik berwarna biru untuk skenario 2. Seperti yang sudah dijelaskan sebelumnya recall merupakan kemampuan model klasifikasi dalam menemukan kembali data yang sesungguhnya. Gambar 4.26 menunjukkan bahwa

nilai recall pada skenario 1 untuk kelas daun sehat sangatlah kecil, hal ini terjadi karena data yang digunakan pada kelas daun sehat lebih sedikit dibanding kelas lainnya sehingga nilai yang dihasilkan menjadi kecil. Sedangkan pada skenario 2 nilai recall pada kelas daun sehat tidak memiliki perbedaan yang jauh karena pada skenario 2 menggunakan data yang telah diseimbangkan.



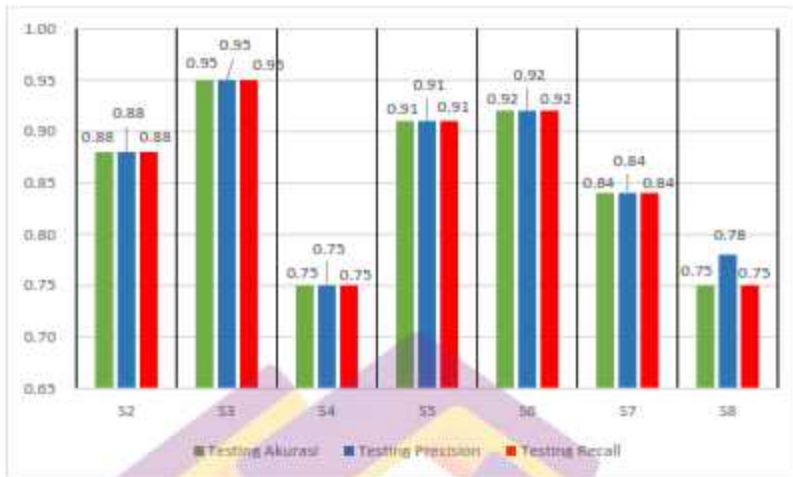
Gambar 4.27 Nilai precision tiap kelas dari skenario 1 dan 2 pada VGG-16

Gambar 4.27 adalah grafik dari perbandingan nilai precision dari tiap kelas dari skenario 1 dan 2 pada arsitektur VGG-16. Pada grafik berwarna hijau adalah skenario dengan data yang tidak seimbang, hasil yang ditunjukkan adalah nilai precision dari kelas daun sehat memiliki nilai paling kecil yaitu sebesar 50%, berbeda dengan nilai precision dari kelas *early blight* dan *late blight* yang memiliki hasil lebih tinggi. Selanjutnya jika dilihat untuk grafik berwarna biru yaitu skenario dengan data seimbang memiliki nilai precision yang baik pada tiap kelasnya karena nilai yang dihasilkan tidak memiliki perbedaan yang jauh.

Sebelumnya telah dijelaskan bahwa nilai precision dan recall saling berhubungan satu sama lainnya dan untuk model klasifikasi yang baik adalah dengan nilai precision dan recall yang sama baiknya, jika ada salah satu atau bahkan keduanya mendapatkan nilai yang kurang bagus maka hasil klasifikasinya akan kurang baik juga. Perbandingan dari skenario 1 dan 2 dengan arsitektur VGG-16 memiliki hasil yang sama dengan skenario 1 dan 2 menggunakan arsitektur CNN sederhana yaitu skenario 2 atau skenario dengan menggunakan data yang seimbang memiliki hasil yang lebih baik.

4.6.3 Perbandingan augmentasi data pada arsitektur VGG-16

Skenario selanjutnya adalah percobaan dengan menggunakan teknik augmentasi data pada data yang telah diseimbangkan dengan arsitektur VGG-16. Seperti pada percobaan sebelumnya teknik augmentasi data yang digunakan adalah random zoom pada skenario 3, random shear pada skenario 4, random brightness pada skenario 5, random rotation pada skenario 6, horizontal flip pada skenario 7, dan yang terakhir vertical flip pada skenario 8.



Gambar 4.28 Perbandingan pemanfaatan augmentasi data pada VGG-16

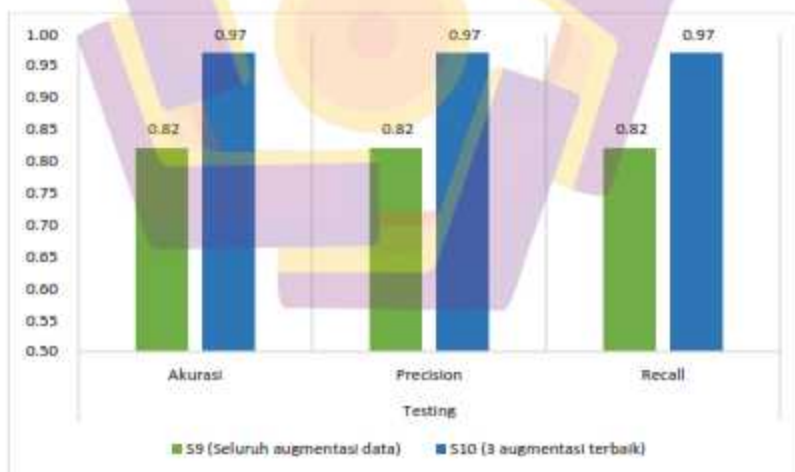
Gambar 4.28 merupakan perbandingan dari penggunaan teknik augmentasi data pada klasifikasi penyakit daun tanaman kentang menggunakan arsitektur VGG-16. Grafik berwarna hijau adalah nilai akurasi dari masing-masing skenario percobaan, grafik berwarna biru adalah nilai precision dari masing-masing skenario percobaan, dan grafik berwarna merah adalah nilai recall dari masing-masing skenario percobaan.

Skenario 2 adalah klasifikasi dengan menggunakan data yang telah diseimbangkan namun tidak menggunakan augmentasi tambahan, selanjutnya pada skenario 3 sampai 8 menggunakan data seimbang dan juga augmentasi tambahan. Seperti pada percobaan menggunakan arsitektur CNN sederhana, skenario yang tidak bisa melebihi nilai dari skenario 2 adalah skenario 4, 7 dan 8. Kemudian 3 teknik augmentasi data terbaik dan bisa lebih baik dari skenario 2 adalah random

zoom (skenario 3), random brightness (skenario 5) dan random rotation (skenario 6).

4.6.4 Augmentasi data lanjutan pada arsitektur VGG-16

Tahap selanjutnya adalah melakukan pengujian augmentasi data lanjutan, terdapat 2 skenario untuk percobaan lanjutan yaitu skenario dengan menggunakan seluruh teknik augmentasi data yang digunakan secara bersamaan, kemudian skenario dengan menggunakan 3 augmentasi data terbaik secara bersamaan. 2 skenario yang akan dilakukan diberi nama skenario 9 untuk uji coba menggunakan seluruh teknik augmentasi data dan skenario 10 untuk uji coba menggunakan 3 teknik augmentasi data terbaik.



Gambar 4.29 Perbandingan skenario 9 dan 10 pada VGG-16

Gambar 4.29 adalah perbandingan dari skenario 9 dan 10 dengan menggunakan arsitektur VGG-16. Hasil yang didapatkan dari kedua percobaan ini

adalah skenario 10 mendapatkan hasil terbaik dengan nilai akurasi sebesar 97%, kemudian nilai precision sebesar 97% dan nilai recall sebesar 97%.

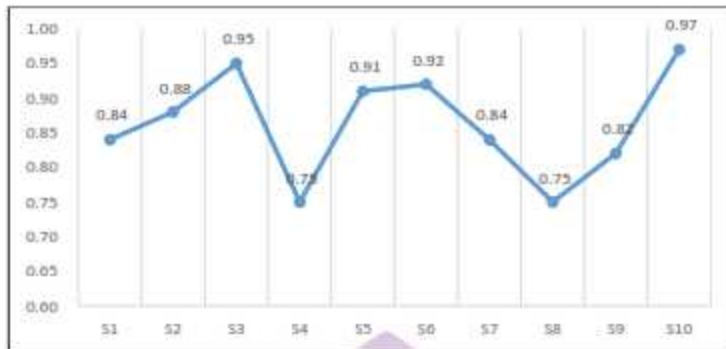
4.6.5 Perbandingan hasil seluruh skenario pada arsitektur VGG-16

Tahap selanjutnya adalah melakukan analisa perbandingan dari keseluruhan skenario yang telah digunakan, sama halnya pada percobaan sebelumnya total skenario yang digunakan adalah 10 skenario.

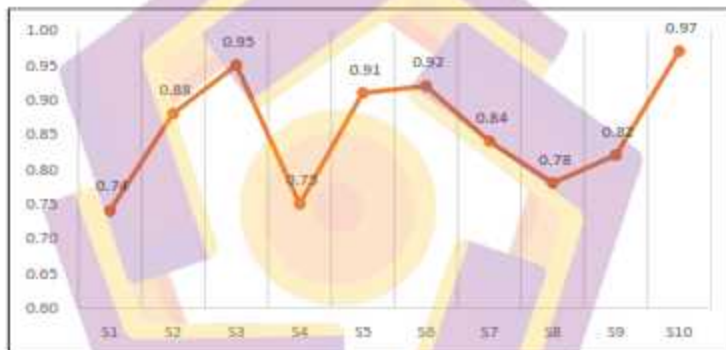


Gambar 4.30 Perbandingan nilai akurasi dari data training dan data validation

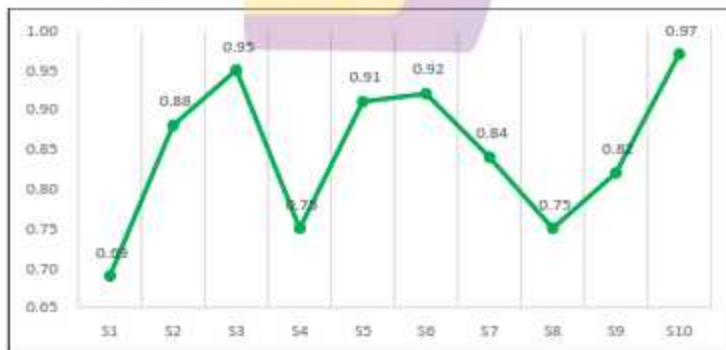
Gambar 4.30 adalah grafik perbandingan dari pelatihan model klasifikasi menggunakan arsitektur VGG-16 pada skenario 1 (S1) sampai dengan skenario 10 (S10). Hasil terbaik ditunjukkan pada skenario 10 dengan tingkat nilai akurasi sebesar 97% pada data training dan 96% pada data validation.



Gambar 4.31 Nilai akurasi dari seluruh skenario dengan VGG-16



Gambar 4.32 Nilai Precision dari seluruh skenario dengan VGG-16



Gambar 4.33 Nilai Recall dari seluruh skenario dengan VGG-16

Gambar 4.31 adalah perbandingan nilai akurasi dari seluruh skenario percobaan menggunakan arsitektur VGG-16, Gambar 4.32 adalah perbandingan untuk nilai precision dan Gambar 4.33 adalah perbandingan untuk nilai recall. Pada legenda grafik maksud dari S1 adalah skenario 1, kemudian S2 adalah skenario 2, dan selanjutnya hingga S10 adalah skenario 10. Nilai terbaik dari seluruh skenario adalah skenario 10 yaitu model klasifikasi dengan menggunakan data yang telah diseimbangkan, kemudian menggunakan teknik augmentasi data tambahan berupa random zoom, random brightness dan random rotation secara bersamaan. Nilai akurasi dari skenario 10 adalah 97%, kemudian nilai precision 97% dan nilai recall juga mencapai 97%.



BAB V

PENUTUP

5.1. Kesimpulan

Berdasarkan pada percobaan penelitian yang telah dilakukan, maka dapat diambil kesimpulan sebagai berikut:

1. Penggunaan data yang memiliki persebaran jumlah data tidak seimbang akan menimbulkan kurangnya ketepatan dalam mengidentifikasi pada kelas yang memiliki data paling sedikit, dibuktikan pada skenario 1 baik menggunakan CNN sederhana ataupun VGG-16 nilai recall pada kelas daun sehat mendapatkan nilai yang sangat rendah.
2. Penyeimbangan data dengan menggunakan teknik data augmentation terbukti dapat meningkatkan kinerja suatu model klasifikasi, dibuktikan pada skenario 2 baik menggunakan CNN sederhana ataupun VGG-16 hasil akurasi, precision dan recall pada tiap kelas menghasilkan nilai yang lebih baik.
3. Penerapan augmentasi data setelah data diseimbangkan menghasilkan bahwa teknik augmentasi data random zoom, random brightness dan random rotation adalah teknik augmentasi terbaik pada klasifikasi penyakit pada daun tanaman kentang.
4. Hasil terbaik dari keseluruhan skenario percobaan adalah skenario 10 dengan menggunakan arsitektur VGG-16. Percobaan ini menggunakan data yang telah diseimbangkan, kemudian menggunakan 3 teknik

augmentasi data terbaik. Hasil yang didapatkan adalah 97% pada nilai akurasi, 97% pada nilai precision dan 97% pada nilai recall.

5.2. Saran

Berdasarkan analisis hasil percobaan yang dilakukan, maka ada beberapa saran yang dapat digunakan untuk penelitian selanjutnya adalah sebagai berikut:

1. Pada penelitian selanjutnya metode penyeimbangan data yang digunakan pada penelitian ini dapat diuji coba dengan menggunakan data yang tidak seimbang lainnya (selain data daun kentang).
2. Sebaiknya pada proses untuk meningkatkan nilai akurasi pada model klasifikasi yang dilakukan dapat mencoba menerapkan preprocessing selain augmentasi data seperti median filter, green channel, CLAHE dan lain sebagainya. Juga pada proses klasifikasi dapat menguji coba dengan penggunaan dropout ataupun L Regularization.
3. Penelitian selanjutnya dapat mengembangkan berupa penentuan penanganan apa yang akan dilakukan ketika penyakit daun tanaman kentang sudah diidentifikasi. Penanganan yang dilakukan didasarkan pada level penyakit yang diidentifikasi, level penyakit dapat mengacu kepada jumlah daun yang berpenyakit pada setiap pohonnya.

DAFTAR PUSTAKA

PUSTAKA BUKU

- Andono, P. N., T. Sutojo and Muljono (2017) *Pengolahan Citra Digital*. Edited by A. Pramesta. Yogyakarta: Percetakan Andi.
- Arhami, M. and Nasir, M. (2020) *Data Mining Algoritma dan Implementasi*. 1st edn. Edited by R. I. Utami. Yogyakarta: Andi Offset.
- Harani, N. H. and Hasanah, M. (2020) *Deteksi Objek dan Pengenalan Karakter Plat Nomor Kendaraan Indonesia Berbasis Python*. Pertama. Edited by M. Y. H. Setyawan. Bandung: Kreatif Industri Nusantara.
- Heryadi, Y. and Irwansyah, E. (2020) *DEEP LEARNING dan Aplikasinya di Bidang Informasi Geospasial*. Depok: Artifisia Wahana Informa Technology Press.
- Hidayatullah, P. (2017) *Pengolahan Citra Digital Teori dan Aplikasi Nyata*. Cetakan Pe. Bandung: Informatika Bandung.
- Nafi'yah, N. and Mujilawati, S. (2018) *Buku Ajar CITRA BINARISASI DAN ENHANCEMENT*. Sleman, Yogyakarta: DEEPUBLISH.
- Rukajat, A. (2018) *Pendekatan Penelitian Kuantitatif: Quantitative Research Approach*. Pertama. Yogyakarta: DEEPUBLISH.
- Shanmugamani, R. (2018) *Deep Learning for Computer Vision*. First Edit. Brimingham: Packt Publishing.
- Siahaan, V. and Sianipar, R. H. (2020) *PANDUAN PRAKTIS DAN KOMPLET Pemrosesan Citra Digital Dengan MATLAB*. North Sumatera: Balige Publishing.
- Suyanto (2017) *Data Mining Untuk Klasifikasi dan Klusterisasi Data*. 1st edn. Bandung: Informatika Bandung.

PUSTAKA MAJALAH, JURNAL ILMIAH ATAU PROSIDING

- Agarwal, M. *et al.* (2020) 'ToLeD: Tomato Leaf Disease Detection using Convolution Neural Network', *Procedia Computer Science*, Elsevier B.V., 167(2019), pp. 293–301. doi: 10.1016/j.procs.2020.03.225.

- Arrofiqoh, E. N. and Harintaka, H. (2018) 'Implementasi Metode Convolutional Neural Network Untuk Klasifikasi Tanaman Pada Citra Resolusi Tinggi', *Jurnal Ilmiah Geomatika-JIG*, 24(2), p. 61. doi: 10.24895/jig.2018.24-2.810.
- Beals, K. A. (2019) 'Potatoes , Nutrition and Health', *American Journal of Potato Research*, 96, pp. 102–110. doi: 10.1007/s12230-018-09705-4.
- Buslaev, A. *et al.* (2018) 'Albumentations: Fast and flexible image augmentations', *Information (Switzerland)*, 11(2). doi: 10.3390/info11020125.
- Fitriana, A. and Hakim, L. (2019) 'Endophytic Fungus effectiveness Origin Root Crops Potatoes in Compost Media Skin Diseases Coffee to Suppress Development Potato Leaf blight (Phytophthora infestans)', *Jurnal Agrista*, 23(1), pp. 9–14.
- Furrer, A. N., Chegeni, M. and Ferruzzi, M. G. (2018) 'Impact of potato processing on nutrients, phytochemicals, and human health', *Critical Reviews in Food Science and Nutrition*, 58(1), pp. 146–168. doi: 10.1080/10408398.2016.1139542.
- Ghosal, S. and Sarkar, K. (2020) 'Rice Leaf Diseases Classification Using CNN With Transfer Learning', in *2020 IEEE Calcutta Conference (CALCON)*, IEEE, pp. 230–236. doi: 10.1109/CALCON49167.2020.9106423.
- Hana, D., Liu, Q. and Fan, W. (2018) 'A new image classification method using CNN transfer learning and web data augmentation', *Expert Systems with Applications*. Elsevier Ltd, 95, pp. 43–56. doi: 10.1016/j.eswa.2017.11.028.
- Iqbal, M. A. and Talukder, K. H. (2020) 'Detection of Potato Disease Using Image Segmentation and Machine Learning', *2020 International Conference on Wireless Communications, Signal Processing and Networking, WiSPNET 2020*, pp. 43–47. doi: 10.1109/WiSPNET48689.2020.9198563.
- Islam, M. *et al.* (2017) 'Detection of potato diseases using image segmentation and multiclass support vector machine', in *2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE)*. IEEE, pp. 1–4. doi: 10.1109/CCECE.2017.7946594.
- Juba, B. and Le, H. S. (2019) 'Precision-Recall versus accuracy and the role of large data sets', *33rd AAAI Conference on Artificial Intelligence, AAAI 2019*, pp. 4039–4048. doi: 10.1609/aaai.v33i01.33014039.
- Kumbar, B. *et al.* (2019) 'Field application of *Bacillus subtilis* isolates for controlling late blight disease of potato caused by *Phytophthora infestans*', *Biocatalysis and Agricultural Biotechnology*. Elsevier Ltd, 22(April), p. 101366. doi: 10.1016/j.bcab.2019.101366.

- Manliguez, G. (2016) 'Generalized Confusion Matrix for Multiple Classes', (November), pp. 5–7. doi: 10.13140/RG.2.2.31150.51523.
- Ni'mah, F. S., Sutojo, T. and Setiadi, D. R. I. M. (2018) 'Identifikasi Tumbuhan Obat Herbal Berdasarkan Citra Daun Menggunakan Algoritma Gray Level Co-occurrence Matrix dan K-Nearest Neighbor', *Jurnal Teknologi dan Sistem Komputer*, 6(2), pp. 51–56. doi: 10.14710/jtsiskom.6.2.2018.51-56.
- Nisa, C., Puspaningrum, E. Y. and Maulana, H. (2020) 'Penerapan Metode Convolutional Neural Network untuk Klasifikasi Penyakit Daun Apel pada Imbalanced Data', in *Prosiding Seminar Nasional Informatika Bela Negara (SANTIKA)*. Surabaya, pp. 169–175.
- Patil, P., Yaligar, N. and Meena, S. (2017) 'Comparison of Performance of Classifiers - SVM, RF and ANN in Potato Blight Disease Detection Using Leaf Images', *IEEE International Conference on Computational Intelligence and Computing Research, ICCIC 2017*. IEEE, pp. 1–5. doi: 10.1109/ICCIC.2017.8524301.
- Rakhmawati, P. U., Pranoto, Y. M. and Setyati, E. (2018) 'Klasifikasi Penyakit Daun Kentang Berdasarkan Fitur Tekstur Dan Fitur Warna Menggunakan Support Vector Machine', *Seminar Nasional Teknologi dan Rekayasa (SENTRA) 2018*, pp. 1–8.
- Sujatha, R. *et al.* (2021) 'Performance of deep learning vs machine learning in plant leaf disease detection', *Microprocessors and Microsystems*. Elsevier B.V., 80(October 2020), p. 103615. doi: 10.1016/j.micpro.2020.103615.
- Suttapakti, U. and Bunpeng, A. (2019) 'Potato Leaf Disease Classification Based on Distinct Color and Texture Feature Extraction', *Proceedings - 2019 19th International Symposium on Communications and Information Technologies, ISCIT 2019*. IEEE, (Mcd), pp. 82–85. doi: 10.1109/ISCIT.2019.8905128.
- Swasono, D. I., Tjandrasa, H. and Fathicah, C. (2019) 'Classification of tobacco leaf pests using VGG16 transfer learning', *International Conference on Information and Communication Technology and Systems, ICTS 2019*. IEEE, pp. 176–181. doi: 10.1109/ICTS.2019.8850946.
- Tafesse, S. *et al.* (2018) 'Farmers' knowledge and practices of potato disease management in Ethiopia', *NJAS - Wageningen Journal of Life Sciences*. Elsevier, 86–87(September 2017), pp. 25–38. doi: 10.1016/j.njas.2018.03.004.
- Tumewu, S. F., Setiabud, D. H. and Sugiarto, I. (2020) 'Klasifikasi Motif Batik menggunakan metode Deep Convolutional Neural Network dengan Data Augmentation', *Jurnal Infra*, 8(2), pp. 189–194.

- Widiyanto, S., Fitrianto, R. and Wardani, D. T. (2019) 'Implementation of Convolutional Neural Network Method for Classification of Diseases in Tomato Leaves', *Proceedings of 2019 4th International Conference on Informatics and Computing, ICIC 2019*, doi: 10.1109/ICIC47613.2019.8985909.
- Wu, B. *et al.* (2017) 'Reducing overfitting in deep convolutional neural networks using redundancy regularizer', *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10614 LNCS, pp. 49–55. doi: 10.1007/978-3-319-68612-7_6.

