

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

<http://dx.doi.org/10.28932/jutisi.vXiX.X>

Riwayat Artikel

Received: xx Bulan 20xx | Final Revision: xx Bulan 20xx | Accepted: xx Bulan 20xx

Athiya Shinta Wulandari<sup>1</sup>, Halomoan Filipus Simarmata<sup>2</sup>, Nyayu Chika Marselina<sup>✉3</sup>,

Sukma Imelda<sup>4</sup>, Wella Novita Andriani<sup>5</sup>

<sup>1</sup>Perbankan Syariah, Universitas Muhammadiyah Surabaya

Jl. Raya Sutorejo No.59, Dukuh Sutorejo, Kec. Mulyorejo, Kota SBY, Jawa Timur, 60113, Indonesia

<sup>1</sup>athiyashinta25@gmail.com

<sup>2</sup>Sistem Informasi, Universitas Telkom

Jl. Telekomunikasi No. 1, Dayeuhkolot, Bandung, Jawa Barat, 40257, Indonesia

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

<sup>3</sup>Matematika, Universitas Sriwijaya

Jl. Raya Palembang-Prabumulih KM. 32 Indralaya, Sumatra Selatan, 30862, Indonesia

<sup>3</sup>nyayuchikamsln@gmail.com

<sup>4</sup>Akuntansi, Universitas Merdeka Malang

Jalan Terusan Dieng No. 62-64 Klojen, Pisang Candi, Kec. Sukun, Kota Malang, Jawa Timur, 65146, Indonesia

<sup>4</sup>sukmaimelda19@gmail.com

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

Jl. Veteran, Kb. Dalem, Sidomoro, Kebomas, Gresik Regency, Jawa Timur, 61122, Indonesia

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

✉Corresponding author: nyayuchikamsln@gmail.com

**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 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*.

# Indonesian Traditional Cake Classification Using Convolutional Neural Network and 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.

## I. 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 [2]. 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, Setiawan, Amalia, Arifin, dan Setiyadi pada tahun 2021 melakukan klasifikasi kue tradisional Indonesia menggunakan *Convolutional Neural Network* (CNN) dengan akurasi yang diperoleh sebesar 65%, tetapi tidak menampilkan nilai presisi, *recall*, dan *f1-score* pada penelitian yang dilakukan [3]. Penelitian Sari pada tahun 2019 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* [4]. Penelitian Fahira, Rahmadhani, Mursanto, Wibisono, dan Wisesa pada tahun 2020 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* [5].

Salah satu cara untuk memperkenalkan kue tradisional Indonesia pada khalayak umum di masa modern, yakni dengan menggunakan *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]. Untuk algoritma CNN menggunakan arsitektur 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, Irawan, dan Setianingsih pada tahun 2021 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* [12]. Adapun penelitian lain menggunakan VGG16 yang dilakukan Setiawan pada tahun 2020 melakukan uji coba dengan akurasi tertinggi yang diperoleh sebesar 89,3% [13].

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,

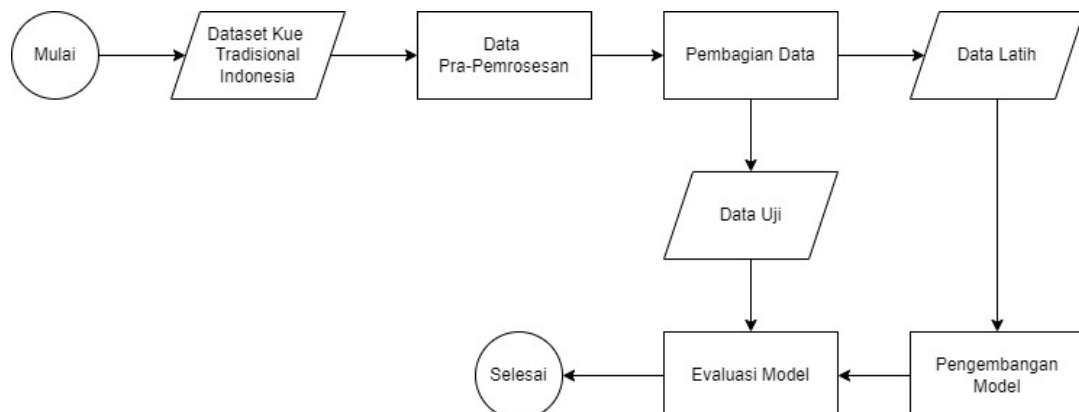
pada penelitian Niswati, Hardatin, Muslimah, dan Hasanah pada tahun 2021 yang melakukan klasifikasi pada citra kanker serviks pada citra *pap smear*. Penelitian tersebut menghasilkan nilai akurasi sebesar 91%. Walaupun akurasi yang diperoleh di atas 90%, penelitian tersebut tidak menampilkan hasil dari presisi, *recall*, dan *f1-score* [15]. Penelitian lainnya yang menggunakan ResNet50, yakni Nahar, Tanwani, dan Chaudhari pada tahun 2018 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 [16].

Selain CNN, algoritma yang dapat digunakan pada klasifikasi data citra, yaitu *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. Salah satu penelitian menggunakan ViT, yakni penelitian yang dilakukan oleh Chen, Fan, dan Panda pada tahun 2021 melakukan klasifikasi menggunakan dataset ImageNet1K yang menghasilkan akurasi sebesar 82%, tetapi tidak menampilkan hasil presisi, *recall*, dan *f1-score* [18]. Penelitian Kamran, Hossain, Tavakkoli, Zuckerbrod, dan Baker pada tahun 2021 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 [19]. Penelitian yang dilakukan oleh Gheflati dan Rivaz pada tahun 2021 yang melakukan klasifikasi pada citra medis *ultrasound* payudara menghasilkan akurasi sebesar 86%, tetapi tidak menampilkan hasil presisi, *recall*, dan *f1-score* [20].

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 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.

## II. METODOLOGI PENELITIAN

Rangkaian proses yang akan dilakukan pada penelitian klasifikasi kue tradisional Indonesia ditunjukkan, seperti pada Gambar 1.











Gambar 1. Rangkaian Proses yang Dilakukan dalam Penelitian

Berdasarkan Gambar 1, rangkaian proses yang dilakukan pada penelitian klasifikasi kue tradisional Indonesia adalah pengumpulan dataset kue tradisional Indonesia, *preprocessing data*, split data menjadi data latih dan data uji, data latih akan digunakan dalam model yang telah dibangun yang nantinya hasil pelatihan dari data tersebut akan dievaluasi, dan data uji akan langsung diuji menggunakan model yang telah latih.

### A. Pengumpulan Data

Dataset yang digunakan diperoleh dari Kaggle, yaitu dataset kue tradisional Indonesia. Dataset ini diperoleh dari link <https://www.kaggle.com/datasets/ilhamfp31/kue-indonesia>. Dataset tersebut terdiri dari 1845 data citra dengan 8 kelas, yakni kue klepon, kue lumpur, kue kastengel, kue putri salju, kue serabi, kue dadar gulung, kue lapis, dan kue risoles. Terdapat sebanyak 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. Beberapa contoh data citra dari 8 kelas dalam dataset kue tradisional Indonesia ditunjukkan pada Tabel 1.

TABEL 1.  
CONTOH DATA CITRA PADA 8 KELAS KUE TRADISIONAL INDONESIA

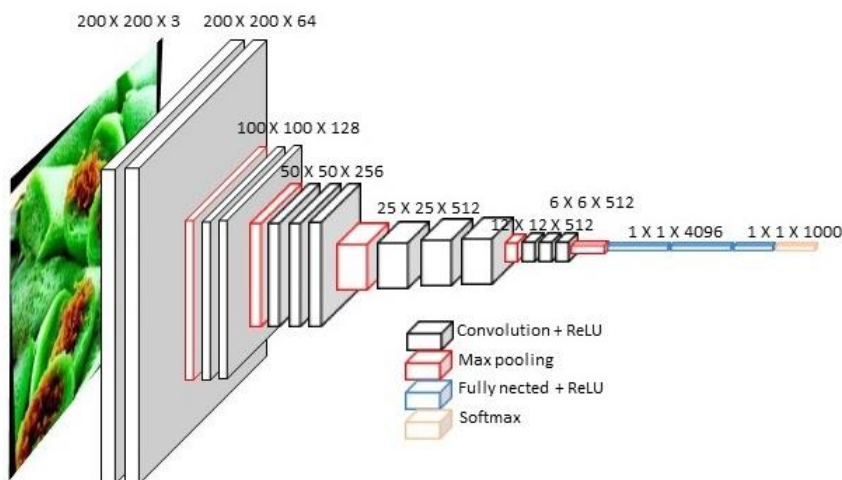
|   |   |  |   |
|---|---|--|---|
|  |  |  |  |
| Kue Klepon<br>(240 Data)  | Kue Lumpur<br>(238 Data)  | Kue Kastengel<br>(223 Data)  | Kue Putri Salju<br>(214 Data)   |
|  |  |  |  |
| Kue Serabi<br>(221 Data)  | Kue Dadar Gulung<br>(232 Data)  | Kue Lapis<br>(241 Data)  | Kue Risoles<br>(236 Data)   |

### B. Preprocessing Data

Pada tahap ini, data citra yang digunakan akan diproses terlebih dahulu dengan mengubah ukuran citra kue tradisional Indonesia menjadi  $256 \times 256$ . Setelah mengubah ukuran citra kue tradisional Indonesia, data citra akan di augmentasi untuk meningkatkan jumlah data citra pada dataset kue tradisional Indonesia. Kemudian, dilakukan split data menjadi 80% data latih dan 20% data uji.

### C. Penerapan Algoritma CNN

1) *Arsitektur VGG16*: Hasil *preprocessing* data sebelumnya akan diimplementasikan ke dalam arsitektur VGG16. Dalam penelitian ini, arsitektur VGG16 yang digunakan terdapat pada Gambar 2.



Gambar 2. Arsitektur VGG16 [12]

Berdasarkan Gambar 2, 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).

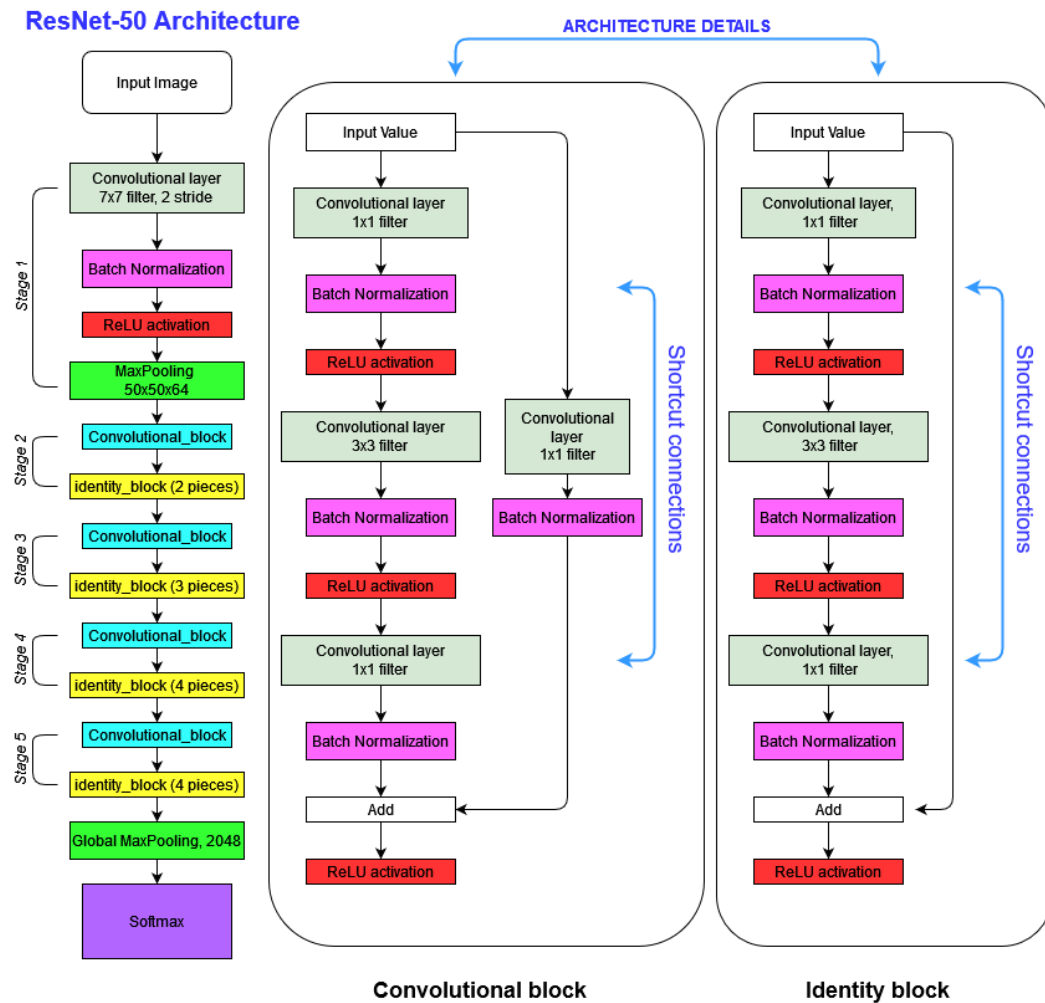
$$f(x_i) = \frac{e^{x_i}}{\sum_{j=1}^N e^{x_j}} \quad (2)$$

Dimana  $x_j$  adalah nilai input pada lapisan softmax dan  $f(x_i)$  adalah peluang 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 probabilitas prediksi,  $d$  adalah target,  $J$  adalah fungsi *cross entropy*,  $k$  adalah label dari data citra. Selain fungsi *loss*, digunakan juga *optimizer*, yakni *Adaptive Moment Optimization* (Adam).

2) *Arsitektur ResNet50*: Selain VGG16, hasil *preprocessing data* juga diimplementasikan ke dalam arsitektur ResNet50. Pada penelitian ini, arsitektur ResNet50 yang digunakan terdapat pada Gambar 3.

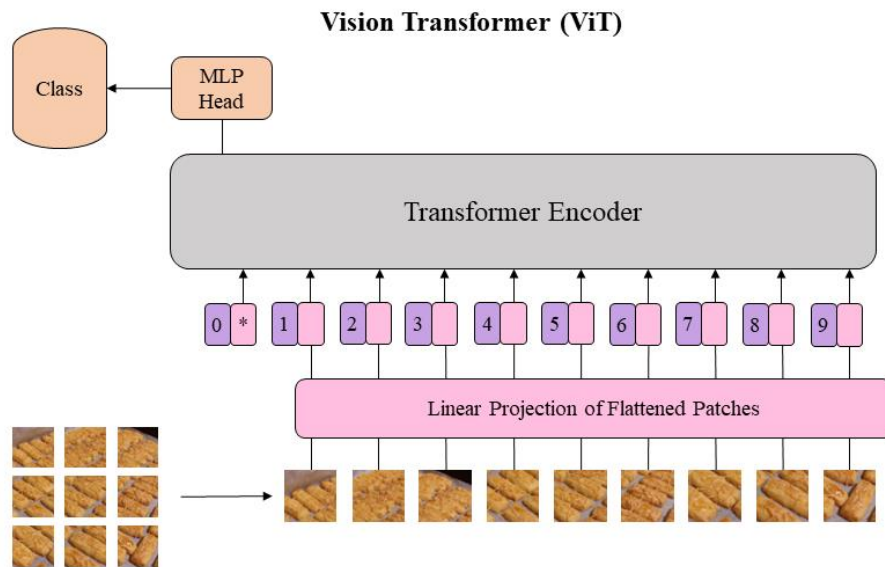


Gambar 3. Arsitektur ResNet50 [24]

Berdasarkan Gambar 3, 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 diterapkan pada arsitektur ResNet50 adalah *categorical cross entropy* dan Adam. Persamaan fungsi *loss categorical cross entropy* dapat dilihat pada Persamaan (3).

#### D. 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 terdapat pada Gambar 2.



Gambar 4. Arsitektur ViT [25]

Dari Gambar 2 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 *mean* atau rata-rata, dan  $\sigma$  merupakan 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]. 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 latih dan 20% data uji. 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 2.

TABEL 2.  
PERBANDINGAN NILAI HYPERPARAMETER 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 2, *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.

#### E. Evaluasi

*Confusion matrix* merupakan tabel yang menyatakan prediksi klasifikasi jumlah *data testing* yang benar dan jumlah *data testing* yang salah [30]. Berikut *confusion matrix* pada klasifikasi dua kelas menurut [31] pada Tabel 3.

TABEL 3.  
CONFUSION MATRIX

| Kelas  |       | Prediksi |       |
|--------|-------|----------|-------|
|        |       | True     | False |
| Aktual | True  | TP       | FN    |
|        | False | FP       | TN    |



Berdasarkan Tabel 3, dalam penaksiran kinerja suatu model dengan *confusion matrix*, terdapat empat istilah, 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 [33], rumus untuk menghitung nilai akurasi, presisi, *recall*, dan *f1-score* terdapat pada Persamaan (7), Persamaan (8), Persamaan (9), dan Persamaan (10).

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

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





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

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

### III. HASIL DAN PEMBAHASAN

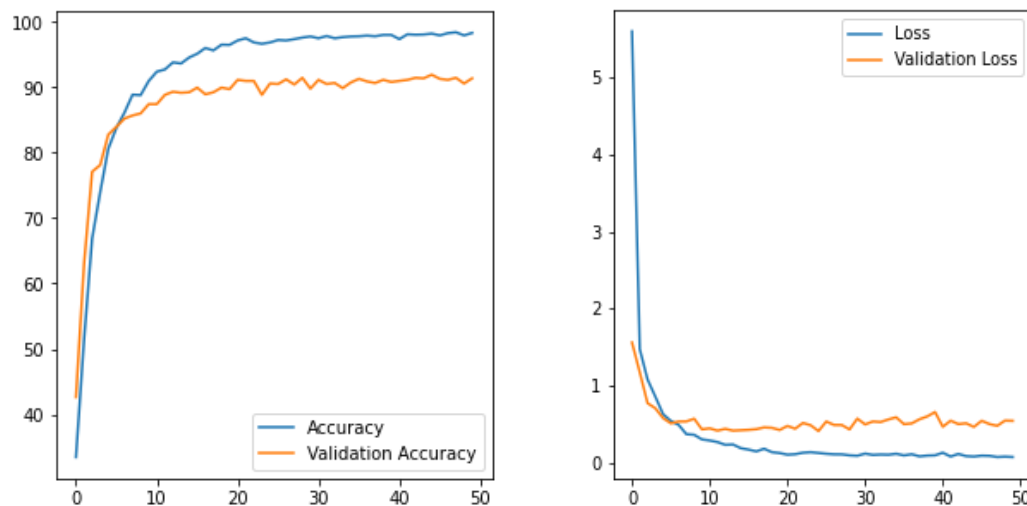
Pada penelitian ini, dataset kue tradisional Indonesia memiliki 8 kelas, yakni kue klepon, kue lumpur, kue kastengel, kue putri salju, kue serabi, kue dadar gulung, kue lapis, dan kue risoles. Dilakukan perubahan ukuran data citra pada dataset kue tradisional Indonesia menjadi  $256 \times 256$ . Setelah itu, melakukan augmentasi data dengan menggunakan *horizontal flip* dan *vertikal flip*. Hasil dari proses augmentasi data kue tradisional Indonesia terdapat pada Tabel 4.

TABEL 4.  
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 4 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 latih dan 20% data uji sehingga terdapat 4222 data latih dan 1055 data uji. 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 latih serta nilai akurasi dan *loss* pada data validasi. Grafik akurasi dan *loss* yang diperoleh pada data latih menggunakan arsitektur VGG16 terdapat pada Gambar 5.





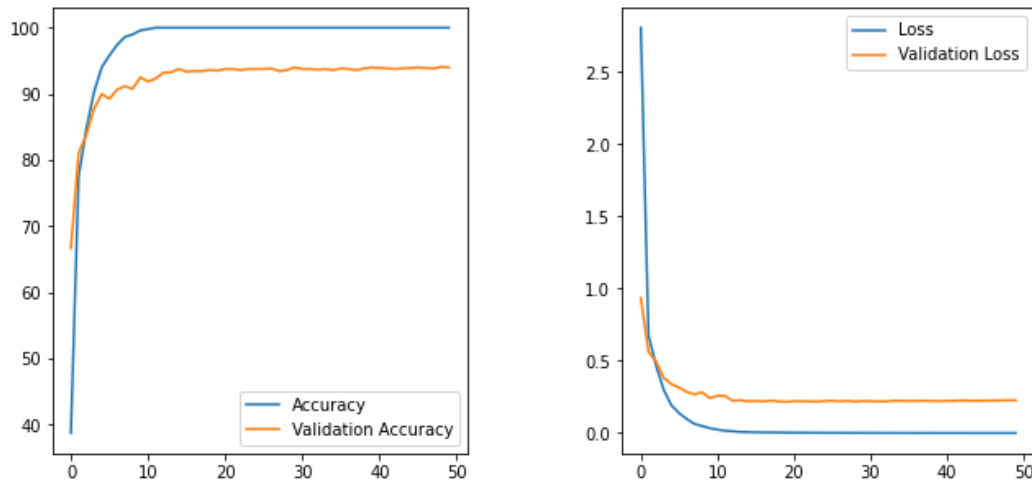
Gambar 5. Grafik Akurasi dan Loss Setiap Epoch VGG16

Berdasarkan Gambar 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* 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 terdapat pada Tabel 5.

TABEL 5.  
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 5, hasil rata-rata akurasi, presisi, *recall*, dan *f1-score* dari klasifikasi dengan 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 latih serta nilai akurasi dan *loss* pada data validasi. Grafik akurasi dan *loss* yang diperoleh pada data latih dapat dilihat pada Gambar 6.



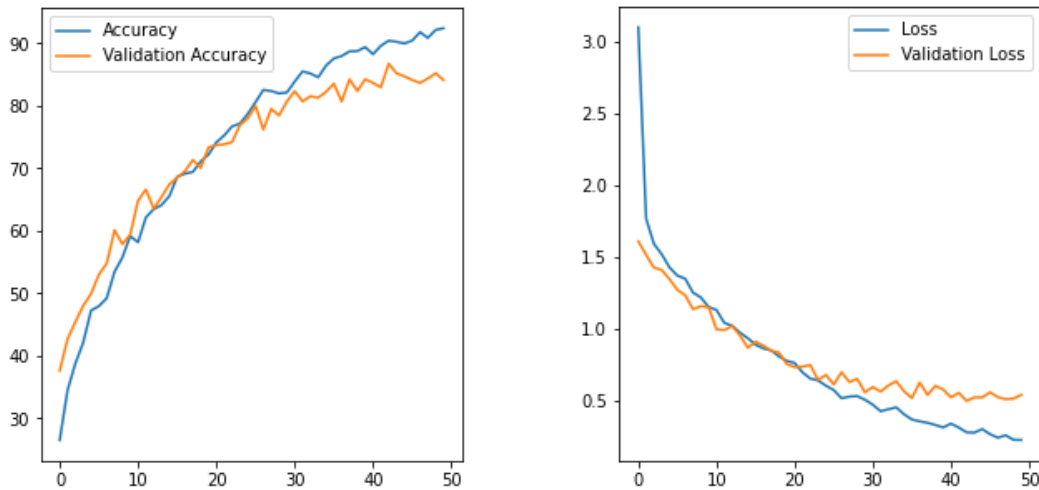
Gambar 6. Grafik Akurasi dan Loss Setiap Epoch ResNet50

Berdasarkan Gambar 6 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*. Nilai akurasi, presisi, *recall*, dan *f1-score* dari klasifikasi dengan arsitektur ResNet50 terdapat pada Tabel 6.

TABEL 6.  
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 6, hasil rata-rata akurasi, presisi, *recall*, dan *f1-score* dari klasifikasi dengan 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 latih serta nilai akurasi dan *loss* pada data validasi. Grafik akurasi dan *loss* yang diperoleh pada data latih arsitektur A algoritma ViT terdapat pada Gambar 7.



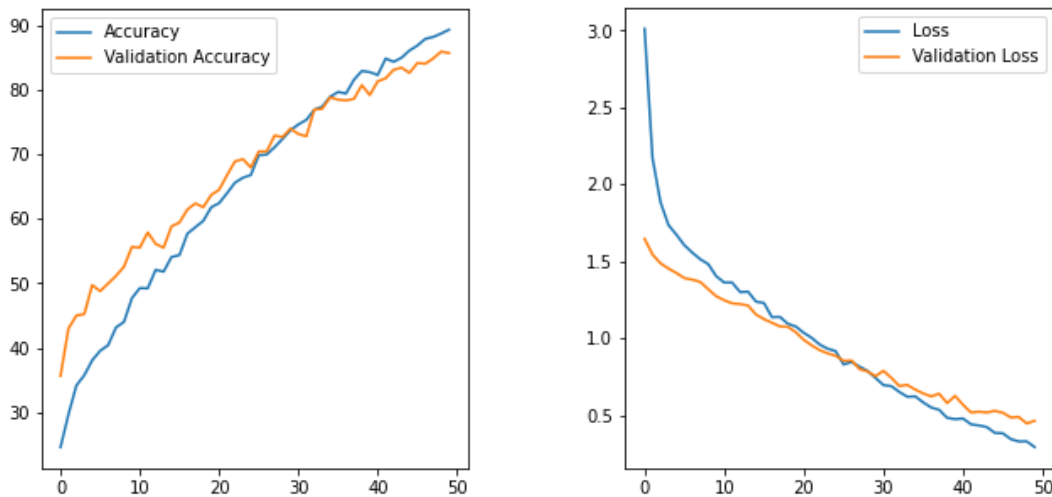
Gambar 7. Grafik Akurasi dan Loss Setiap Epoch Arsitektur A ViT

Berdasarkan Gambar 7 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*. Nilai akurasi, presisi, *recall*, dan *f1-score* dari klasifikasi menggunakan algoritma ViT arsitektur A terdapat pada Tabel 7.

TABEL 7.  
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 7, hasil rata-rata akurasi, presisi, *recall*, dan *f1-score* dari klasifikasi dengan 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 latih serta nilai akurasi dan *loss* pada data validasi. Grafik akurasi dan *loss* yang diperoleh pada data latih terdapat pada Gambar 8.



Gambar 8. Grafik Akurasi dan Loss Setiap Epoch Arsitektur B ViT

Berdasarkan Gambar 8 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*. Nilai akurasi, presisi, *recall*, dan *f1-score* dari klasifikasi dengan algoritma ViT arsitektur B terdapat pada Tabel 8.

TABEL 8.  
HASIL AKURASI, PRESISI, *RECALL*, DAN *F1-SCORE* KUE TRADISIONAL INDONESIA ARSITEKTUR B ViT

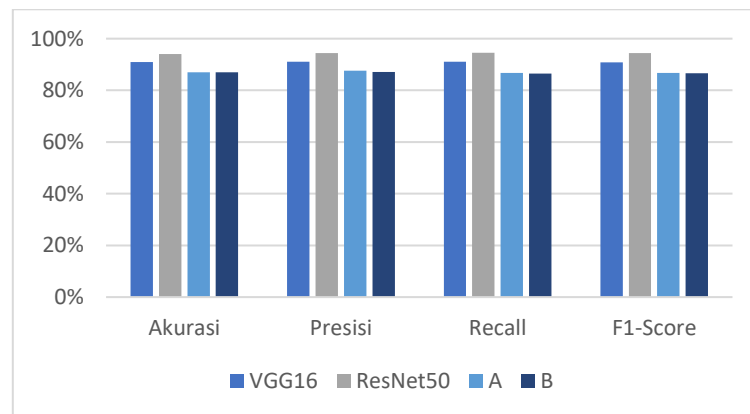
| Nama Kelas       | Akurasi    | Presisi       | <i>Recall</i> | <i>F1-Score</i> |
|------------------|------------|---------------|---------------|-----------------|
| Kue Dadar Gulung | 87%        | 89%           | 86%           | 87%             |
| Kue Kastengel    |            | 85%           | 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 8, hasil rata-rata akurasi, presisi, *recall*, dan *f1-score* dari klasifikasi dengan 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 terdapat pada Tabel 9.

TABEL 9.  
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 9 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. Nilai akurasi, presisi, *recall*, dan *f1-score* dari VGG16, ResNet50, dan ViT dapat dilihat secara ringkas pada Gambar 9.



Gambar 9. Perbandingan Hasil Akurasi, Presisi, *Recall*, dan *F1-Score* Model Kue Tradisional Indonesia

Dari Gambar 9 dapat dilihat bahwa nilai akurasi presisi, *recall*, dan *f1-score* yang dihasilkan 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 kinerja yang lebih baik di antara model lainnya sehingga ResNet50 lebih baik dalam mengklasifikasikan kue tradisional Indonesia dibandingkan model lainnya.

#### IV. SIMPULAN

Berdasarkan penelitian yang dilakukan, didapatkan nilai akurasi, presisi, *recall*, dan *f1-score* menggunakan VGG16, ResNet50, dan ViT berada di atas 80% yang berarti bahwa keempat model memiliki kinerja 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 adalah 94% untuk akurasi dan presisi, 94,50% untuk *recall*, dan 94,38% untuk *f1-score*. Hal ini menunjukkan, bahwa model arsitektur ResNet50 memiliki kinerja yang lebih baik dibandingkan arsitektur lainnya dalam mengklasifikasikan kue tradisional Indonesia.

#### UCAPAN TERIMA KASIH

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

# 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 Deep 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. Sagioglu, "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.