

**LAPORAN AKHIR**  
**STUDI INDEPENDEN BERSERTIFIKAT**  
**Foundations of AI and Life Skills for Gen-Z**  
**Di Orbit Future Academy**

Diajukan untuk memenuhi persyaratan kelulusan  
Program MSIB MBKM

oleh :  
Nyayu Chika Marselina / 08011281924027



**MATEMATIKA UNIVERSITAS SRIWIJAYA**  
**2022**

**Lembar Pengesahan Matematika Universitas Sriwijaya**

**PEMBELAJARAN STUDI INDEPENDEN**

**Di Orbit Future Academy**

oleh :

Nyayu Chika Marselina / 08011281924027

disetujui dan disahkan sebagai

Laporan Magang atau Studi Independen Bersertifikat Kampus Merdeka

Indralaya, 15 Juni 2022

Pembimbing Magang atau Studi Independen Matematika Universitas Sriwijaya



Anita Desiani, S.Si., M.Kom.

NIP: 197712112003122002

**Lembar Pengesahan**  
**PEMBELAJARAN STUDI INDEPENDEN**  
**Di Orbit Future Academy**

oleh :

Nyayu Chika Marselina / 08011281924027

disetujui dan disahkan sebagai  
Laporan Magang atau Studi Independen Bersertifikat Kampus Merdeka

Bandung, 16 Juni 2022

AI Coach



Angel Metanosa Afinda, S.Kom.

NIP: 2201043

## Abstraksi

Orbit Future Academy (OFA) merupakan salah satu instansi yang bekerja sama dengan program Magang dan Studi Independen Bersertifikat (MSIB) Kampus Merdeka. OFA bergerak di bidang perkembangan inovasi dan teknologi pendidikan di Indonesia. Salah satu misi dari OFA, yakni membangun jaringan Orbit Transformation Center (OTC) secara nasional untuk menyampaikan kurikulum keterampilan masa depan berbasis sertifikasi melalui Platform Konten Digital sehingga misi tersebut berkesinambungan secara langsung dengan salah satu tujuan dari program MSIB Kampus Merdeka. Pada pelaksanaan MSIB, OFA berfokus dalam pemanfaatan *Artificial Intelligence* (AI) sebagai topik utama yang dipelajari oleh peserta MSIB. Pada akhir pembelajaran, OFA memberikan sebuah Proyek Akhir (PA) kepada para peserta MSIB agar peserta dapat menerapkan langsung materi pembelajaran selama mengikuti program MSIB di OFA. Peserta diharapkan dapat menerapkan salah satu dari ketiga domain AI yang dipelajari dengan melakukan pengembangan model menggunakan algoritma tertentu melalui PA yang dikerjakan secara berkelompok. Adapun basis dari PA, yakni aplikasi atau penelitian dimana kelompok PA dapat memilih salah satu basis dari PA yang dikerjakan.

Kata Kunci: Orbit Future Academy, Magang dan Studi Independen Bersertifikat, *Artificial Intelligence*, Proyek Akhir.

## Kata Pengantar

Puji dan syukur atas kehadiran Allah SWT, sehingga laporan akhir program Magang dan Studi Independen Bersertifikat (MSIB) *Foundations of AI and Life Skills for Gen-Z* di Orbit Future Academy (OFA) 2022 dapat diselesaikan dengan baik. Tidak lupa ucapan terima kasih dihaturkan kepada:

1. Orang tua yang selalu mendukung jalannya program ini dari awal hingga akhir.
2. OFA selaku instansi pada program MSIB beserta para coach yang lain dan juga staff terkait.
3. Coach Angel Metanosa Afinda selaku *homeroom coach* dari kelas Jupyter XXI dan juga Coach Ipin Sugiyarto selaku *homeroom coach* kelas Cordoba di OFA.
4. Coach Fauzan selaku coach pembimbing domain *Artificial Intelligence Computer Vision* di OFA.
5. Teman-teman tim yang sudah berjuang untuk menyelesaikan penelitian ini.

Laporan ini disusun sebagai syarat laporan akhir kegiatan MSIB yang diikuti oleh penulis. Oleh karena itu, penulis memohon maaf apabila terdapat kesalahan dan kekurangan dalam pembuatan laporan. Akhir kata, penulis mengharapkan semoga laporan ini memberikan manfaat maupun inspirasi kepada pembacanya.

Penulis,



Nyayu Chika Marselina

## Daftar Isi

Bab I	Pendahuluan	1
I.1	Latar belakang	1
I.2	Lingkup	2
I.3	Tujuan	2
Bab II	Orbit Future Academy	1
II.1	Struktur Organisasi	1
II.2	Lingkup Pekerjaan	2
II.3	Deskripsi Pekerjaan	3
II.4	Jadwal Kerja	6
Bab III	Perbandingan <i>Convolutional Neural Network</i> dan <i>Vision Transformer</i> Pada Klasifikasi Kue Tradisional Indonesia	1
III.1	Latar Belakang Proyek Akhir	1
III.2	Proses Pelaksanaan Proyek Akhir	3
III.3	Hasil Proyek Akhir	5
Bab IV	Penutup	1
IV.1	Kesimpulan	1
IV.2	Saran	1
Bab V	Referensi	vii
Bab VI	Lampiran A. TOR	1
Bab VII	Lampiran B. Log Activity	1
Bab VIII	Lampiran C. Jurnal	1

## Daftar Tabel

Tabel 2.1 Perbandingan nilai <i>hyperparameter</i> Arsitektur A dan Arsitektur B ViT	4
Tabel 2.2 Hasil Perbandingan Nilai Akurasi, Presisi, <i>Recall</i> , dan <i>F1-Score</i> Kue Tradisional Indonesia	5
Tabel 2.3 Agenda Kelas	6
Tabel 3.1 Contoh Hasil Proses Augmentasi Data Citra Kue Tradisional Indonesia	5
Tabel 3.2 Hasil Akurasi, Presisi, <i>Recall</i> , dan <i>F1-Score</i> Kue Tradisional Indonesia Arsitektur VGG16	7
Tabel 3.3 Hasil Akurasi, Presisi, <i>Recall</i> , dan <i>F1-Score</i> Kue Tradisional Indonesia Arsitektur ResNet50	8
Tabel 3.4 Hasil Akurasi, Presisi, <i>Recall</i> , dan <i>F1-Score</i> Kue Tradisional Indonesia Arsitektur A ViT	10
Tabel 3.5 Hasil Akurasi, Presisi, <i>Recall</i> , dan <i>F1-Score</i> Kue Tradisional Indonesia Arsitektur B ViT	11
Tabel 3.6 Perbandingan Hasil Akurasi, Presisi, <i>Recall</i> , dan <i>F1-Score</i> Model Kue Tradisional Indonesia	12

## Daftar Gambar

Gambar 2.1 Logo Orbit Future Academy	1
Gambar 2.2 Struktur Organisasi OFA	2
Gambar 2.3 Grafik Akurasi dan <i>Loss</i> Arsitektur A ViT	4
Gambar 2.4 Grafik Akurasi dan <i>Loss</i> Arsitektur B ViT	5
Gambar 2.5 Perbandingan Akurasi, Presisi, <i>Recall</i> , dan <i>F1-Score</i> Kue Tradisional Indonesia	6
Gambar 3.1 Rangkaian Proses Proyek Akhir Kue Tradisional Indonesia	3
Gambar 3.2 Grafik Akurasi dan <i>Loss</i> Setiap <i>Epoch</i> VGG16	6
Gambar 3.3 Grafik Akurasi dan <i>Loss</i> Setiap <i>Epoch</i> ResNet50	8
Gambar 3.4 Grafik Akurasi dan <i>Loss</i> Setiap <i>Epoch</i> Arsitektur A ViT	9
Gambar 3.5 Grafik Akurasi dan <i>Loss</i> Setiap <i>Epoch</i> Arsitektur B ViT	11
Gambar 3.6 Perbandingan Hasil Akurasi, Presisi, <i>Recall</i> , dan <i>F1-Score</i> Model Kue Tradisional Indonesia	13



## **Bab I Pendahuluan**

### **I.1 Latar belakang**

Pada saat ini dan di masa depan, wawasan serta keterampilan *Artificial Intelligence* merupakan kompetensi masa depan yang mutlak harus dikuasai oleh sebanyak mungkin generasi penerus bangsa Indonesia. Era industri 4.0 menuntut sumber daya manusia yang menguasai komunikasi dengan mesin dan rekayasa otomatisasi, demi mengoptimalkan sinergi konstruktif antara manusia dengan mesin beserta perangkat lunaknya dan juga menuntut sumber daya manusia yang menguasai komunikasi multi-dimensi dengan manusia lain, berbagai perangkat, dan sistem-sistem pendukung lain sebagai bagian dari kemampuan pemecahan masalah yang semakin kompleks di masa depan.

Di Indonesia hingga tahun 2025, terdapat kebutuhan tenaga kerja sebanyak lebih dari 40 juta orang untuk menempati berbagai posisi dan fungsi pekerjaan masa depan yang sebelumnya tidak ada di masa lalu lintas industri. Mayoritas dari posisi dan fungsi pekerjaan tersebut sangat erat berkaitan dengan penguasaan teknologi dan rekayasa *Artificial Intelligence* (AI). Inilah sebabnya mengapa secara global, kebutuhan akan tenaga ahli Teknik Informatika dan praktisi AI terus meningkat tajam. Terdapat banyak kebutuhan talenta teknologi yang belum dapat dipenuhi oleh sistem pendidikan formal.

Orbit Future Academy (OFA) lahir untuk menjembatani dan memecahkan masalah tersebut, dengan bersinergi bersama dunia akademik formal untuk menyediakan paket-paket pembelajaran berwawasan teknologi maupun rekayasa secara intensif dan mendalam, bersama program Kampus Merdeka. Berkat dari dukungan dari para mitra teknologi kelas dunia serta Kementrian Pendidikan, Kebudayaan, Riset, dan Teknologi, OFA percaya bahwa peningkatan kualitas dan perluasan dampak program Orbit Jago AI dapat membawa transformasi signifikan bagi dunia pendidikan, dunia industri, maupun bagi kehidupan sehari-hari masyarakat Indonesia pada saat ini dan di masa depan. *Foundation of AI and Life Skills for Gen-Z* adalah program pelatihan *Artificial Intelligence* dari OFA secara *online* untuk peserta program Magang dan Studi Independen Bersertifikat (MSIB)

yang bertujuan bukan hanya untuk memperkenalkan teknologi AI ke pelajar, tapi juga untuk memungkinkan mereka bisa mengangkat perangkat AI, sehingga bisa membuat sesuatu produk yang menciptakan dampak sosial. Dalam hal ini, OFA berfokus pada beberapa domain AI, seperti *Data Science*, *Natural Language Processing*, dan *Computer Vision*. Selain keterampilan AI, mempelajari keterampilan hidup yang bisa dipergunakan untuk mencari kerja atau menciptakan kerja. Pada akhir program, peserta MSIB diarahkan untuk membuat Proyek Akhir (PA) dengan menggunakan domain AI yang telah dipelajari selama proses pembelajaran. Proses pembuatan PA dilakukan secara berkelompok. PA yang dibuat memiliki dua pilihan, yakni berbasis aplikasi atau jurnal. Dalam hal ini, PA yang dibuat adalah klasifikasi kue tradisional Indonesia. Dataset yang digunakan berupa data gambar sehingga PA menggunakan domain AI *Computer Vision*. Adapun algoritma yang digunakan dalam PA, yakni *Convolutional Neural Network* (CNN) dan *Vision Transformer* (ViT). Pada algoritma CNN digunakan arsitektur VGG16 dan ResNet50. Pada algoritma ViT akan dibuat dibuat dua buah arsitektur dengan membedakan nilai *hyperparameter* yang digunakan pada kedua arsitektur. Kedua algoritma yang digunakan akan dibandingkan satu sama lain dan dilihat arsitektur yang lebih baik dalam melakukan klasifikasi kue tradisional Indonesia.

## **I.2 Lingkup**

MSIB berfokus pada pembelajaran mahasiswa yang menghasilkan lulusan sesuai perkembangan zaman dan mengikuti kemajuan Ilmu Pengetahuan dan Teknologi (IPTEK). Hal tersebut relevan dalam pengerjaan PA team yang berfokus pada teknologi pendidikan dan mengikuti kemajuan IPTEK mengenai kecerdasan buatan dan implementasinya melalui PA.

## **I.3 Tujuan**

Tujuan yang diharapkan setelah mengikuti program ini:

1. Memiliki wawasan tentang AI dan perkembangannya.

2. Mampu menggunakan bahasa pemrograman Python untuk mengembangkan aplikasi AI.
3. Mampu merancang dan mengimplementasikan AI Project Cycle melalui mini project dan PA.
4. Mampu mengembangkan salah satu dari 3 domain AI (*Data Science*, *Natural Language Processing*, dan *Computer Vision*) hingga tahap *deployment* yang diimplementasikan melalui PA.
5. Mengetahui hal-hal yang dibutuhkan seorang wirausahawan yang bergerak di bidang *start-up* dalam mentransformasikan ide ke dalam bentuk produk atau jasa.

## Bab II Orbit Future Academy

### II.1 Struktur Organisasi



Gambar 2.1 Logo Orbit Future Academy

Orbit Future Academy (OFA) didirikan pada tahun 2016 dengan tujuan untuk meningkatkan kualitas hidup melalui inovasi, edukasi, dan pelatihan keterampilan. Label atau *brand* Orbit merupakan kelanjutan dari warisan mendiang Prof. Dr. Ing. B. J. Habibie (presiden Republik Indonesia ke-3) dan istrinya, Dr. Hasri Ainun Habibie. Mereka berdua telah menjadi penggerak dalam mendukung perkembangan inovasi dan teknologi pendidikan di Indonesia. OFA mengkurasi dan melokalkan program/kursus internasional untuk *upskilling* atau *reskilling* pemuda dan tenaga kerja menuju pekerjaan masa depan. Hal ini sesuai dengan slogan OFA, yakni “*Skills-for-Future-Jobs*”.

#### **Visi:**

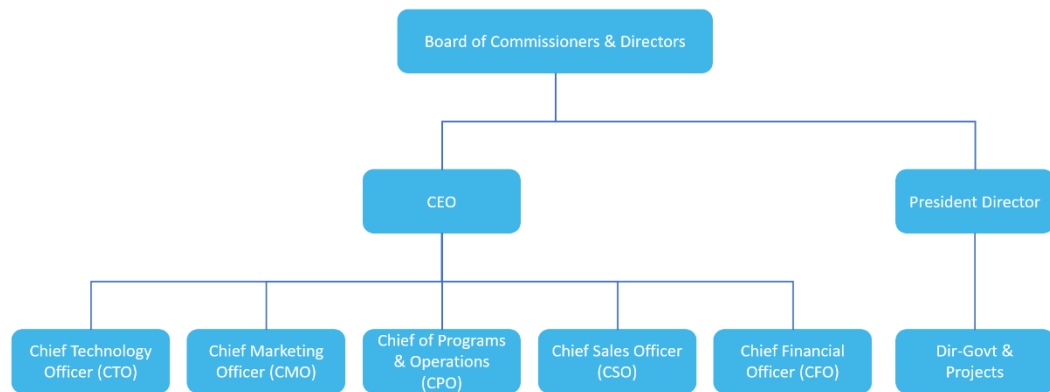
Memberikan pembelajaran berbasis keterampilan transformatif terbaik untuk para pencari kerja & pencipta lapangan kerja.

#### **Misi:**

1. Membangun jaringan Orbit Transformation Center (OTC) secara nasional untuk menyampaikan kurikulum keterampilan masa depan berbasis sertifikasi melalui Platform Konten Digital.

2. Secara proaktif bekerja dengan pemerintah & organisasi dengan mengubah tenaga kerja mereka agar sesuai dengan perubahan pekerjaan yang terjadi karena Industri 4.0.
3. Melatih pemuda dengan keterampilan kewirausahaan & mencocokkan mereka dengan peluang masa depan yang muncul di berbagai industri.
4. Menghubungkan jaringan inkubator dan akselerator yang dikurasi ke industri, investor, dan ekosistem start-up global.

Struktur organisasi OFA dapat dilihat pada Gambar 2.2.



Gambar 2.2 Struktur Organisasi OFA

## II.2 Lingkup Pekerjaan

Seorang fasilitator akan mendampingi kurang lebih 40 peserta MSIB (student) dalam satu kelas. Terdapat dua jenis fasilitator, yakni:

### a. Homeroom Coach

Homeroom coach bertugas menyampaikan materi tentang dasar-dasar AI, memberikan penilaian pada student, dan mengarahkan *student* saat pengerjaan Proyek Akhir (PA).

### **b. Domain Coach**

Domain coach bertugas menyampaikan materi tentang domain AI atau life skills dan memberikan penilaian pada student.

Lingkup pekerjaan student adalah mengikuti kelas bersama homeroom atau domain coach, sesuai agenda kelas, hingga program selesai.

## **II.3 Deskripsi Pekerjaan**

Berikut adalah deskripsi pekerjaan student sebelum pengerjaan PA:

- a. Mengikuti pre-test.
- b. Mengikuti kelas sesi pagi pada pukul 08.00 hingga 11.30 WIB.
- c. Mengikuti kelas sesi siang pada pukul 13.00 hingga 16.30 WIB.
- d. Mengulang materi yang telah disampaikan di kelas sesi pagi dan siang, setelah kelas sesi siang, selama 1 jam (*self-study*).
- e. Mengerjakan latihan individu atau kelompok yang diberikan oleh homeroom atau domain coach saat kelas berlangsung.
- f. Mengerjakan tugas yang diberikan homeroom atau domain coach hingga batas waktu tertentu.
- g. Mengerjakan *mini project* yang diberikan homeroom atau domain coach hingga batas waktu tertentu
- h. Mengikuti post-test.

Student memiliki peran menentukan domain AI PA, menentukan algoritma yang digunakan, menentukan arsitektur yang digunakan pada masing-masing algoritma yang dipilih, membuat dan mengembangkan model kue tradisional Indonesia, melakukan *deployment* hasil model dengan flask, dan mengusulkan perubahan basis PA. Selama pengerjaan PA, dengan deskripsi pekerjaan sebagai berikut:

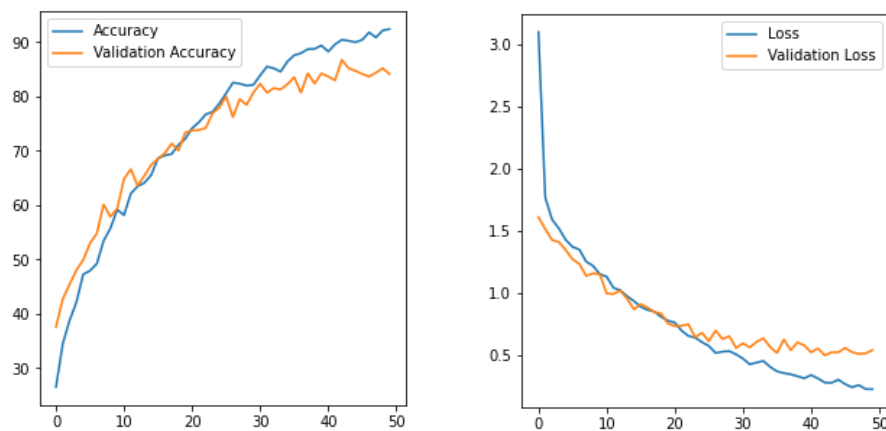
- a. Memilih domain AI yang digunakan dalam mengolah data kue tradisional Indonesia. Dalam hal ini, dipilih domain AI CV karena CV mampu mengolah data citra.
- b. Menentukan algoritma yang akan digunakan dalam melakukan klasifikasi kue tradisional Indonesia. Algoritma yang dipilih adalah CNN dan ViT.

- c. Membentuk dua buah arsitektur ViT dengan membedakan nilai *hyperparameter* pada masing-masing arsitektur ViT. Perbedaan nilai *hyperparameter* dapat dilihat pada Tabel 2.1.

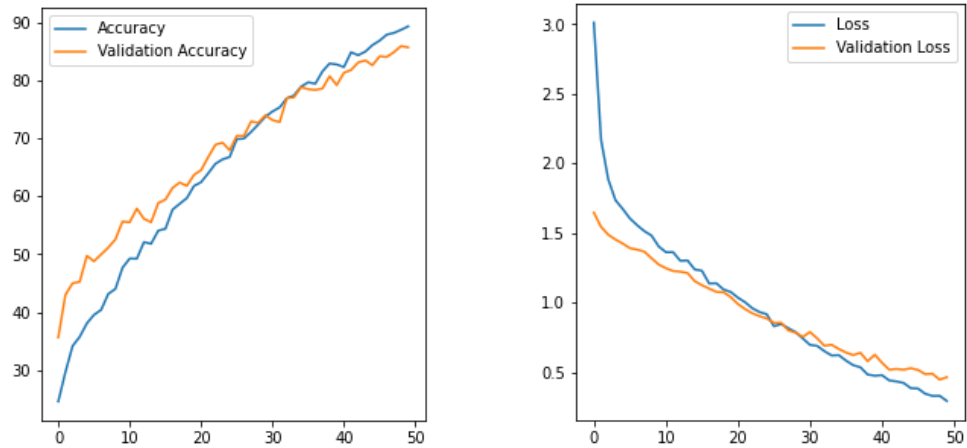
Tabel 2.1 Perbandingan nilai *hyperparameter* Arsitektur A dan Arsitektur B ViT

<i>Hyperparameter</i>	Arsitektur A	Arsitektur B
learning_rate	0,001	0,0001
weight_decay	0,0001	0,00001
batch_size	32	64
image_size	72	144
patch_size	6	12
projection_dim	64	128
num_heads	4	8

- d. Untuk evaluasi model digunakan *confusion matrix* dimana parameter yang digunakan adalah akurasi, presisi, *recall* dan *f1-score*. *Metrics* yang digunakan untuk melakukan proses *training* data adalah akurasi. Adapun grafik *epoch* akurasi dan *loss* dari masing-masing arsitektur ViT, seperti pada Gambar 2.3 dan Gambar 2.4.



Gambar 2.3 Grafik Akurasi dan *Loss* Arsitektur A ViT



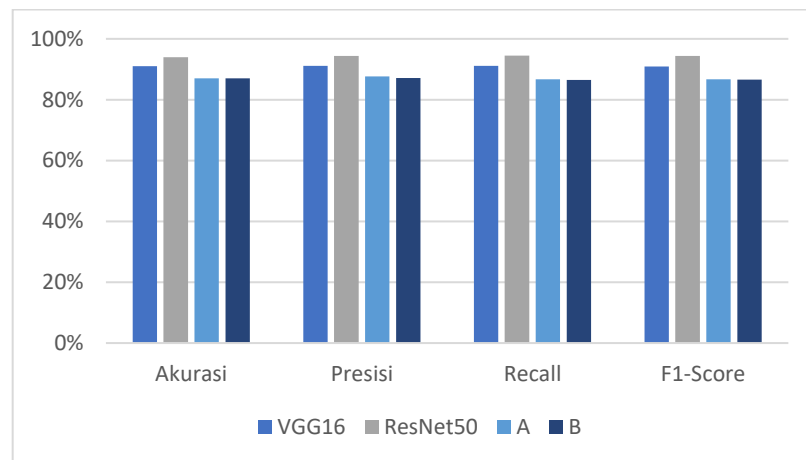
Gambar 2.4 Grafik Akurasi dan *Loss* Arsitektur B ViT

- e. Mencoba melakukan *deployment*, tetapi saat melakukan *connect* antara hasil model dengan web yang dibuat, hasil model tidak terbaca.
- f. Mengusulkan perubahan basis PA yang semula berbasis aplikasi web menjadi berbasis penelitian dikarenakan keterbatasan waktu.
- g. Membuat perbandingan keempat arsitektur yang telah dibuat untuk melihat arsitektur yang memiliki nilai akurasi, presisi, *recall*, dan *f1-score* terbaik di antara arsitektur lainnya. Hasil perbandingan keempat arsitektur dapat dilihat pada Tabel 2.2 dan Gambar 2.5.

Tabel 2.2 Hasil Perbandingan Nilai Akurasi, Presisi, *Recall*, dan *F1-Score* Kue Tradisional Indonesia

Algoritma	Arsitektur	Akurasi	Presisi	<i>Recall</i>	<i>F1-Score</i>
CNN	VGG16	91%	91,13%	91,13%	90,88%
	<b>ResNet50</b>	<b>94%</b>	<b>94,38%</b>	<b>94,5%</b>	<b>94.38%</b>
ViT	A	87%	87,63%	86,75%	86,75%
	B	87%	87,13%	86,5%	86,63%





Gambar 2.5 Perbandingan Akurasi, Presisi, *Recall*, dan *F1-Score* Kue Tradisional Indonesia

## II.4 Jadwal Kerja

Program ini berlangsung setiap hari kerja (Senin sampai dengan Jumat) selama 8 jam per harinya, dengan rincian sebagai berikut:

Tabel 2.3 Agenda Kelas

Pukul (WIB)	Durasi (jam)	Aktivitas
08.00 s.d. 11.30	3.5	Kelas Sesi Pagi
13.00 s.d. 16.30	3.5	Kelas Sesi Siang
16.30 s.d. 17.30	1	<i>Self-Study</i>

Program ini berlangsung dari bulan Februari 2022 sampai dengan bulan Juli 2022.

### **Bab III      Perbandingan *Convolutional Neural Network* dan *Vision Transformer* Pada Klasifikasi Kue Tradisional Indonesia**

#### **III.1 Latar Belakang Proyek Akhir**

Kue tradisional Indonesia adalah salah satu bentuk dari ciri khas dan kebudayaan yang dimiliki oleh Indonesia. Terdapat banyak jenis-jenis kue tradisional Indonesia, yakni kue klepon, kue lumpur, kue kastengel, kue putri salju, kue serabi, kue dadar gulung, kue lapis, dan kue risoles [1]. Di era sekarang, kue tradisional mulai digantikan dengan kue modern. Kompetisi antara kue tradisional dengan kue modern sangat ketat apalagi kue-kue modern memiliki proses pembuatan yang lebih efisien dibandingkan kue tradisional. Tak hanya itu, terdapat banyak kue modern yang memiliki harga lebih murah sehingga kue tradisional semakin tertinggal dan minat masyarakat jauh berkurang [3]. Kue tradisional Indonesia dapat digunakan sebagai salah satu bentuk pengenalan ciri khas atau kebudayaan Indonesia pada masyarakat. Beberapa penelitian mengenai kue atau makanan tradisional Indonesia, yakni Kurnia et al, 2021 [1] yang menggunakan algoritma *Convolutional Neural Network* (CNN) dengan akurasi yang diperoleh sebesar 65%, tetapi tidak menampilkan nilai presisi, *recall*, dan *f1-score* pada penelitian yang dilakukan. Penelitian Sari et al, 2019 [4] melakukan identifikasi pada makanan tradisional Indonesia menggunakan *Random Forest* (RF). Hasil akurasi yang diperoleh adalah 59,7%. Pada penelitian tersebut tidak menampilkan nilai presisi, *recall*, dan *f1-score*. Penelitian Fahira et al, 2020 [5] melakukan klasifikasi pada makanan tradisional Jawa Indonesia menggunakan KNN. Hasil akurasi dan *f1-score* yang diperoleh adalah 89,17% dan 85,58%.

Salah satu cara untuk memperkenalkan kue tradisional Indonesia pada khalayak umum di masa modern, yakni dengan memanfaatkan *Artificial Intelligence* (AI) atau kecerdasan buatan [6]. Salah satu kelebihan dari AI adalah dapat menyebarkan informasi secara cepat dan mudah [7]. Domain AI yang dapat mengolah data citra adalah *Computer Vision* (CV). Salah satu tugas dari CV, yakni melakukan klasifikasi pada data citra [8]. Salah satu algoritma yang dapat melakukan klasifikasi data citra, yakni CNN [9]. Algoritma yang memiliki

performa yang baik dalam melakukan pelatihan dataset dalam jumlah besar adalah CNN [10]. Adapun arsitektur pada CNN adalah VGG16 dan ResNet50. VGG16 merupakan model yang menunjukkan bahwa jaringan yang berada di dalam merupakan faktor penting untuk kinerja yang baik. Keunggulan model ini adalah arsitektur sangat homogen, hanya melakukan akumulasi  $3 \times 3$  dan  $2 \times 2$  *end-to-end* [11]. VGG16 telah diterapkan di beberapa penelitian. Penelitian Gunawan et al, 2021 [12] melakukan pengenalan ekspresi wajah dengan dataset *Facial Expression Recognition* 2013 yang menghasilkan akurasi di bawah 70%. Pada penelitian tersebut tidak ditunjukkan nilai presisi, *recall*, dan *f1-score*. Adapun penelitian lain menggunakan VGG16 yang dilakukan Setiawan, 2020 [13] melakukan uji coba dengan akurasi tertinggi yang diperoleh sebesar 89,3%.

Arsitektur ResNet-50 sendiri merupakan jaringan residual yang terdiri dari 50 *layer*. ResNet50 bukan berfokus menyelidiki beberapa fitur, tetapi menyelidiki beberapa residual. Maksud dari residual adalah pengurangan fitur yang dipelajari berdasarkan masukan suatu *layer* [14]. Arsitektur ResNet50 telah diterapkan pada beberapa penelitian salah satunya, pada penelitian Niswati et al, 2021 [15] yang melakukan klasifikasi kanker serviks pada citra *pap smear*. Penelitian tersebut menghasilkan nilai akurasi sebesar 91%. Walaupun akurasi yang diperoleh di atas 90%, penelitian tersebut tidak menampilkan hasil dari presisi, *recall*, dan *f1-score*. Penelitian lainnya yang menggunakan ResNet50, yakni Nahar et al, 2018 [16] yang melakukan klasifikasi citra sidik jari dengan hasil akurasi yang diperoleh sebesar 90%, tetapi tidak menampilkan nilai presisi, *recall*, dan *f1-score* pada penelitian yang dilakukan.

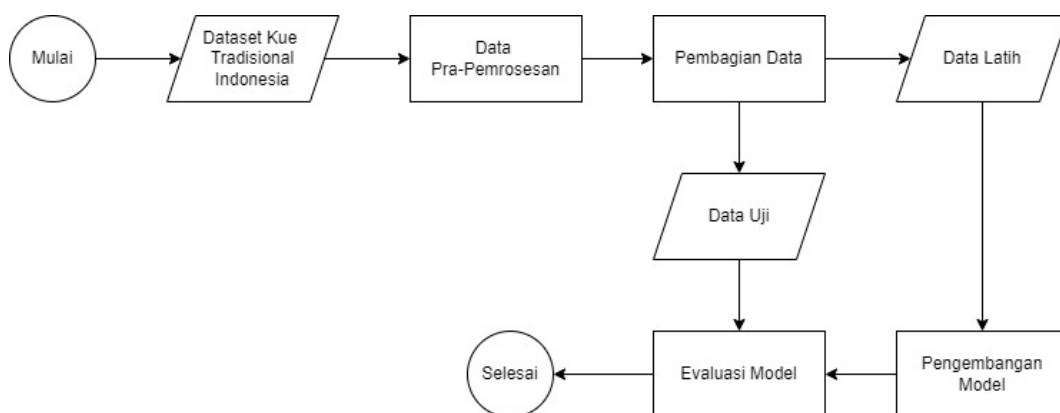
Selain CNN, algoritma yang dapat digunakan untuk klasifikasi data citra adalah *Vision Transformer* (ViT) yang memiliki ukuran model lebih efisien dibandingkan konvolusi konvensional dan dapat memprediksi fitur dengan resolusi yang tinggi dalam waktu yang lebih singkat [17]. Sebelumnya, ViT hanya digunakan dalam *Natural Language Processing* (NLP), namun saat ini ViT menjadi salah satu *transformer* yang dimanfaatkan pada pengolahan citra. Algoritma ViT telah digunakan pada beberapa penelitian salah satunya penelitian yang dilakukan oleh Chen et al, 2021 [18] melakukan klasifikasi menggunakan dataset

ImageNet1K yang menghasilkan akurasi sebesar 82%, tetapi tidak menampilkan hasil presisi, *recall*, dan *f1-score*. Penelitian Kamran et al, 2021 [19] mengklasifikasikan citra retina menggunakan ViT dengan akurasi, presisi, dan *recall* yang diperoleh sebesar 85,7%, 90%, dan 83,3%, tetapi tidak menampilkan nilai *f1-score* pada penelitian yang dilakukan. Penelitian yang dilakukan oleh Gheflati dan Rivaz, 2021 [20] yang melakukan klasifikasi pada citra medis *ultrasound* payudara menghasilkan akurasi sebesar 86%, tetapi tidak menampilkan hasil presisi, *recall*, dan *f1-score*.

Melihat penelitian-penelitian sebelumnya, hasil yang diperoleh belum maksimal sehingga pada PA ini dilakukan klasifikasi kue tradisional Indonesia menggunakan algoritma CNN dan ViT. Pada PA, arsitektur CNN yang digunakan adalah VGG16 dan ResNet50. Algoritma ViT yang digunakan akan dibagi menjadi dua arsitektur yang dibentuk dengan perombakan arsitektur yang mengubah nilai-nilai *hyperparameter* tertentu. Parameter yang akan digunakan untuk melakukan evaluasi model yang dibuat adalah akurasi, presisi, *recall*, dan *f1-score*. Kedua algoritma yang digunakan akan dibandingkan untuk menentukan algoritma yang lebih baik dalam melakukan klasifikasi citra kue tradisional Indonesia.

### III.2 Proses Pelaksanaan Proyek Akhir

Rangkaian proses pengerjaan model PA yang dibuat dapat dilihat pada Gambar 3.1.



Gambar 3.1 Rangkaian Proses Proyek Akhir Kue Tradisional Indonesia

Dari Gambar 3.1 proses yang dilakukan dalam pengerjaan model AI adalah pencarian dataset kue tradisional Indonesia, data prapemrosesan, pembagian data menjadi data uji dan data latih dimana data latih akan digunakan sebagai data dalam pengembangan model yang kemudian akan dievaluasi melalui model yang dibuat dan data uji akan digunakan langsung pada tahap evaluasi model. Dataset yang digunakan adalah dataset kue tradisional Indonesia berupa data citra yang bersumber dari Kaggle dengan total data sebanyak 1845 data dimana terdapat 240 data pada kelas kue klepon, 238 data pada kelas kue lumpur, 223 data pada kelas kue kastengel, 214 data pada kelas kue putri salju, 221 data pada kelas kue serabi, 232 data pada kelas kue dadar gulung, 241 data pada kelas kue lapis, dan 236 data pada kelas kue risoles.

Data yang digunakan akan di augmentasi sehingga menghasilkan sebanyak 5277 data kue tradisional Indonesia. Dalam pengembangan model, digunakan algoritma CNN dan ViT dimana masing-masing algoritma akan dibentuk dua buah arsitektur. Pada algoritma CNN, arsitektur yang digunakan adalah VGG16 dan ResNet50. Pada algoritma ViT, arsitektur dibuat dengan membedakan beberapa nilai dari *hyperparameter* yang digunakan pada masing-masing arsitektur. *Hyperparameter* yang digunakan sebagai pembeda dalam arsitektur ViT adalah *learning rate*, *weight decay*, *batch size*, *image size*, *patch size*, *projection dimension*, dan *num heads*.

Adapun hambatan penyelesaian PA menurut pandangan tim adalah pada saat persiapan penentuan fiksasi ide project yang akan kami kerjakan nantinya dan penentuan basis PA yang dipilih antara aplikasi web atau jurnal. Karena minimnya referensi membuat tim bingung untuk memutuskan proyek apa yang nantinya akan dijalankan. Tim sudah berdiskusi dan mengambil keputusan, tapi akhirnya terdapat beberapa kelompok yang sama dengan ide PA kami dan memutuskan untuk mengubah ide awal tersebut. Ketika berjalan pengerjaan PA kami juga kesulitan membagi waktu untuk diskusi, mengingat setiap anggota tim berasal dari instansi yang berbeda dan pastinya memiliki jadwal yang berbeda, tetapi itu semua dapat disiasati dengan cara setiap anggota tim menyempatkan waktu untuk diskusi meskipun di malam hari. Dalam suatu hambatan atau masalah yang dihadapi suatu

tim dalam mengerjakan PA, pastinya ada solusi yang dapat membantu meringankan hambatan tersebut.

Pertama, pastinya harus menyadari bahwa setiap anggota harus membangun kepercayaan antar sesama dan pastinya memberikan apresiasi untuk performa yang sudah baik serta *feedback* spesifik dan kritik yang membangun untuk performa yang perlu diperbaiki. Selain itu, mengenali gaya perilaku dan kepribadian setiap anggota tim, karena sangat penting agar pengerjaan akan terlaksana dengan baik jika saling menghargai dan memahami satu sama lain. Kedua, berani untuk menghadapi suatu konflik saat pengerjaan PA. Tentunya, menyadari bahwa tujuan dari konflik adalah untuk menghasilkan solusi terbaik dalam waktu sesingkat mungkin, maka harus berani berdiskusi sehingga menyelesaikan masalah dengan cepat dan menyeluruh. Ketiga, menumbuhkan komitmen dalam tim. Komitmen akan terwujud ketika ada kejelasan dan penerimaan terhadap perencanaan dan keputusan tim. Salah satu cara agar memaksimalkan sebuah komitmen adalah membuat *deadline* yang jelas mengenai waktu yang telah ditetapkan untuk menjalankan rencana tim.

### III.3 Hasil Proyek Akhir

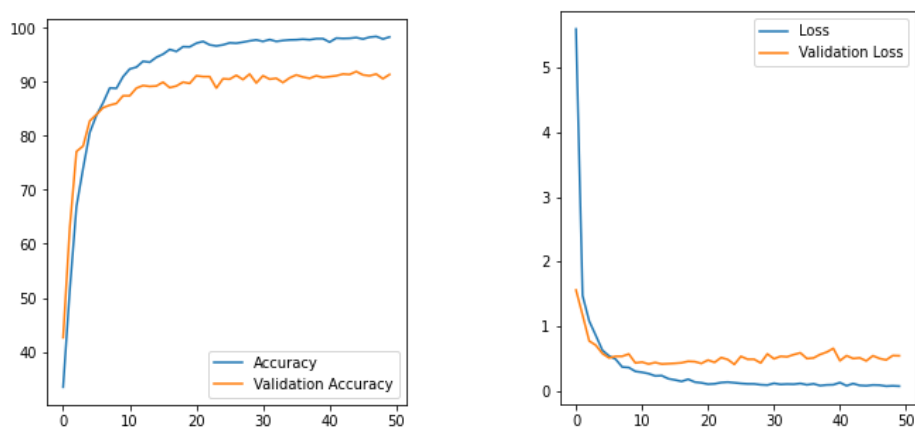
Pada penelitian ini, dataset kue tradisional Indonesia memiliki 8 kelas, yakni kue klepon, kue lumpur, kue kastengel, kue putri salju, kue serabi, kue dadar gulung, kue lapis, dan kue risoles. Dilakukan perubahan ukuran data citra pada dataset kue tradisional Indonesia menjadi  $256 \times 256$ . Setelah itu, melakukan augmentasi data dengan menggunakan *horizontal flip* dan *vertikal flip*. Hasil dari proses augmentasi data dapat dilihat pada Tabel 3.1.

Tabel 3.1 Contoh Hasil Proses Augmentasi Data Citra Kue Tradisional Indonesia

			
Kue Klepon	Kue Lumpur	Kue Kastengel	Kue Putri Salju

			
Kue Serabi	Kue Dadar Gulung	Kue Lapis	Kue Risoles

Pada Tabel 3.1 dapat dilihat, bahwa terdapat 8 contoh data citra yang diperoleh dari proses augmentasi data. Dengan augmentasi, dataset kue tradisional meningkat menjadi 5227 data citra. Setelah melakukan augmentasi data, melakukan split data pada dataset kue tradisional Indonesia dengan membagi 5277 data citra ke dalam 80% *data training* dan 20% *data testing* sehingga terdapat 4222 *data training* dan 1055 *data testing*. Dilakukan proses pelatihan data menggunakan algoritma CNN dan ViT. Pada algoritma CNN, digunakan arsitektur VGG16 dan arsitektur ResNet50, sedangkan pada algoritma ViT dilakukan perombakan arsitektur sehingga menghasilkan dua arsitektur, yakni arsitektur A dan arsitektur B. Pada algoritma CNN digunakan optimasi dan fungsi *loss* untuk mengestimasi nilai *loss*. Fungsi *loss* yang digunakan adalah *categorical cross entropy*. Dari proses *training* data, diperoleh nilai akurasi dan *loss* pada *data training* serta nilai akurasi dan *loss* pada data validasi. Grafik akurasi dan *loss* yang diperoleh pada *data training* menggunakan arsitektur VGG16 dapat dilihat pada Gambar 3.2.



Gambar 3.2 Grafik Akurasi dan *Loss* Setiap *Epoch* VGG16

Berdasarkan Gambar 3.2 dapat dilihat, bahwa akurasi yang diperoleh mengalami peningkatan pada setiap *epoch* dan nilai *loss* mengalami penurunan pada setiap *epoch*. Kenaikan akurasi secara signifikan terjadi pada *epoch* 4 ke *epoch* 5. Akurasi stabil di kisaran 90% terdapat pada *epoch* 37 dan seterusnya. Pada grafik *loss*, penurunan nilai *loss* secara signifikan juga terjadi pada *epoch* 4 ke *epoch* 5. Untuk mengevaluasi performa klasifikasi model kue tradisional Indonesia dari arsitektur VGG16, digunakan parameter akurasi, presisi, *recall*, dan *f1-score*. Hasil akurasi, presisi, *recall*, dan *f1-score* dari klasifikasi menggunakan arsitektur VGG16 dapat dilihat pada Tabel 3.2.

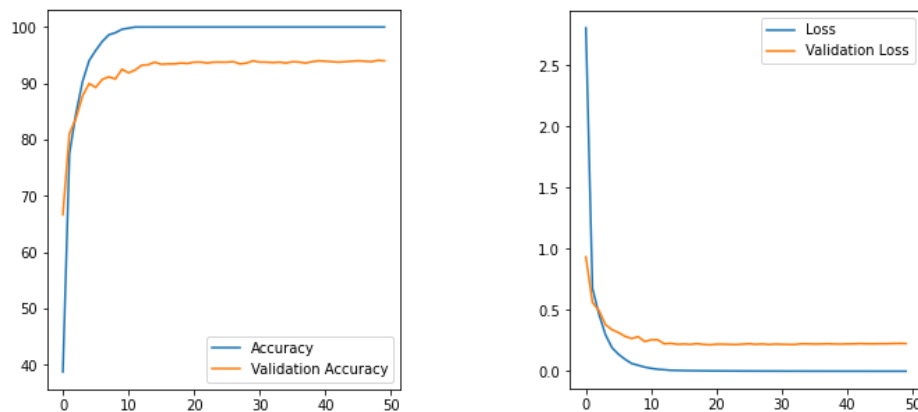
Tabel 3.2 Hasil Akurasi, Presisi, *Recall*, dan *F1-Score* Kue Tradisional Indonesia Arsitektur VGG16

Nama Kelas	Akurasi	Presisi	<i>Recall</i>	<i>F1-Score</i>
Kue Dadar Gulung	91%	95%	82%	88%
Kue Kastengel		88%	97%	92%
Kue Klepon		98%	96%	97%
Kue Lapis		91%	96%	93%
Kue Lumpur		89%	91%	90%
Kue Putri Salju		93%	88%	90%
Kue Risoles		91%	91%	91%
Kue Serabi		84%	88%	86%
<b>Rata-Rata</b>	<b>91%</b>	<b>91,13%</b>	<b>91,13%</b>	<b>90,88%</b>

Berdasarkan Tabel 3.2, hasil rata-rata akurasi, presisi, *recall*, dan *f1-score* dari klasifikasi menggunakan arsitektur VGG16 adalah 91%, 91,13%, 91,13%, dan 90,88% sehingga klasifikasi kue tradisional Indonesia pada model yang dibuat memiliki performa yang baik. Nilai presisi tertinggi terdapat pada kelas kue klepon, yakni sebesar 98%. Nilai *recall* tertinggi terdapat pada kelas kue kastengel, yakni 97%. Nilai *f1-score* tertinggi juga terdapat pada kelas kue klepon, yakni sebesar 97%. Pada algoritma CNN digunakan optimasi dan fungsi *loss* untuk mengestimasi



nilai *loss*. Fungsi *loss* yang digunakan adalah *categorical cross entropy*. Dari proses *training* data, diperoleh nilai akurasi dan *loss* pada *data training* serta nilai akurasi dan *loss* pada data validasi. Grafik akurasi dan *loss* yang diperoleh pada *data training* dapat dilihat pada Gambar 3.3.



Gambar 3.3 Grafik Akurasi dan *Loss* Setiap *Epoch* ResNet50

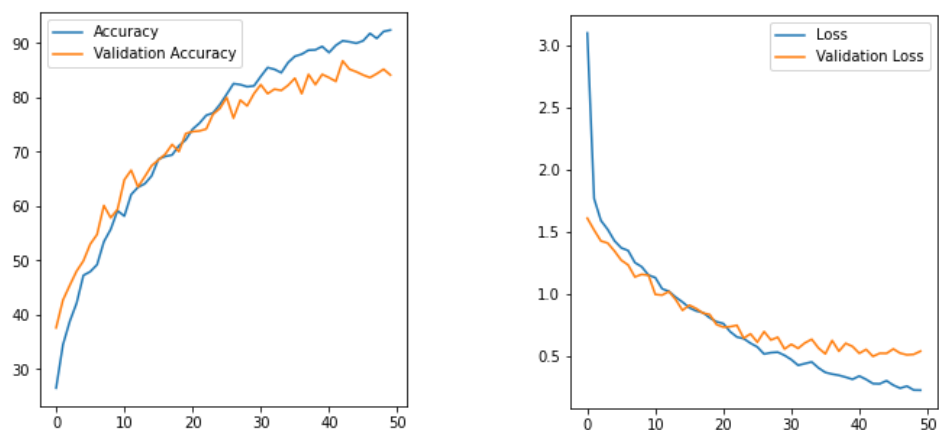
Berdasarkan Gambar 3.3 dapat dilihat, bahwa akurasi yang diperoleh mengalami peningkatan pada setiap *epoch* dan nilai *loss* mengalami penurunan pada setiap *epoch*. Kenaikan akurasi secara signifikan terjadi pada *epoch* 3 ke *epoch* 4. Akurasi stabil di kisaran 93% terdapat pada *epoch* 15 dan seterusnya. Pada grafik *loss*, penurunan nilai *loss* secara signifikan juga terjadi pada *epoch* 3 ke *epoch* 4. Untuk mengevaluasi performa klasifikasi model kue tradisional Indonesia dari arsitektur ResNet50, digunakan parameter akurasi, presisi, *recall*, dan *f1-score*. Hasil akurasi, presisi, *recall*, dan *f1-score* dari klasifikasi menggunakan arsitektur ResNet50 dapat dilihat pada Tabel 3.3.

Tabel 3.3 Hasil Akurasi, Presisi, *Recall*, dan *F1-Score* Kue Tradisional Indonesia Arsitektur ResNet50

Nama Kelas	Akurasi	Presisi	<i>Recall</i>	<i>F1-Score</i>
Kue Dadar Gulung	94%	92%	92%	92%
Kue Kastengel		95%	96%	96%

Kue Klepon		96%	99%	97%
Kue Lapis		96%	96%	96%
Kue Lumpur		95%	92%	94%
Kue Putri Salju		95%	94%	94%
Kue Risoles		93%	96%	94%
Kue Serabi		93%	91%	92%
<b>Rata-Rata</b>	<b>94%</b>	<b>94,38%</b>	<b>94,5%</b>	<b>94,38%</b>

Berdasarkan Tabel 3.3, hasil rata-rata akurasi, presisi, *recall*, dan *f1-score* dari klasifikasi menggunakan arsitektur ResNet50 adalah 94%, 94,38%, 94,5%, dan 94,38% sehingga klasifikasi kue tradisional Indonesia pada model yang dibuat memiliki performa yang baik. Nilai presisi tertinggi terdapat pada kelas kue klepon dan kue lapis yakni sebesar 96%. Nilai *recall* tertinggi terdapat pada kelas kue klepon, yakni 99%. Nilai *f1-score* tertinggi juga terdapat pada kelas kue klepon, yakni sebesar 97%. Pada algoritma ViT arsitektur A, digunakan optimasi dan fungsi *loss* untuk mengestimasi nilai *loss*. Fungsi *loss* yang digunakan adalah *sparse categorical cross entropy*. Dari proses *training* data, diperoleh nilai akurasi dan *loss* pada *data training* serta nilai akurasi dan *loss* pada data validasi. Grafik akurasi dan *loss* yang diperoleh pada *data training* arsitektur A algoritma ViT dapat dilihat pada Gambar 3.4.



Gambar 3.4 Grafik Akurasi dan *Loss* Setiap *Epoch* Arsitektur A ViT

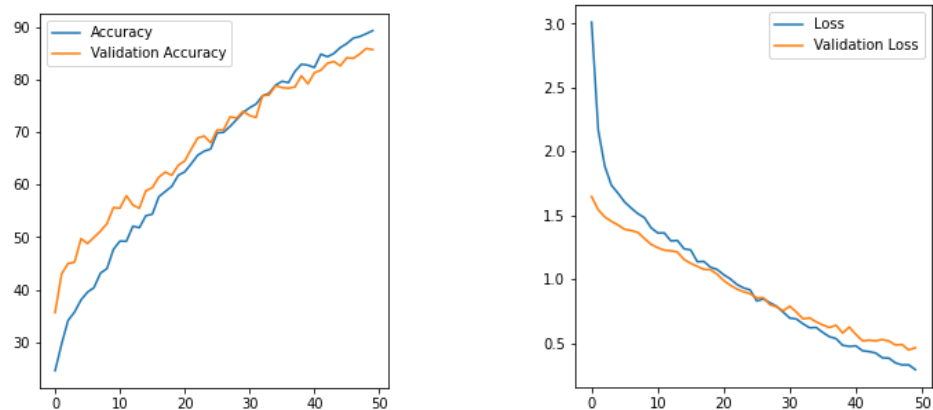
Berdasarkan Gambar 3.4 dapat dilihat, bahwa akurasi yang diperoleh mengalami peningkatan pada setiap *epoch* dan nilai *loss* mengalami penurunan pada setiap *epoch*. Kenaikan akurasi secara signifikan terjadi pada *epoch* 10 ke *epoch* 11. Akurasi stabil di kisaran 80% terdapat pada *epoch* 35 dan seterusnya. Pada grafik *loss*, penurunan nilai *loss* secara signifikan terjadi pada *epoch* 23 ke *epoch* 24. Untuk mengevaluasi performa klasifikasi model kue tradisional Indonesia dari algoritma ViT, digunakan parameter akurasi, presisi, *recall*, dan *f1-score*. Hasil akurasi, presisi, *recall*, dan *f1-score* dari klasifikasi menggunakan algoritma ViT arsitektur A dapat dilihat pada Tabel 3.4.

Tabel 3.4 Hasil Akurasi, Presisi, *Recall*, dan *F1-Score* Kue Tradisional Indonesia Arsitektur A ViT

Nama Kelas	Akurasi	Presisi	<i>Recall</i>	<i>F1-Score</i>
Kue Dadar Gulung	87%	92%	84%	88%
Kue Kastengel		93%	79%	85%
Kue Klepon		96%	89%	92%
Kue Lapis		88%	79%	83%
Kue Lumpur		80%	89%	84%
Kue Putri Salju		91%	95%	93%
Kue Risoles		79%	92%	85%
Kue Serabi		82%	87%	84%
<b>Rata-Rata</b>	<b>87%</b>	<b>87,63%</b>	<b>86,75%</b>	<b>86,75%</b>

Berdasarkan Tabel 3.4, hasil rata-rata akurasi, presisi, *recall*, dan *f1-score* dari klasifikasi menggunakan algoritma ViT arsitektur A adalah 87%, 87,63%, 86,75%, dan 86,75%. Nilai presisi tertinggi terdapat pada kelas kue klepon, yakni sebesar 96%. Nilai *recall* tertinggi terdapat pada kelas kue putri salju, yakni 95%. Nilai *f1-score* tertinggi juga terdapat pada kelas kue putri salju, yakni sebesar 93%. Pada algoritma ViT menggunakan arsitektur B, digunakan optimasi dan fungsi *loss*

untuk mengestimasi nilai *loss*. Fungsi *loss* yang digunakan adalah *sparse categorical cross entropy*. Dari proses *training* data, diperoleh nilai akurasi dan *loss* pada *data training* serta nilai akurasi dan *loss* pada data validasi. Grafik akurasi dan *loss* yang diperoleh pada *data training* dapat dilihat pada Gambar 3.5.



Gambar 3.5 Grafik Akurasi dan *Loss* Setiap *Epoch* Arsitektur B ViT

Berdasarkan Gambar 3.5 dapat dilihat, bahwa akurasi yang diperoleh mengalami peningkatan pada setiap *epoch* dan nilai *loss* mengalami penurunan pada setiap *epoch*. Kenaikan akurasi secara signifikan terjadi pada *epoch* 32 ke *epoch* 33. Akurasi stabil di kisaran 80% terdapat pada *epoch* 37 dan seterusnya. Pada grafik *loss*, penurunan nilai *loss* secara signifikan terjadi pada *epoch* 27 ke *epoch* 28. Untuk mengevaluasi performa klasifikasi model kue tradisional Indonesia dari algoritma ViT, digunakan parameter akurasi, presisi, *recall*, dan *f1-score*. Hasil akurasi, presisi, *recall*, dan *f1-score* dari klasifikasi menggunakan algoritma ViT arsitektur B dapat dilihat pada Tabel 3.5.

Tabel 3.5 Hasil Akurasi, Presisi, *Recall*, dan *F1-Score* Kue Tradisional Indonesia Arsitektur B ViT

Nama Kelas	Akurasi	Presisi	<i>Recall</i>	<i>F1-Score</i>
Kue Dadar Gulung	87%	89%	86%	87%
Kue Kastengel		85%	86%	86%

Kue Klepon		91%	91%	91%
Kue Lapis		92%	74%	82%
Kue Lumpur		82%	86%	84%
Kue Putri Salju		92%	95%	93%
Kue Risoles		87%	87%	87%
Kue Serabi		79%	87%	83%
<b>Rata-Rata</b>	<b>87%</b>	<b>87,13%</b>	<b>86,5%</b>	<b>86,63%</b>

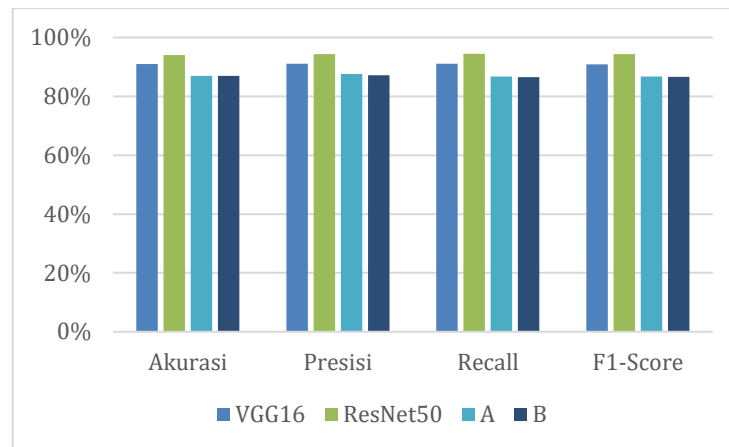
Berdasarkan Tabel 3.5, hasil rata-rata akurasi, presisi, *recall*, dan *f1-score* dari klasifikasi menggunakan algoritma ViT arsitektur B adalah 87%, 87,13%, 86,5%, dan 86,63% sehingga klasifikasi kue tradisional Indonesia pada model yang dibuat memiliki performa yang baik. Nilai presisi tertinggi terdapat pada kelas kue lapis dan kue putri salju, yakni sebesar 92%. Nilai *recall* tertinggi terdapat pada kelas kue putri salju, yakni 95%. Nilai *f1-score* tertinggi terdapat pada kelas kue klepon, yakni sebesar 91%. Perbandingan nilai akurasi, presisi, *recall*, dan *f1-score* yang diperoleh dari keempat model yang dibuat dapat dilihat pada Tabel 3.6.

Tabel 3.6 Perbandingan Hasil Akurasi, Presisi, *Recall*, dan *F1-Score* Model Kue Tradisional Indonesia

Algoritma	Arsitektur	Akurasi	Presisi	<i>Recall</i>	<i>F1-Score</i>
CNN	VGG16	91%	91,13%	91,13%	90,88%
	<b>ResNet50</b>	<b>94%</b>	<b>94,38%</b>	<b>94,5%</b>	<b>94,38%</b>
ViT	A	87%	87,63%	86,75%	86,75%
	B	87%	87,13%	86,5%	86,63%

Berdasarkan Tabel 3.6 dapat dilihat bahwa nilai akurasi presisi, *recall*, dan *f1-score* tertinggi terdapat pada model yang menggunakan arsitektur ResNet50, sedangkan nilai akurasi presisi, *recall*, dan *f1-score* terendah diperoleh oleh model

yang menggunakan algoritma ViT. Hasil akurasi, presisi, *recall*, dan *f1-score* dari VGG16, ResNet50, dan ViT dapat dilihat secara ringkas pada Gambar 3.6.



Gambar 3.6 Perbandingan Hasil Akurasi, Presisi, *Recall*, dan *F1-Score* Model Kue Tradisional Indonesia

Dari Gambar 3.6 dapat dilihat bahwa nilai akurasi presisi, *recall*, dan *f1-score* yang diperoleh berada di atas 80% sehingga model yang dibuat menggunakan VGG16, ResNet50, dan ViT memiliki performa yang baik dalam melakukan klasifikasi kue tradisional Indonesia. Jika keempat model yang dibuat tersebut dibandingkan, maka dapat diketahui bahwa model arsitektur ResNet50 memiliki performa yang terbaik di antara model lainnya sehingga ResNet50 lebih baik dalam melakukan klasifikasi kue tradisional Indonesia dibandingkan keempat model lainnya.

Kelebihan pada model kue tradisional yang dikembangkan adalah menghasilkan nilai akurasi, presisi, *recall*, *f1-score* di atas 80%. Kelemahan pada model tersebut, yakni tidak dilakukannya *deployment* dari hasil pengembangan model sehingga rencana pengembangan model kue tradisional di kemudian hari, yaitu mencoba melakukan *deployment* pada hasil model yang telah diperoleh.

## **Bab IV      Penutup**

### **IV.1 Kesimpulan**

Dari proses pelaksanaan MSIB di OFA, dapat disimpulkan beberapa hal, yakni sebagai berikut:

1. Peserta mempelajari mengenai domain AI, seperti *Data Science*, NLP, dan CV.
2. Peserta menerapkan pelajaran domain AI melalui PA secara berkelompok.
3. PA yang dibuat adalah membandingkan algoritma CNN dan algoritma ViT pada dataset Kue Tradisional Indonesia.
4. Pada PA yang dikerjakan, algoritma CNN menggunakan dua arsitektur, yaitu VGG16 dan ResNet50. Begitu juga dengan algoritma ViT yang mana nilai *hyperparameter* yang digunakan dibedakan pada masing-masing arsitektur.
5. Pada algoritma CNN, arsitektur ResNet50 menghasilkan nilai akurasi, presisi, dan *recall* yang lebih baik dibandingkan VGG16.
6. Algoritma ViT arsitektur A dan B memiliki nilai akurasi, *recall*, dan *f1-score* yang sama, kecuali presisi.
7. PA yang dikerjakan berbasis penelitian atau jurnal yang disubmit pada *jurnal Science and Technology Index (SINTA)* 4.

### **IV.2 Saran**

Setelah Pelaksanaan MSIB di Mitra Orbit future academy selama kurang lebih 5 bulan, banyak pengalaman dan juga ilmu yang sangat bermanfaat di masa yang akan datang. Demi membangun dan memajukan program MSIB khususnya di Mitra OFA. Saran yang membangun akan sangat dibutuhkan. Saran yang dapat disampaikan, yakni sebagai berikut:

1. Ada baiknya pihak MSIB khususnya mitra OFA untuk lebih memperbaiki jam belajar atau jam mengajar kepada *student*, dikarenakan jam belajar sangat berpengaruh terhadap kinerja otak *student* atau peserta.
2. Mengingat banyak *student* yang bukan berasal langsung dari bidang sistem informasi, maka keahlian mengajar dari para coach sangat diperlukan.

Dikarenakan dasar pelajaran yang kami dapat dari kelas Studi Independen OFA adalah mengenai AI. Yang dimaksud keahlian mengajar disini adalah *student* atau peserta sangat berharap pengajar yang terkait benar-benar paham apa yang akan disampaikan kepada *student*, tidak hanya menjelaskan berdasarkan materi yang dibagikan kepada *student*.

3. Pada *batch* selanjutnya diharapkan adanya kekompakan dan juga tanggung jawab dari para coach dan juga *student* mengenai jalannya pembelajaran, agar para coach dan juga *student* dapat sama-sama mendapatkan manfaat dan ilmu yang merata.
4. Model pada penelitian ini dapat dikembangkan menjadi lebih baik lagi, misalnya hasil model dapat digunakan untuk *deployment*.



## Bab V Referensi

- [1] D. A. Kurnia, A. Setiawan, D. R. Amalia, R. W. Arifin, and D. Setiyadi, “Image Processing Identification for Indonesian Cake Cuisine using CNN Classification Technique,” *J. Phys. Conf. Ser.*, vol. 1783, no. 1, p. 12047, 2021, doi: 10.1088/1742-6596/1783/1/012047.
- [2] T. Karlita, B. P. Afif, and I. Prasetyaningrum, “Indonesian Traditional Cake Classification Using Convolutional Neural Networks,” *iCAST-SS 2021*, vol. 647, pp. 924–929, 2022, doi: 10.2991/assehr.k.220301.153.
- [3] I. Daud *et al.*, “Marketing Management Training and Guidance for Gempita Berkarya Group, Banua Rantau Village, Batang Alai District, HST District,” *OMNICODE J.*, vol. 1, no. 2, pp. 48–52, 2022.
- [4] Y. A. Sari *et al.*, “Indonesian Traditional Food Image Identification using Random Forest Classifier based on Color and Texture Features,” in *2019 4th International Conference on Sustainable Information Engineering and Technology (SIET)*, 2019, pp. 206–211, doi: 10.1109/SIET48054.2019.8986058.
- [5] P. K. Fahira, Z. P. Rahmadhani, P. Mursanto, A. Wibisono, and H. A. Wisesa, “Classical Machine Learning Classification for Javanese Traditional Food Image,” in *2020 4th International Conference on Informatics and Computational Sciences (ICICoS)*, 2020, pp. 1–5, doi: 10.1109/ICICoS51170.2020.9299039.
- [6] T. Lin, “Analysis on the Development Strategies of Traditional Retail Enterprises’ Digitalization under the Background of Big Data,” 2021, doi: 10.1145/3465631.3465791.
- [7] S.-L. Wamba-Taguimdje, S. Fosso Wamba, J. R. Kala Kamdjoug, and C. E. Tchatchouang Wanko, “Influence of Artificial Intelligence (AI) on Firm Performance: The Business Value of AI-Based Transformation Projects,” *Bus. Process Manag. J.*, vol. 26, no. 7, pp. 1893–1924, Jan. 2020, doi: 10.1108/BPMJ-10-2019-0411.
- [8] X. Feng, Y. Jiang, X. Yang, M. Du, and X. Li, “Computer Vision Algorithms and Hardware Implementations: A Survey,” *Integration*, vol. 69, pp. 309–

- 320, 2019, doi: <https://doi.org/10.1016/j.vlsi.2019.07.005>.
- [9] R. Chauhan, K. K. Ghanshala, and R. C. Joshi, “Convolutional Neural Network (CNN) for Image Detection and Recognition,” in *2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC)*, 2018, pp. 278–282, doi: 10.1109/ICSCCC.2018.8703316.
  - [10] A. Desiani, Erwin, B. Suprihatin, S. Yahdin, A. I. Putri, and F. R. Husein, “Bi-Path Architecture of CNN Segmentation and Classification Method for Cervical Cancer Disorders Based on Pap-smear Images,” *Int. J. Comput. Sci.*, vol. 48, no. 3, 2021.
  - [11] A. M. Rizki and N. Marina, “Klasifikasi Kerusakan Bangunan Sekolah Menggunakan Metode Convolutional Neural Network Dengan Pre-Trained Model Vgg-16,” *J. Ilm. Teknol. dan Rekayasa*, vol. 24, no. 3, pp. 197–206, 2019, doi: 10.35760/tr.2019.v24i3.2396.
  - [12] R. J. Gunawan, B. Irawan, and C. Setianingsih, “Pengenalan Ekspresi Wajah Berbasis Convolutional Neural Network Dengan Model Arsitektur VGG16,” in *e-Proceeding of Engineering*, 2021, vol. 8, no. 5, pp. 6442–6454.
  - [13] W. Setiawan, “Perbandingan Arsitektur Convolutional Neural Network Untuk Klasifikasi Fundus,” *J. Simantec*, vol. 7, no. 2, pp. 48–53, 2020, doi: 10.21107/simantec.v7i2.6551.
  - [14] D. M. Wonohadidjojo, “Perbandingan Convolutional Neural Network pada Transfer Learning Method untuk Mengklasifikasikan Sel Darah Putih,” *Ultim. J. Tek. Inform.*, vol. 13, no. 1, pp. 51–57, 2021, doi: 10.31937/ti.v13i1.2040.
  - [15] Z. Niswati, R. Hardatin, M. N. Muslimah, and S. N. Hasanah, “Perbandingan Arsitektur ResNet50 dan ResNet101 dalam Klasifikasi Kanker Serviks pada Citra Pap Smear,” *Fakt. Exacta*, vol. 14, no. 3, p. 160, 2021, doi: 10.30998/faktorexacta.v14i3.10010.
  - [16] P. Nahar, S. Tanwani, and N. S. Chaudhari, “Fingerprint Classification Using Deep Neural Network Model ResNet50,” *Int. J. Res. Anal. Rev.*, vol. 5, no. 04, pp. 1521–1535, 2018.
  - [17] Q. Yu, Y. Xia, Y. Bai, Y. Lu, A. L. Yuille, and W. Shen, “Glance-and-Gaze

- Vision Transformer,” in *Advances in Neural Information Processing Systems*, 2021, vol. 34, pp. 12992–13003, [Online]. Available: <https://proceedings.neurips.cc/paper/2021/file/6c524f9d5d7027454a783c841250ba71-Paper.pdf>.
- [18] C. F. Chen, Q. Fan, and R. Panda, “CrossViT: Cross-Attention Multi-Scale Vision Transformer for Image Classification,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2021, pp. 347–356, doi: 10.1109/ICCV48922.2021.00041.
- [19] S. A. Kamran, K. F. Hossain, A. Tavakkoli, S. L. Zuckerbrod, and S. A. Baker, “VTGAN: Semi-Supervised Retinal Image Synthesis and Disease Prediction Using Vision Transformers,” *Proc. IEEE Int. Conf. Comput. Vis. Work.*, pp. 3235–3245, 2021, doi: 10.1109/ICCVW54120.2021.00362.
- [20] B. Gheflati and H. Rivaz, “Vision Transformer for Classification of Breast Ultrasound Images,” *Comput. Vis. Pattern Recognit.*, vol. 2, pp. 1–5, 2021, [Online]. Available: <http://arxiv.org/abs/2110.14731>.
- [21] N. E. M. Khalifa, M. Loey, M. H. N. Taha, and H. N. E. T. Mohamed, “Deep Transfer Learning Models for Medical Diabetic Retinopathy Detection,” *Acta Inform. Medica*, vol. 27, no. 5, pp. 327–332, 2019, doi: 10.5455/aim.2019.27.327-332.

## Bab VI      Lampiran A. TOR

### TERM OF REFERENCE (TOR) STUDI INDEPENDEN BERSERTIFIKAT FOUNDATION OF AI AND LIFE SKILLS FOR GEN-Z DI ORBIT FUTURE ACADEMY

#### A. Rincian Program

Foundation of AI and Life Skills for Gen-Z adalah program pelatihan *Artificial Intelligence* (AI) daring yang bertujuan untuk memperkenalkan teknologi dan perangkat AI kepada pelajar, sehingga diharapkan mereka dapat mengembangkan produk AI yang memiliki dampak sosial. Program ini berfokus pada komponen utama AI, seperti Data Science (DS), Natural Language Processing (NLP), dan Computer Vision (CV). Selain keterampilan AI, pelajar juga akan mendapat *life skills* yang bermanfaat untuk mencari atau menciptakan lapangan kerja.

#### B. Tujuan Program

Tujuan yang diharapkan setelah peserta mengikuti program ini:

6. Memiliki wawasan tentang AI dan perkembangannya.
7. Mampu merancang dan mengimplementasikan AI Project Cycle.
8. Mampu menggunakan bahasa pemrograman Python untuk mengembangkan aplikasi AI.
9. Mampu mengembangkan salah satu dari 3 domain AI (DS, NLP, dan CV) hingga tahap *deployment*.
10. Mampu menggunakan *soft skills* dan *hard skills* dalam dunia industri dan lingkungan perusahaan.
11. Mampu mengaplikasikan kiat-kiat yang dibutuhkan seorang wirausahawan yang bergerak di bidang *start-up* dalam mentransformasikan ide ke dalam bentuk produk/jasa sehingga dapat menciptakan peluang bisnis yang terus berinovasi, berevolusi, dan berkelanjutan.

### C. Jadwal dan Tempat Pelaksanaan

Jadwal pelaksanaan tertera dalam tabel berikut:

Pukul (WIB)	Durasi (jam)	Aktivitas
08.00 s.d. 11.30	3.5	Kelas Sesi Pagi
13.00 s.d. 16.30	3.5	Kelas Sesi Siang
16.30 s.d. 17.30	1	<i>Self-Study</i>

Kelas akan diselenggarakan secara daring melalui aplikasi *video conference*.

### D. Peserta

Peserta program ini adalah mahasiswa yang berasal dari Perguruan Tinggi di bawah Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia.

### E. Uraian Tugas Peserta

Selama mengikuti program ini, peserta diharuskan:

1. Mengikuti program dari awal hingga selesai.
2. Mematuhi aturan program.
3. Mematuhi aturan kelas yang dibuat bersama *homeroom* atau *domain coach*.
4. Mengikuti kelas dengan presensi minimal 85%.
5. Membuat laporan harian dan mingguan di *website* Kampus Merdeka.
6. Menyelesaikan Proyek Akhir (PA) beserta laporannya.

Homeroom Coach,



Angel Metanosa Afinda, S.Kom.  
2201043

Jambi, 21 Februari 2022

Peserta Program,



Nyayu Chika Marselina  
08011281924027

## Bab VII      Lampiran B. Log Activity

Minggu/Tgl	Kegiatan	Hasil
Minggu 10 / 25-29 April 2022	Diskusi terkait PA yang akan kami jalankan	Pada minggu ini tim kami melakukan perubahan ide PA, setelah itu merencanakan domain AI yang akan digunakan dan basis hasil yang diterapkan. Mencari referensi jurnal dan artikel yang terkait mengenai ide dan mencari dataset yang akan kami gunakan. Lalu kami berdiskusi bersama terkait jadwal konsultasi bersama coach sesuai domain AI yang digunakan dalam PA. Dan juga berdiskusi mengenai fiksasi ide PA dan pembagian tugas per anggota tim dalam kelompok untuk pengerjaannya.
Minggu 12 / 9-13 Mei 2022	Minggu ini tim berfokus pada pengerjaan PA dan melakukan konsultasi bersama coach sesuai domain AI dari PA	Pada minggu ini tim membuat AI project cycle yang meliputi problem scoping (4W),

		<p>data acquisition, dan data exploration. Kami juga mendiskusikan perihal AI project cycle bagian modeling, terkait algoritma apa nantinya yang akan digunakan dalam pengerjaan PA. Lalu kami melakukan konsultasi bersama coach mengenai PA dan menyampaikan apa progres kami yang sudah dikerjakan. Setelah disampaikan kepada coach dan diberi masukan terkait PA, kami segera memperbaiki dan melakukan pembagian tugas anggota tim. Setelah itu kami melakukan fiksasi model yang digunakan, mencari referensi tampilan website yang akan dibuat, mendiskusikan hasil model yang diperoleh dari proses modeling yang dilakukan, dan mulai</p>
--	--	---

		merancang front end web.
Minggu 13 / 16-20 Mei 2022	Mengembangkan model dan melakukan Bimbingan bersama Coach Domain AI	Dilakukan pembuatan model dengan algoritma lain, sebagai pembandingan. Lalu mendiskusikan perihal connect hasil bobot model yang diperoleh bersama tim. Dan menampilkan kepada coach terkait hasil model pada colab kami.
Minggu 14 / 23-27 Mei 2022	Diskusi mengenai deployment aplikasi dan perkembangan laporan PA Bab 1	Membahas kemajuan dari laporan PA pada bab 1, lalu dilakukan juga pembuatan logo untuk aplikasi yang sedang di kembangkan, selanjutnya melakukan pengecekan ulang pada model yang kami gunakan dan mencoba running epoch kembali
Minggu 15 / 30 Mei-3 Juni 2022	Membahas mengenai deployment menggunakan flask dengan hasil model yang masih belum terbaca sehingga basis PA diubah. Melaporkan progress PA kepada Coach Ipin selaku homeroom coach kelas Cordoba	Membahas mengenai deployment, tetapi hasil model masih belum terbaca dan mengubah basis PA dari aplikasi menjadi jurnal. Mencoba



		<p>melakukan deployment menggunakan flask tetapi terkendala karena model tidak terbaca. Mencoba ulang deployment menggunakan flask tetapi model masih belum terbaca.</p>
<p>Minggu 16 / 6-10 Juni 2022</p>	<p>Membuat jurnal dengan mengubah basis PA dari aplikasi menjadi jurnal. Melakukan koreksi jurnal bersama Coach Angel. Serta mencari algoritma pembandingan ViT.</p>	<p>Mengubah basis PA dari aplikasi menjadi jurnal sehingga melakukan proses pembuatan jurnal untuk bab pendahuluan dan mencari referensi untuk submit jurnal. Membuat jurnal bagian metodologi penelitian, hasil dan pembahasan, dan kesimpulan. Mencari algoritma lain yang dapat digunakan sebagai pembandingan dengan algoritma vision transformer. Mencoba menggunakan algoritma CNN menghasilkan akurasi di bawah 60% sehingga berusaha mencari algoritma lain agar menghasilkan</p>

		akurasi model di atas 75%.
Minggu 17 / 13-16 Juni 2022	Pembuatan laporan PA dan video. Presentasi progres PA bersama Coach Angel selaku homeroom coach kelas Jupyter XXI. Melakukan submit jurnal dan laporan PA	Presentasi dengan coach angel mengenai progres PA lalu membuat video presentasi untuk PA dan menyelesaikan revisi jurnal hingga submit laporan PA dan jurnal.

1. Judul

# **Perbandingan *Convolutional Neural Network* dan *Vision Transformer* Pada Klasifikasi Kue Tradisional Indonesia**

2. Data penulis

Athiya Shinta Wulandari<sup>1</sup>

<sup>1</sup>Perbankan Syariah, Universitas Muhammadiyah Surabaya  
Jawa Timur, Indonesia

Halomoan Filipus Simarmata<sup>2</sup>

<sup>2</sup>Sistem Informasi, Universitas Telkom  
Jawa Barat, Indonesia

Nyayu Chika Marselina<sup>3</sup>

<sup>3</sup>Matematika, Universitas Sriwijaya  
Sumatra Selatan, Indonesia

Sukma Imelda<sup>4</sup>

<sup>4</sup>Akuntansi, Universitas Merdeka Malang  
Jawa Timur, Indonesia

Wella Novita Andriani<sup>5</sup>

<sup>5</sup>Desain Komunikasi Visual, Universitas Internasional Semen Indonesia  
Jawa Timur, Indonesia

Email: <sup>1</sup>athiyashinta25@gmail.com,

<sup>2</sup>halomoansimarmata@student.telkomuniversity.ac.id,

<sup>3</sup>nyayuchikamsln@gmail.com, <sup>4</sup>sukmaimelda19@gmail.com,

<sup>5</sup>wellanovita37@gmail.com

3. Abstrak

**Abstrak** — Kue tradisional Indonesia merupakan ciri khas budaya yang dimiliki Indonesia. Saat ini kue tradisional mulai digantikan dengan kue modern. Persaingan antara kue tradisional dengan kue modern sangat ketat, hal ini diperkuat dengan kue-kue modern memiliki proses pembuatan yang

lebih efisien dibandingkan kue tradisional sehingga kue tradisional semakin ditinggalkan dan dapat membuat masyarakat kedepannya kurang mengenal apa-apa saja kue tradisional yang ada di Indonesia. Salah satu cara untuk memperkenalkan kue tradisional Indonesia kepada khalayak umum dengan melakukan pengenalan objek melalui klasifikasi kue tradisional Indonesia. Beberapa algoritma yang digunakan dalam klasifikasi data citra, yakni Convolutional Neural Network (CNN) dan Vision Transformer (ViT). CNN mampu mengklasifikasikan objek data citra dalam jumlah yang besar, tetapi proses dalam menjalankan model cenderung lebih lama, sedangkan ViT memiliki ukuran model yang lebih efisien dibandingkan konvolusi konvensional dan dapat memprediksi fitur beresolusi tinggi. Penelitian ini bertujuan untuk membangun model klasifikasi pada citra kue tradisional Indonesia menggunakan algoritma CNN dan ViT. Arsitektur CNN yang digunakan, yakni VGG16 dan ResNet50. Dari algoritma ViT, dibentuk dua arsitektur dengan memilih nilai yang berbeda dari hyperparameter yang digunakan. Dataset yang digunakan memiliki 8 kelas dengan total gambar 1845 data. Parameter yang digunakan dalam mengevaluasi model, yaitu akurasi, presisi, recall, dan f1-score. Dari penelitian yang dilakukan hasil akurasi, presisi, recall, dan f1-score tertinggi diperoleh pada model ResNet50 dengan nilai secara berturut-turut adalah 94%, 94%, 94,50%, dan 94,38%. Hal tersebut menunjukkan bahwa ResNet50 memiliki performa yang lebih baik dalam klasifikasi kue Tradisional Indonesia dibandingkan arsitektur lainnya.

**Kata kunci**— *Convolutional Neural Network*; Kue Tradisional Indonesia; ResNet50; VGG16; *Vision Transformer*.

**Abstract** — Indonesian traditional cakes are a characteristic of Indonesia's culture. Currently, traditional cakes are starting to be replaced with modern cakes so traditional cakes are increasingly being abandoned and can make people in the future less familiar with traditional cakes in Indonesia. One way to introduce Indonesian traditional cakes to the general public is by

introducing objects through the classification of traditional Indonesian cakes. Several algorithms are used in image data classification, namely Convolutional Neural Network (CNN) and Vision Transformer (ViT). CNN can classify a large number of image data objects, but the process of running the model tends to be longer, while ViT has a more efficient model size than conventional convolution and can predict high-resolution features. This study aims to build a classification model on the image of traditional Indonesian cakes using CNN and ViT algorithms. CNN architectures used are VGG16 and ResNet50. From the ViT algorithm, two architectures were formed by choosing different values of hyperparameters used. Dataset used has 8 classes with 1845 images of data. Parameters used in evaluating the model are accuracy, precision, recall, and f1-score. From the research conducted, the results of the highest accuracy, precision, recall, and f1-score were obtained in the ResNet50 model with values of 94%, 94%, 94.50%, and 94.38% respectively. This shows that ResNet50 has a better performance in the classification of Traditional Indonesian cakes than other architectures.

**Keywords**— Convolutional Neural Network; Indonesian Traditional Cake; ResNet50; VGG16; Vision Transformer.

#### **4. Pendahuluan**

Kue tradisional Indonesia adalah salah satu bentuk dari ciri khas dan kebudayaan yang dimiliki oleh Indonesia. Terdapat banyak jenis-jenis kue tradisional Indonesia, yakni kue klepon, kue lumpur, kue kastengel, kue putri salju, kue serabi, kue dadar gulung, kue lapis, dan kue risoles [1]. Di era sekarang, kue tradisional mulai digantikan dengan kue modern. Kompetisi antara kue tradisional dengan kue modern sangat ketat apalagi kue-kue modern memiliki proses pembuatan yang lebih efisien dibandingkan kue tradisional. Tak hanya itu, terdapat banyak kue modern yang memiliki harga lebih murah sehingga kue tradisional semakin tertinggal dan minat masyarakat jauh berkurang [3]. Kue tradisional Indonesia dapat digunakan sebagai salah satu

bentuk pengenalan ciri khas atau kebudayaan Indonesia pada masyarakat. Beberapa penelitian mengenai kue atau makanan tradisional Indonesia, yakni Kurnia et al, 2021 [1] yang menggunakan algoritma *Convolutional Neural Network* (CNN) dengan akurasi yang diperoleh sebesar 65%, tetapi tidak menampilkan nilai presisi, *recall*, dan *f1-score* pada penelitian yang dilakukan. Penelitian Sari et al, 2019 [4] melakukan identifikasi pada makanan tradisional Indonesia menggunakan *Random Forest*. Hasil akurasi yang diperoleh adalah 59,7%. Pada penelitian tersebut tidak menampilkan nilai presisi, *recall*, dan *f1-score*. Penelitian Fahira et al, 2020 [5] melakukan klasifikasi pada makanan tradisional Jawa Indonesia menggunakan KNN. Hasil akurasi dan *f1-score* yang diperoleh adalah 89,17% dan 85,58%, tetapi tidak menampilkan hasil presisi dan *recall*.

Salah satu cara untuk memperkenalkan kue tradisional Indonesia pada khalayak umum di masa modern, yakni dengan memanfaatkan *Artificial Intelligence* (AI) atau kecerdasan buatan [6]. Salah satu kelebihan dari AI adalah dapat menyebarkan informasi secara cepat dan mudah [7]. Domain AI yang dapat mengolah data citra adalah *Computer Vision* (CV). Salah satu tugas dari CV, yakni melakukan klasifikasi pada data citra [8]. Salah satu algoritma yang dapat melakukan klasifikasi data citra, yakni CNN [9]. Algoritma yang memiliki performa yang baik dalam melakukan pelatihan dataset dalam jumlah besar adalah CNN [10]. Adapun arsitektur pada CNN adalah VGG16 dan ResNet50. VGG16 merupakan model yang menunjukkan bahwa jaringan yang berada di dalam merupakan faktor penting untuk kinerja yang baik. Keunggulan model ini adalah arsitektur sangat homogen, hanya melakukan akumulasi  $3 \times 3$  dan  $2 \times 2$  end-to-end [11].

Arsitektur ResNet-50 sendiri merupakan jaringan residual yang terdiri dari 50 layer. ResNet50 bukan berfokus menyelidiki beberapa fitur, tetapi menyelidiki beberapa residual. Maksud dari residual adalah pengurangan fitur yang dipelajari berdasarkan masukan suatu layer [14]. Selain CNN, algoritma yang dapat digunakan untuk klasifikasi data citra adalah *Vision Transformer* (ViT) yang memiliki ukuran model lebih efisien dibandingkan konvolusi

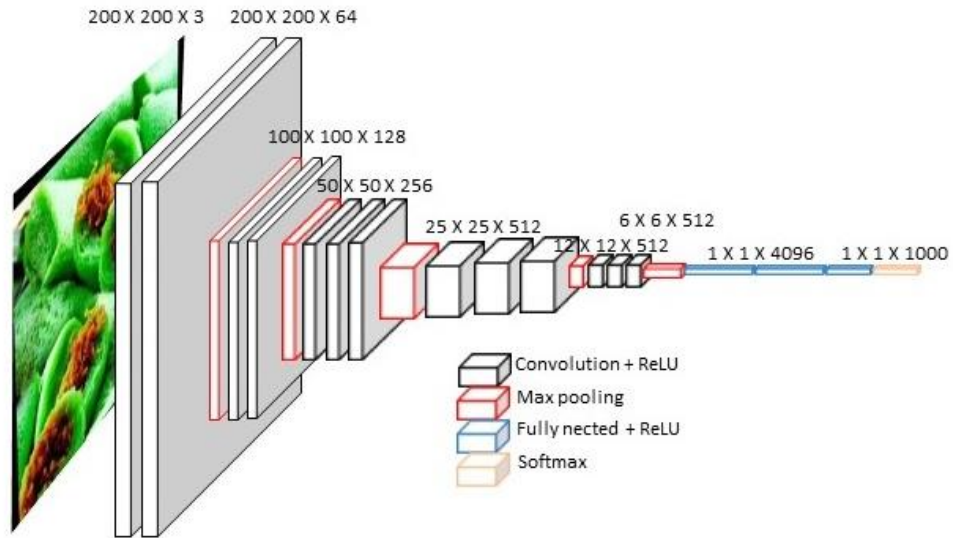
konvensional dan dapat memprediksi fitur dengan resolusi yang tinggi dalam waktu yang lebih singkat [17]. Sebelumnya, ViT hanya digunakan dalam *Natural Language Processing* (NLP), namun saat ini ViT menjadi salah satu *transformer* yang dimanfaatkan pada pengolahan citra. Algoritma ViT telah digunakan pada beberapa penelitian salah satunya penelitian yang dilakukan oleh Chen et al, 2021 [18] melakukan klasifikasi menggunakan dataset ImageNet1K yang menghasilkan akurasi sebesar 82%, tetapi tidak menampilkan hasil presisi, *recall*, dan *f1-score*. Penelitian Kamran et al, 2021 [19] mengklasifikasikan citra retina menggunakan ViT dengan akurasi, presisi, dan *recall* yang diperoleh sebesar 85,7%, 90%, dan 83,3%, tetapi tidak menampilkan nilai *f1-score* pada penelitian yang dilakukan. Penelitian yang dilakukan oleh Gheflati dan Rivaz, 2021 [20] yang melakukan klasifikasi pada citra medis *ultrasound* payudara menghasilkan akurasi sebesar 86%, tetapi tidak menampilkan hasil presisi, *recall*, dan *f1-score*.

Melihat penelitian-penelitian sebelumnya, hasil yang diperoleh belum maksimal sehingga pada penelitian ini dilakukan klasifikasi kue tradisional Indonesia menggunakan algoritma CNN dan ViT. Pada penelitian ini arsitektur CNN yang digunakan adalah VGG16 dan ResNet50. Algoritma ViT yang digunakan akan dibagi menjadi dua arsitektur yang dibentuk dengan perombakan arsitektur yang mengubah nilai-nilai *hyperparameter* tertentu. Parameter yang akan digunakan untuk melakukan evaluasi model yang dibuat adalah akurasi, presisi, *recall*, dan *f1-score*. Kedua algoritma yang digunakan akan dibandingkan untuk menentukan algoritma yang lebih baik dalam melakukan klasifikasi citra kue tradisional Indonesia.

## 5. Studi Literatur

### A. Penerapan Algoritma CNN

- 1) *Arsitektur VGG16*: Hasil *preprocessing* data sebelumnya akan diimplementasikan ke dalam arsitektur VGG16. Dalam penelitian ini, arsitektur VGG16 yang digunakan dapat dilihat pada Gambar 1.



Gambar 1. Arsitektur VGG16 [12]

Berdasarkan Gambar 1, terlihat bahwa arsitektur VGG16 memiliki 5 blok konvolusi yang mana pada lapisan konvolusi pertama merupakan *input* gambar berukuran  $200 \times 200$  dengan *channel* RGB (*Red, Green, Blue*). Pada setiap blok konvolusi terdapat lapisan *pooling* dengan jenis *pooling* yang digunakan adalah *max pooling*. Setelah 5 blok konvolusi terdapat lapisan *fully connected* yang menerapkan fungsi aktivasi *Rectified Linear Unit* (ReLU). Berdasarkan [21], persamaan dari fungsi aktivasi ReLU dituliskan pada Persamaan (1).

$$f(x) = \max(0, x) \quad (1)$$

Pada lapisan *fully connected* terakhir, dilakukan klasifikasi pada data citra ke dalam 8 kelas sehingga setiap data gambar memiliki label atau kelasnya masing-masing. Dalam penentuan kelas atau label, digunakan fungsi aktivasi softmax. Berdasarkan [22], persamaan softmax dapat dilihat pada Persamaan (2).

$$f(x_i) = \frac{e^{x_i}}{\sum_{j=1}^N e^{x_j}} \quad (2)$$

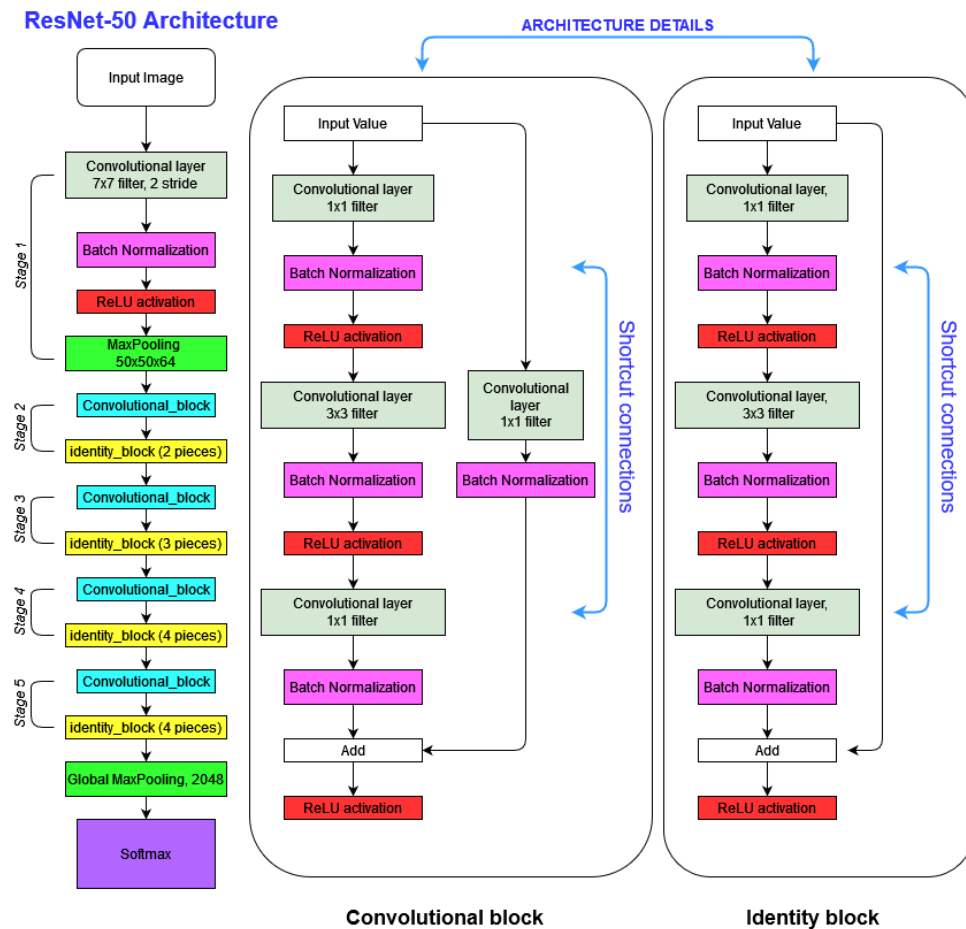


Dimana  $x_j$  adalah nilai input pada lapisan softmax dan  $f(x_i)$  adalah probabilitas yang diperoleh pada kategori ke-i. Untuk menghitung *loss* pada proses *training*, digunakan fungsi *categorical cross entropy*. Berdasarkan [23], persamaan *categorical cross entropy* dapat dilihat pada Persamaan (3).

$$J_{cce} = - \sum_{q=1}^l \sum_{k=1}^p d_{qk} \log(y_{qk}) \quad (3)$$

Dimana  $y_{qk}$  merupakan peluang prediksi, d adalah target, J adalah fungsi *cross entropy*, k adalah kelas dari gambar. Selain fungsi *loss*, digunakan juga *optimizer*, yakni *Adaptive Moment Optimization* (Adam). VGG16 telah diterapkan di beberapa penelitian. Penelitian Gunawan et al, 2021 [12] melakukan pengenalan ekspresi wajah dengan dataset Facial Expression Recognition 2013 yang menghasilkan akurasi di bawah 70%. Pada penelitian tersebut tidak ditunjukkan nilai presisi, recall, dan f1-score. Adapun penelitian lain menggunakan VGG16 yang dilakukan Setiawan, 2020 [13] melakukan uji coba dengan akurasi tertinggi yang diperoleh sebesar 89,3%.

2) *Arsitektur ResNet50*: Selain VGG16, hasil *preprocessing data* juga diimplementasikan ke dalam arsitektur ResNet50. Pada penelitian ini, arsitektur ResNet50 yang digunakan dapat dilihat pada Gambar 2.



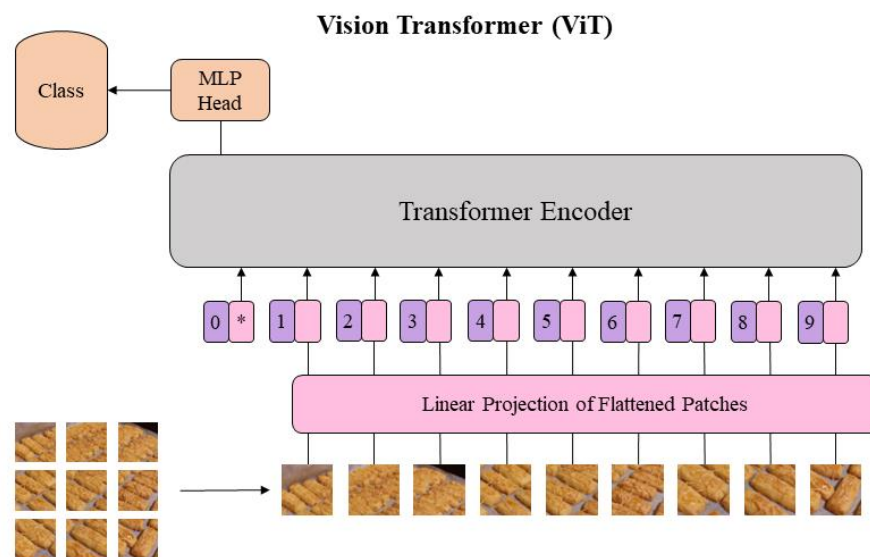
Gambar 2. Arsitektur ResNet50 [24]

Berdasarkan Gambar 2, ResNet50 memiliki lima *stage* konvolusi di mana di dalam setiap *stage* terdapat blok konvolusi dan blok *identity*. Pada tiap blok konvolusi dan blok *identity* terdapat *batch normalization* dan fungsi aktivasi ReLU. Persamaan ReLU dapat dilihat pada Persamaan (1). Sebelum masuk ke *stage 2* dan *fully connected layer*, terdapat *pooling layer*. Pooling yang digunakan adalah *max pooling*. *Fully connected layer* terletak setelah *stage* konvolusi yang mana dihasilkannya *output* yang dibagi menjadi 8 kelas dan penentuan kelas dilakukan menggunakan fungsi aktivasi softmax. Rumus perhitungan fungsi aktivasi softmax dapat dilihat pada Persamaan (2). Pada proses *compile* model ResNet50, digunakan fungsi *loss* dan *optimizer*. Fungsi *loss* dan *optimizer* yang digunakan adalah *categorical cross entropy* dan Adam. Persamaan fungsi *loss categorical cross entropy* dapat dilihat pada Persamaan

(3). Arsitektur ResNet50 telah diterapkan pada beberapa penelitian salah satunya, pada penelitian Niswati et al, 2021 [15] yang melakukan klasifikasi kanker serviks pada citra pap smear. Penelitian tersebut menghasilkan nilai akurasi sebesar 91%. Walaupun akurasi yang diperoleh di atas 90%, penelitian tersebut tidak menampilkan hasil dari presisi, recall, dan f1-score. Penelitian lainnya yang menggunakan ResNet50, yakni Nahar et al, 2018 [16] yang melakukan klasifikasi citra sidik jari dengan hasil akurasi yang diperoleh sebesar 90%, tetapi tidak menampilkan nilai presisi, recall, dan f1-score pada penelitian yang dilakukan.

### B. Penerapan Algoritma Vision Transformer

Selain diterapkan menggunakan algoritma CNN, hasil *preprocessing data* sebelumnya juga akan diimplementasikan ke dalam algoritma ViT. Dalam penelitian ini, arsitektur ViT yang digunakan dapat dilihat pada Gambar 3.



Gambar 3. Arsitektur ViT [25]

Dari Gambar 3 dapat dilihat, bahwa pada algoritma ViT terdapat beberapa langkah dalam mengklasifikasikan kelas, khususnya pada dataset kue tradisional Indonesia. Langkah awal pada ViT adalah *input* data citra kue tradisional Indonesia. Kemudian, mendefinisikan *hyperparameter* yang akan

digunakan selama proses *modeling*. Dalam hal ini, digunakan sebanyak 8 lapisan *transformer*. Data citra kue tradisional Indonesia akan di augmentasi dengan normalisasi dan mengubah ukuran citra menjadi  $256 \times 256$ , dan dilakukan *flip* acak. Berdasarkan [26] persamaan normalisasi data terdapat pada Persamaan (4).

$$Z = \frac{x_i - \bar{x}}{\sigma} \quad (4)$$

Dimana,  $i$  menunjukkan nilai  $x$  ke- $i$ ,  $\bar{x}$  adalah rata-rata, dan  $\sigma$  adalah standar deviasi atau varians. Setelah dilakukan augmentasi, dihitung nilai rata-rata dan varians dari *data training* untuk normalisasi. Tahap selanjutnya, membangun sebuah jaringan. Pada penelitian ini akan digunakan jaringan *Multilayer Perceptron* (MLP) dan *layer* akan memisahkan citra ke dalam *patch*. Tak hanya itu, digunakan juga *path encoder* untuk mengubah *patch* yang akan memproyeksikan *patch* menjadi vektor. Pada jaringan MLP, fungsi aktivasi yang digunakan adalah *Gaussian Error Linear Unit* (GELU). Berdasarkan [27] fungsi aktivasi GELU dapat dilihat pada Persamaan (5).

$$GELU(v) = v \cdot P(V \leq v) = v \cdot \Phi(v) = v \cdot \frac{1}{2} \left[ 1 + \operatorname{erf} \left( \frac{v}{\sqrt{2}} \right) \right] \quad (5)$$

Dimana  $v$  dan  $\Phi(v)$  merupakan *input* pada fungsi aktivasi dan fungsi distribusi kumulatif  $N(0,1)$ . Mendefinisikan jaringan yang dapat mengubah citra menjadi *patch* yang bertujuan agar ViT belajar mengklasifikasikan gambar kue tradisional Indonesia. *Path encoder* akan melakukan transformasi linear dari citra *patch* dan menambahkan posisi pembelajaran ke vektor yang diproyeksikan. Setelah membentuk jaringan, dilakukan pembentukan model ViT. Dibentuk blok ViT. Pada blok *transformer*, digunakan lapisan *self-attention* pada urutan *patch*. *Output* dari blok *transformer* akan diklasifikasikan agar membantu proses *output*. Dari langkah-langkah sebelumnya, telah terdapat

algoritma untuk augmentasi, pembuatan *patch*, dan *patch encoding*. *Patch* yang dikodekan akan menjadi input akhir sebagai representasi citra ke *transformer*.

Terdapat lapisan *flatten* yang berfungsi untuk mengubah bentuk *output*. Tahap berikutnya, dilakukan proses kompilasi dan *training* model yang telah dibuat. Dalam proses kompilasi, *optimizer* yang digunakan adalah Adam yang merupakan algoritma pengoptimalan yang berbasis gradien orde pertama dimana dalam proses komputasi membutuhkan memori yang rendah [28] dan fungsi *loss* yang digunakan adalah *sparse categorical cross entropy* yang mana tidak perlu melakukan *hot encoding* terlebih dahulu. Berdasarkan [29], persamaan *sparse categorical cross entropy* dapat dilihat pada Persamaan (6).

$$L(\theta) = -\sum_{i=1}^k y_i \log \hat{y}_i \quad (6)$$

Dimana,  $k$  adalah jumlah *training*  $y_i$ ,  $\hat{y}_i$  adalah fungsi aktivasi yang digunakan. Sebelum melakukan *training* model, dataset kue tradisional Indonesia dibagi menjadi 80% *data training* dan 20% *data testing*. Lalu, dilakukan evaluasi model menggunakan beberapa parameter, seperti akurasi, presisi, *recall*, dan *f1-score*. Dilakukan perombakan arsitektur ViT sehingga terdapat dua arsitektur dengan perbandingan nilai *hyperparameter* yang dapat dilihat pada Tabel 1.

Tabel 1. Perbandingan Arsitektur A dan Arsitektur B Algoritma ViT

<i>Hyperparameter</i>	Arsitektur A	Arsitektur B
learning_rate	0,001	0,0001
weight_decay	0,0001	0,00001
batch_size	32	64
image_size	72	144
patch_size	6	12
projection_dim	64	128
num_heads	4	8

Berdasarkan Tabel 1, *learning\_rate* dan *weight\_decay* yang digunakan pada arsitektur B lebih kecil dibandingkan arsitektur A. *Hyperparameter* *batch\_size*, *image\_size*, *patch\_size*, *projection\_dim*, dan *num\_heads* pada arsitektur B bernilai dua kali lipat dibandingkan arsitektur A. *Hyperparameter* lainnya, seperti *num\_epochs*, *transformer\_layers*, dan *mlp\_head\_units* tidak mengalami perubahan pada kedua arsitektur sehingga ketiga variabel tersebut memiliki nilai yang sama pada kedua arsitektur ViT.

### C. Evaluasi

*Confusion matrix* adalah tabel yang menyatakan klasifikasi jumlah data uji yang benar dan jumlah data uji yang salah [30]. Berikut *confusion matrix* untuk klasifikasi dua kelas menurut [31] pada Tabel 2.

Tabel 2. Confusion Matrix

		Prediksi	
		True	False
Aktual	True	TP	FN
	False	FP	TN

Berdasarkan Tabel 2, dalam penaksiran kinerja suatu model menggunakan *confusion matrix*, terdapat empat istilah sebagai representasi hasil proses klasifikasi, yakni *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN) [32]. Adapun ukuran evaluasi kinerja yang digunakan dalam klasifikasi, yakni akurasi, presisi, *recall*, dan *f1-score*. Berdasarkan [21] akurasi, presisi, *recall*, dan *f1-score* dapat dihitung menggunakan Persamaan (7), Persamaan (8), Persamaan (9), dan Persamaan (10).

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (7)$$

$$Presisi = \frac{TP}{TP+FN} \times 100\% \quad (8)$$

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (9)$$

$$F1 - Score = 2 \times \frac{Presisi \times Recall}{Presisi + Recall} \times 100\% \quad (10)$$

## 6. Hasil dan Analisa

Pada penelitian ini, dataset kue tradisional Indonesia memiliki 8 kelas, yakni kue klepon, kue lumpur, kue kastengel, kue putri salju, kue serabi, kue dadar gulung, kue lapis, dan kue risoles. Dilakukan perubahan ukuran data citra pada dataset kue tradisional Indonesia menjadi  $256 \times 256$ . Setelah itu, melakukan augmentasi data dengan menggunakan *horizontal flip* dan *vertikal flip*. Hasil dari proses augmentasi data dapat dilihat pada Tabel 3.

Tabel 3. Contoh Hasil Proses Augmentasi Data Citra Kue Tradisional Indonesia

			
Kue Klepon	Kue Lumpur	Kue Kastengel	Kue Putri Salju
			
Kue Serabi	Kue Dadar Gulung	Kue Lapis	Kue Risoles

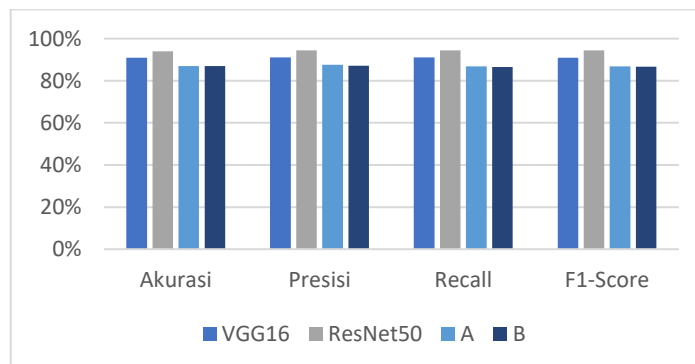
Pada Tabel 3 dapat dilihat, bahwa terdapat 8 contoh data citra yang diperoleh dari proses augmentasi data. Dengan augmentasi, dataset kue tradisional meningkat menjadi 5227 data citra. Setelah melakukan augmentasi data, melakukan split data pada dataset kue tradisional Indonesia dengan membagi 5277 data citra ke dalam 80% *data training* dan 20% *data testing* sehingga terdapat 4222 *data training* dan 1055 *data testing*. Dilakukan proses pelatihan data menggunakan algoritma CNN dan ViT. Pada algoritma CNN, digunakan arsitektur VGG16 dan arsitektur ResNet50, sedangkan pada

algoritma ViT dilakukan perombakan arsitektur sehingga menghasilkan dua arsitektur, yakni arsitektur A dan arsitektur B. Dilakukan proses pelatihan data terlebih dahulu. Hasil dari pelatihan atau *training* akan dievaluasi menggunakan *confusion matrix* yang mana parameter yang digunakan adalah akurasi, presisi, *recall*, dan *f1-score*. Hasil perbandingan akurasi, presisi, *recall*, dan *f1-score* dapat dilihat pada Tabel 4.

Tabel 4. Perbandingan Hasil Akurasi, Presisi, *Recall*, dan *F1-Score* Model Kue Tradisional Indonesia

Algoritma	Arsitektur	Akurasi	Presisi	<i>Recall</i>	<i>F1-Score</i>
CNN	VGG16	91%	91,13%	91,13%	90,88%
	<b>ResNet50</b>	<b>94%</b>	<b>94,38%</b>	<b>94.5%</b>	<b>94.38%</b>
ViT	A	87%	87,63%	86,75%	86,75%
	B	87%	87,13%	86,5%	86,63%

Berdasarkan Tabel 4 dapat dilihat bahwa nilai akurasi presisi, *recall*, dan *f1-score* tertinggi terdapat pada model yang menggunakan arsitektur ResNet50, sedangkan nilai akurasi presisi, *recall*, dan *f1-score* terendah diperoleh oleh model yang menggunakan algoritma ViT. Hasil akurasi, presisi, *recall*, dan *f1-score* dari VGG16, ResNet50, dan ViT dapat dilihat secara ringkas pada Gambar 4.



Gambar 4. Perbandingan Hasil Akurasi, Presisi, *Recall*, dan *F1-Score* Model Kue Tradisional Indonesia



Dari Gambar 4 dapat dilihat bahwa nilai akurasi presisi, *recall*, dan *f1-score* yang diperoleh berada di atas 80% sehingga model yang dibuat menggunakan VGG16, ResNet50, dan ViT memiliki performa yang baik dalam melakukan klasifikasi kue tradisional Indonesia. Jika keempat model yang dibuat tersebut dibandingkan, maka dapat diketahui bahwa model arsitektur ResNet50 memiliki performa yang terbaik di antara model lainnya sehingga ResNet50 lebih baik dalam melakukan klasifikasi kue tradisional Indonesia dibandingkan keempat model lainnya.

## 7. Kesimpulan

Berdasarkan penelitian yang dilakukan, diperoleh akurasi, presisi, recall, dan f1-score dibuat menggunakan VGG16, ResNet50, dan ViT berada di atas 80% yang berarti bahwa keempat model memiliki performa yang baik dalam melakukan klasifikasi kue tradisional Indonesia. Hasil akurasi, presisi, recall, dan f1-score tertinggi diperoleh pada arsitektur ResNet50. Akurasi, presisi, recall, dan f1-score yang dihasilkan oleh model dengan arsitektur ResNet50, yakni 94% untuk akurasi dan presisi, 94,50% untuk recall, dan 94,38% untuk f1-score. Hal tersebut menunjukkan, bahwa model arsitektur ResNet50 memiliki performa yang lebih baik dibandingkan arsitektur lainnya dalam pengklasifikasian kue tradisional Indonesia.

## 8. Ucapan terima kasih

Puji syukur dicurahkan ke hadirat Allah SWT karena atas rahmat dan karunia-Nya penelitian ini selesai pada waktunya dan terselesaikan dengan baik. Tidak lupa ucapan terima kasih diucapkan kepada:

1. Orang tua yang selalu mendukung jalannya program ini dari awal hingga akhir.
2. Orbit Future Academy selaku instansi pada program Magang dan Studi Independen Bersertifikat (MSIB) beserta para coach yang lain dan juga staff terkait.

3. Coach Angel Metanosa Afinda selaku *homeroom coach* dari kelas Jupyter XXI dan juga Coach Ipin Sugiyarto selaku *homeroom coach* kelas Cordoba di Orbit Future Academy .
4. Coach Fauzan selaku coach pembimbing domain AI CV di Orbit Future Academy.
5. Teman-teman tim yang sudah berjuang untuk menyelesaikan penelitian ini.

### **9. Author Contributions Statement\***

Pada perumusan ide pada PA diusulkan oleh Sukma Imelda. Ide tersebut dilandasi oleh kue tradisional Indonesia yang sudah mulai mengalami ketertinggalan karena terdapat banyak kue-kue modern yang lebih mudah dibuat dan didapatkan dibandingkan kue tradisional Indonesia sehingga kue tradisional Indonesia memiliki kemungkinan untuk tidak dikenali oleh khalayak umum kedepannya. Untuk mengatasi hal tersebut dapat dilakukan dengan memanfaatkan AI karena AI memiliki salah satu keunggulan, yakni menyebarkan informasi secara cepat dan mudah.

Dalam hal ini pengumpulan data mengenai kue tradisional Indonesia dilakukan oleh Wella Novita Andriani yang melakukan pencarian data melalui Kaggle. Data yang diperoleh berupa data citra sebanyak 1845 data dengan 8 kelas, yakni kue klepon, kue lumpur, kue kastengel, kue putri salju, kue serabi, kue dadar gulung, kue lapis, dan kue risoles. Pada setiap kelas kue tradisional Indonesia terdapat lebih dari 200 data. Salah satu domain AI yang dapat digunakan dalam mengolah data citra adalah CV dimana memiliki algoritma dalam melakukan proses pengolahan. Halomoan Filipus Simarmata mengusulkan algoritma yang digunakan adalah CNN dan ViT karena CNN memiliki kelebihan yang dapat melakukan proses pengolahan data citra dalam jumlah banyak dan ViT merupakan algoritma yang dikembangkan dari CNN.

Selain itu, CNN memiliki berbagai arsitektur, tetapi arsitektur yang dipilih dalam melakukan penelitian adalah VGG16 dan ResNet50. Nyayu Chika Marselina melakukan pembentukan dua buah arsitektur ViT dengan

membedakan nilai beberapa *hyperparameter* yang digunakan pada masing-masing arsitektur. Pembentukan model CNN dan ViT dilakukan oleh Halomoan Filipus Simarmata dan Nyayu Chika Marselina. Namun, selama proses pelatihan data, hasil akurasi yang diperoleh berada di bawah 60% yang berarti model yang dibentuk belum cukup baik dalam melakukan klasifikasi kue tradisional Indonesia. Athiya Shinta Wulandari mengusulkan untuk dilakukan penambahan dataset karena memiliki kemungkinan bahwa dataset yang digunakan masih sedikit untuk proses pelatihan model.

Setelah dilakukan augmentasi data, diperoleh sebanyak 5277 data gambar kue tradisional Indonesia dan model yang dilatih menghasilkan akurasi di atas 80% yang berarti model sudah memiliki performa yang baik dalam klasifikasi kue tradisional Indonesia. Pada penulisan jurnal, Sukma Imelda menuliskan bagian latar belakang dan ucapan terima kasih, Wella Novita Andriani dan Athiya Shinta Wulandari menuliskan bagian metodologi penelitian, Halomoan Filipus Simarmata dan Nyayu Chika Marselina menuliskan bagian hasil pembahasan dan kesimpulan dari penelitian yang dilakukan.

## **10. Daftar pustaka**

- [1] T. Karlita, B. P. Afif, and I. Prasetyaningrum, "Indonesian Traditional Cake Classification Using Convolutional Neural Networks," *iCAST-SS 2021*, vol. 647, pp. 924–929, 2022, doi: 10.2991/assehr.k.220301.153.
- [2] I. Daud et al., "Marketing Management Training and Guidance for Gempita Berkarya Group, Banua Rantau Village, Batang Alai District, HST District," *OMNICODE J.*, vol. 1, no. 2, pp. 48–52, 2022.
- [3] D. A. Kurnia, A. Setiawan, D. R. Amalia, R. W. Arifin, and D. Setiyadi, "Image Processing Identification for Indonesian Cake Cuisine using CNN Classification Technique," *J. Phys. Conf. Ser.*, vol. 1783, no. 1, p. 12047, 2021, doi: 10.1088/1742-6596/1783/1/012047.
- [4] Y. A. Sari et al., "Indonesian Traditional Food Image Identification using Random Forest Classifier based on Color and Texture Features," in *2019 4th International Conference on Sustainable Information Engineering and*

- Technology (SIET), 2019, pp. 206–211, doi: 10.1109/SIET48054.2019.8986058.
- [5] P. K. Fahira, Z. P. Rahmadhani, P. Mursanto, A. Wibisono, and H. A. Wisesa, “Classical Machine Learning Classification for Javanese Traditional Food Image,” in 2020 4th International Conference on Informatics and Computational Sciences (ICICoS), 2020, pp. 1–5, doi: 10.1109/ICICoS51170.2020.9299039.
- [6] T. Lin, “Analysis on the Development Strategies of Traditional Retail Enterprises’ Digitalization under the Background of Big Data,” 2021, doi: 10.1145/3465631.3465791.
- [7] S.-L. Wamba-Taguimdje, S. Fosso Wamba, J. R. Kala Kamdjoug, and C. E. Tchatchouang Wanko, “Influence of Artificial Intelligence (AI) on Firm Performance: The Business Value of AI-Based Transformation Projects,” *Bus. Process Manag. J.*, vol. 26, no. 7, pp. 1893–1924, Jan. 2020, doi: 10.1108/BPMJ-10-2019-0411.
- [8] X. Feng, Y. Jiang, X. Yang, M. Du, and X. Li, “Computer Vision Algorithms and Hardware Implementations: A Survey,” *Integration*, vol. 69, pp. 309–320, 2019, doi: <https://doi.org/10.1016/j.vlsi.2019.07.005>.
- [9] R. Chauhan, K. K. Ghanshala, and R. C. Joshi, “Convolutional Neural Network (CNN) for Image Detection and Recognition,” in 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC), 2018, pp. 278–282, doi: 10.1109/ICSCCC.2018.8703316.
- [10] A. Desiani, Erwin, B. Suprihatin, S. Yahdin, A. I. Putri, and F. R. Husein, “Bi-Path Architecture of CNN Segmentation and Classification Method for Cervical Cancer Disorders Based on Pap-smear Images,” *Int. J. Comput. Sci.*, vol. 48, no. 3, 2021.
- [11] A. M. Rizki and N. Marina, “Klasifikasi Kerusakan Bangunan Sekolah Menggunakan Metode Convolutional Neural Network Dengan Pre-Trained Model Vgg-16,” *J. Ilm. Teknol. dan Rekayasa*, vol. 24, no. 3, pp. 197–206, 2019, doi: 10.35760/tr.2019.v24i3.2396.

- [12] R. J. Gunawan, B. Irawan, and C. Setianingsih, "Pengenalan Ekspresi Wajah Berbasis Convolutional Neural Network Dengan Model Arsitektur VGG16," in *e-Proceeding of Engineering*, 2021, vol. 8, no. 5, pp. 6442–6454.
- [13] W. Setiawan, "Perbandingan Arsitektur Convolutional Neural Network Untuk Klasifikasi Fundus," *J. Simantec*, vol. 7, no. 2, pp. 48–53, 2020, doi: 10.21107/simantec.v7i2.6551.
- [14] D. M. Wonohadidjojo, "Perbandingan Convolutional Neural Network pada Transfer Learning Method untuk Mengklasifikasikan Sel Darah Putih," *Ultim. J. Tek. Inform.*, vol. 13, no. 1, pp. 51–57, 2021, doi: 10.31937/ti.v13i1.2040.
- [15] Z. Niswati, R. Hardatin, M. N. Muslimah, and S. N. Hasanah, "Perbandingan Arsitektur ResNet50 dan ResNet101 dalam Klasifikasi Kanker Serviks pada Citra Pap Smear," *Fakt. Exacta*, vol. 14, no. 3, p. 160, 2021, doi: 10.30998/faktorexacta.v14i3.10010.
- [16] P. Nahar, S. Tanwani, and N. S. Chaudhari, "Fingerprint Classification Using Deep Neural Network Model ResNet50," *Int. J. Res. Anal. Rev.*, vol. 5, no. 04, pp. 1521–1535, 2018.
- [17] Q. Yu, Y. Xia, Y. Bai, Y. Lu, A. L. Yuille, and W. Shen, "Glance-and-Gaze Vision Transformer," in *Advances in Neural Information Processing Systems*, 2021, vol. 34, pp. 12992–13003, [Online]. Available: <https://proceedings.neurips.cc/paper/2021/file/6c524f9d5d7027454a783c841250ba71-Paper.pdf>.
- [18] C. F. Chen, Q. Fan, and R. Panda, "CrossViT: Cross-Attention Multi-Scale Vision Transformer for Image Classification," in *Proceedings of the IEEE International Conference on Computer Vision*, 2021, pp. 347–356, doi: 10.1109/ICCV48922.2021.00041.
- [19] S. A. Kamran, K. F. Hossain, A. Tavakkoli, S. L. Zuckerbrod, and S. A. Baker, "VTGAN: Semi-Supervised Retinal Image Synthesis and Disease Prediction Using Vision Transformers," *Proc. IEEE Int. Conf. Comput.*

- Vis. Work., pp. 3235–3245, 2021, doi: 10.1109/ICCVW54120.2021.00362.
- [20] B. Gheflati and H. Rivaz, “Vision Transformer for Classification of Breast Ultrasound Images,” *Comput. Vis. Pattern Recognit.*, vol. 2, pp. 1–5, 2021, [Online]. Available: <http://arxiv.org/abs/2110.14731>.
- [21] C. Baldassi, E. M. Malatesta, and R. Zecchina, “Properties of the Geometry of Solutions and Capacity of Multilayer Neural Networks with Rectified Linear Unit Activations,” *Am. Phys. Soc.*, vol. 123, no. 17, pp. 1–6, 2019, doi: 10.1103/PhysRevLett.123.170602.
- [22] M. Wang, S. Lu, D. Zhu, J. Lin, and Z. Wang, “A High-Speed and Low-Complexity Architecture for Softmax Function in Deep Learning,” in *2018 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS)*, 2018, pp. 223–226, doi: 10.1109/APCCAS.2018.8605654.
- [23] C. H. Chen, P. H. Lin, J. G. Hsieh, S. L. Cheng, and J. H. Jeng, “Robust Multi-Class Classification Using Linearly Scored Categorical Cross-Entropy,” *Proc. 3rd IEEE Int. Conf. Knowl. Innov. Invent.* 2020, no. 4, pp. 200–203, 2020, doi: 10.1109/ICKII50300.2020.9318835.
- [24] F. Nashrullah, S. A. Wibowo, and G. Budiman, “The Investigation of Epoch Parameters in ResNet-50 Architecture for Pornographic Classification,” *J. Comput. Electron. Telecommun.*, vol. 1, no. 1, pp. 1–8, 2020, doi: 10.52435/complete.v1i1.51.
- [25] A. Dosovitskiy et al., “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale,” in *ICLR*, 2021, pp. 1–22, [Online]. Available: <http://arxiv.org/abs/2010.11929>.
- [26] S. S. Kim, M. L. Aprahamian, and S. Lindert, “Improving Inverse Docking Target Identification with Z-Score Selection,” *Chem. Biol. Drug Des.*, vol. 93, no. 6, pp. 1105–1116, 2019, doi: 10.1111/cbdd.13453.
- [27] A. kr. Sarkar and Z.-H. Tan, “On Training Targets and Activation Functions for Deep Representation Learning in Text-Dependent Speaker Verification,” *Comput. Sci.*, pp. 1–12, 2022, [Online]. Available: <http://arxiv.org/abs/2201.06426>.

- [28] S. H. Haji and A. M. Abdulazeez, "Comparison Of Optimization Techniques Based On Gradient Descent Algorithm: A Review," *PalArch's J. Archaeol. Egypt/Egyptology*, vol. 18, no. 4, pp. 2715–2743, 2021.
- [29] E. Gordon-Rodriguez, G. Loaiza-Ganem, G. Pleiss, and J. P. Cunningham, "Uses and Abuses of the Cross-Entropy Loss: Case Studies in Modern Deep Learning," in *ICBINB*, 2020, pp. 1–10, [Online]. Available: <http://arxiv.org/abs/2011.05231>.
- [30] N. W. S. Wardhani, M. Y. Rochayani, A. Iriany, A. D. Sulistyono, and P. Lestantyo, "Cross-validation Metrics for Evaluating Classification Performance on Imbalanced Data," in *2019 International Conference on Computer, Control, Informatics and its Applications (IC3INA)*, 2019, pp. 14–18, doi: 10.1109/IC3INA48034.2019.8949568.
- [31] J. Riany, M. Fajar, and M. P. Lukman, "Penerapan Deep Sentiment Analysis pada Angket Penilaian Terbuka Menggunakan K-Nearest Neighbor," *Sisfo*, vol. 6, no. 1, pp. 147–156, 2016, doi: 10.24089/j.sisfo.2016.09.011.
- [32] G. Canbek, T. T. Temizel, and S. Sagiroglu, "TasKar: A Research and Education Tool for Calculation and Representation of Binary Classification Performance Instruments," in *2021 International Conference on Information Security and Cryptology (ISCTURKEY)*, 2021, pp. 105–110, doi: 10.1109/ISCTURKEY53027.2021.9654359.
- [33] N. E. M. Khalifa, M. Loey, M. H. N. Taha, and H. N. E. T. Mohamed, "Deep Transfer Learning Models for Medical Diabetic Retinopathy Detection," *Acta Inform. Medica*, vol. 27, no. 5, pp. 327–332, 2019, doi: 10.5455/aim.2019.27.327-332.