



## Perbandingan Deteksi Alzheimer: ViT, CNN dan ViT dengan Bobot pada Citra Medis

Aisyah Nur Salsabila<sup>1</sup>, Muhaza Liebenlito<sup>2</sup>, Dhea Urfina Zulkifli<sup>3</sup>

aisyah.salsabila18@mhs.uinjkt.ac.id, muhazaliebenlito@uinjkt.ac.id, dhea.urfina@uinjkt.ac.id

<sup>1,2,3</sup> UIN Syarif Hidayatullah Jakarta

---

### Informasi Artikel

Diterima : 13 Feb 2024

Direview : 18 Feb 2024

Disetujui : 29 Feb 2024

---

### Kata Kunci

Deteksi Alzheimer, CNN, Vision Transformer, Transfer Learning, Bobot

---

### Abstrak

Penyakit Alzheimer merupakan suatu tipe demensia yang berpengaruh terhadap ingatan, cara berpikir, dan perilaku. Gejala-gejala tersebut dapat menjadi cukup parah sehingga dapat mempengaruhi kegiatan sehari-hari. Dalam penelitian ini, diperkenalkan aplikasi Convolutional Neural Network (CNN) sederhana dan *pre-trained* model Vision Transformer (ViT) untuk menganalisis data MRI Scan Alzheimer dengan empat kelas, yaitu *Mild Demented*, *Moderate Demented*, *Non-Demented*, dan *Very Mild Demented*. Pada penelitian ini, dilakukan perbandingan pengaplikasian CNN dengan bobot dan ViT yang dilakukan dengan menggunakan dua cara, yaitu dengan bobot dan tidak. Hasil dari penelitian ini menunjukkan bahwa pengaplikasian ViT dengan bobot menghasilkan akurasi yang lebih tinggi dibanding dengan metode lainnya. Dari penelitian ini, diharapkan dapat menganalisa dan mendeteksi penyakit Alzheimer dalam bidang kesehatan dengan efisien.

---

### Keywords

Alzheimer Detection, CNN, Vision Transformer, Transfer Learnin, Weight

---

### Abstract

Alzheimer's disease is a type of dementia that affects memory, the way of thinking and behavior. These symptoms can be severe enough to impact daily activities. In this study, a simple Convolutional Neural Network (CNN) and pre-trained Vision Transformers model are introduced to analyze Alzheimer MRI scan data with four classes: *Mild Demented*, *Moderate Demented*, *Non-Demented*, and *Very Mild Demented*. In this research, a comparison is made between the application of CNN with weights and ViT, performed both with and without weights. The results of the study indicate that applying the pre-trained Vision Transformers model with weights yields higher accuracy compared to other methods. This research is expected to contribute to the efficient analysis and detection of Alzheimer's disease in the field of healthcare.

## A. Pendahuluan

Penyakit Alzheimer atau yang biasa disebut dengan *Alzheimer's Disease*, penyebab paling umum dari demensia, merupakan penyakit neurodegeneratif yang mempengaruhi ingatan, berpikir, bahasa dan keterampilan kognitif yang lain, sehingga mengganggu aktivitas sehari-hari[1].

Berdasarkan penelitian sebelumnya, penderita demensia di seluruh dunia diestimasi mencapai 50 juta orang, yang akan diekspektasi akan meningkat menjadi 152.8 juta pada tahun 2050. Di Indonesia sendiri, terdapat 4.2 juta penderita demensia. Untuk penyakit Alzheimer, data yang diambil di Pulau Jawa dan Bali, prevalensi penyakit Alzheimer melebihi 20%, yang jauh lebih tinggi dibandingkan dengan perhitungan estimasi internasional sebesar 4-9% untuk usia 60 ke atas, yaitu 27,9% [2].

Penyakit Alzheimer ditandai dengan penumpukan plak *amyloid- $\beta$*  dan *tau tangles*. Seseorang yang tidak menderita demensia namun memiliki neuropatologi Alzheimer disebut juga dengan *Non Demented Alzheimer's Disease* (NDAN)[3].

Telah dilakukan penelitian[4] dengan menggunakan *The Clinical Dementia Rating Scale* (CDR), sebuah skala yang digunakan untuk mengklasifikasi tingkat keparahan penyakit Alzheimer sebagai ukuran hasil utama dalam tahap awal penyakit tersebut, yaitu *Mild Cognitive Impairment* (MCI) atau Gangguan Kognitif Ringan dan *mild AD dementia*. Penelitian tersebut dilakukan terhadap 28.220 individu penderita Alzheimer dengan 13.692 (48,5%) merupakan individu yang tidak mengalami gangguan kognitif, 7.075 (25,1%) individu yang mengalami *Mild Cognitive Impairment* (MCI) atau Gangguan Kognitif Ringan, 4.905 (17,4%) penderita *mild dementia*, 1.706 (6%) *moderate dementia*, dan 842 (3%) mengalami *severe AD dementia*.

Hasil penelitian mengenai karakteristik sosiodemografis dan klinis pada kunjungan pertama menunjukkan bahwa proporsi individu yang memerlukan bantuan penuh dalam kehidupan sehari-hari meningkat dari 0,09% dalam kelompok MCI menjadi 0,3% pada MCI, 1,7% pada *mild dementia*, 17,1% *moderate dementia*, dan 71,4% pada *severe AD dementia*. Sebaliknya, proporsi individu yang hidup secara mandiri menurun dari 98,1% dalam kelompok MCI menjadi 58%, 15,2% pada *mild dementia*, 2,3% pada *moderate dementia*, dan 0,4% pada *severe AD dementia*. Sebagian besar dari individu yang menderita *mild AD dementia* yaitu 59% membutuhkan bantuan pada aktivitas kompleks. Oleh karena itu, dapat disimpulkan bahwa semakin memburuknya penyakit Alzheimer yang dialami dalam skala CDR, penderita semakin mengalami hasil yang lebih buruk dalam beraktivitas sehari-hari dan fungsi kognitif, sehingga memerlukan bantuan dan perawatan yang lebih intensif.

Dalam penelitian sebelumnya, metode CNN *end-to-end framework* telah digunakan dalam klasifikasi Alzheimer. *Framework* yang telah dibuat menghasilkan akurasi klasifikasi pada dataset *Alzheimer's disease Neuroimaging Initiative* (ADNI) untuk kelas *binary*, yaitu *Alzheimer's Disease* (AD) dan *Cognitively Normal* (CN). Pada percobaan untuk *multi-classification*, *framework* tersebut meraih akurasi klasifikasi 97,5% pada data yang sama[5].

Namun, berdasarkan penelitian *review* yang dilakukan oleh Vijeeta Patil dan kawan-kawan[6], CNN membatasi deteksi penyakit Alzheimer pada tahap awal penyakit. *Multi-layered CNN* menjadi lebih kompleks ketika mengidentifikasi area

otak orang tua yang terpengaruh. CNN sendiri tidak efektif pada kehilangan memori pada pasien sebagaimana tidak ada indikasi tersebut pada area otak yang sensitif.

Dengan adanya limitasi CNN pada penyakit Alzheimer, Saman Sarraf dan kawan-kawan[7] membuktikan bahwa pengaplikasian *Optimized Vision Transformer (OViTAD)* untuk memprediksi otak sehat, *Mild Cognitive Impairment (MCI)*, dan penyakit Alzheimer menggunakan data rs-MRI dan MRI struktural. Kinerja dari OViTAD dari percobaan tiga kali pengulangan (pemisahan data acak) sebesar  $97\% \pm 0,0$  dan  $99,55\% \pm 0,39$  untuk kedua modalitas pada percobaan klasifikasi multi-kelas, yang melebihi kinerja sebagian besar model deep learning dan CNN yang sudah ada.

Penelitian klasifikasi Alzheimer menggunakan Vision Transformer juga telah dilakukan oleh Maram Fahaad dan kawan-kawan[8] pada dataset Alzheimer *Open Access Series of Imaging Studies-1 (OASIS-1)* yang memiliki empat kelas, yaitu *Mild-Demented (MID)*, *Moderate Demented (MOD)*, *Non-Demented (ND)*, and *Very Mild Demented (VMD)*. Pada penelitian ini menghasilkan bahwa Vision Transformer dapat mengklasifikasi Alzheimer dengan *accuracy* 99,06%, *precision* 99,06%, *recall* 99,14%, dan *F-1 score* 99,1%.

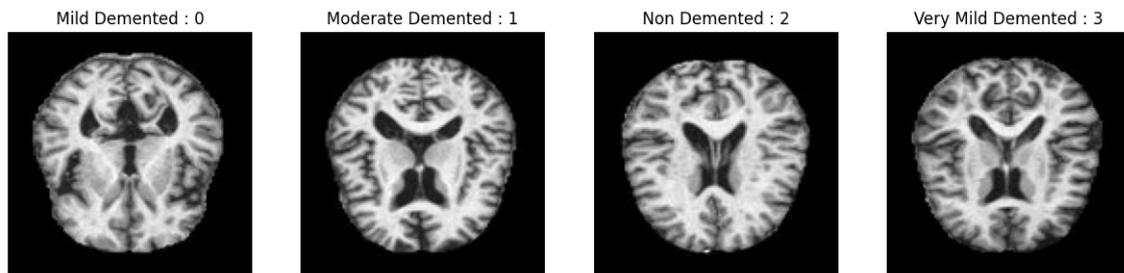
Namun, pernyataan itu bertolak belakang dengan hasil penelitian Hyunji Shin dan kawan-kawan [9] yang mengaplikasikan model *pre-trained ViT* pada data augmentasi Florbetaben PET. Pada kasus binari, akurasi validasi VGG19, yaitu 0,9671, lebih tinggi dibanding akurasi validasi ViT senilai 0,7689. Hal yang sama terjadi pada kasus tiga-kelas, yaitu akurasi validasi VGG19 sebanyak 0,8260 lebih tinggi dibanding akurasi ViT, yaitu 0,7778.

Sebagai pengembangan dari penelitian sebelumnya, penelitian ini melakukan perbandingan kinerja antara model CNN dan ViT dengan penambahan bobot pada setiap kelas di *imbalance data* dalam mendeteksi penyakit Alzheimer yang terdiri atas empat kelas, yaitu *Mild Demented*, *Moderate Demented*, *Non Demented*, dan *Very Mild Demented*. Adapun manfaat dari penelitian adalah untuk mempermudah klasifikasi penyakit Alzheimer sehingga dapat menentukan perawatan yang sesuai dengan kelas Alzheimer yang menimpa penderita.

## B. Metode Penelitian

Penelitian ini menggunakan dataset citra MRI berjumlah 5.120 data yang sudah terlabeli empat kelas, yaitu *Mild Demented* sejumlah 896 data, *Moderate Demented* 64 data, *Non Demented* 3.200 data, dan *Very Mild Demented* 2.240 data. Dataset tersebut dibagi menjadi dua, yaitu data *training* dan data *testing* sebanyak 20%. Pada penerapan CNN dan ViT, dilakukan pengubahan jenis data citra menjadi RGB dan normalisasi terhadap data tersebut. Khusus pada ViT, ukuran dataset diubah menjadi  $224 \times 224$  untuk digunakan pada *pre-trained* model. Data dan kode Python dari penelitian ini dapat diakses di Github[10].

Pada penderita Alzheimer, terjadi penyusutan otak, atau biasa disebut atrofi otak. Seiring memburuknya penyakit Alzheimer, semakin buruk pula gejalanya. Hal ini ditunjukkan dalam pemindaian *longitudinal magnetic resonance*, yang memperlihatkan suatu pola dari atrofi progresif pada *hippocampus*, seperti pelebaran sulci corticalis, perluasan pada ventrikel, penipisan pada kortikal, dan penyusutan pada *hippocampus*[11]. Berikut merupakan contoh sampel dari masing-masing kelas :



**Gambar 1.** Contoh Sampel Data

### **Convolutional Neural Network (CNN)**

*Convolutional Neural Network (CNN)* adalah salah satu tipe *Artificial Neural Network (ANN)* yang merupakan metode *supervised learning*. CNN merupakan metode yang digunakan untuk menemukan dan mengartikan pola pada suatu objek, yang bermanfaat terhadap deteksi pola untuk analisis citra[12].

Pada penelitian ini, *layer* yang digunakan untuk membentuk *framework CNN* terdiri atas *convolutional layers*, di mana filter atau *set kernel* diaplikasikan dengan data *input* citra yang dapat berupa metrik *n*-dimensi untuk menghasilkan *feature map* sebagai *output*.

*Output* pada *convolutional layers* dapat dihasilkan dengan persamaan berikut dengan *h* melambangkan *kernel* dan *f* sebagai *input* citra, sedangkan *m* dan *n* melambangkan jumlah baris dan kolom.

$$G[m, n] = (fx) [m, n] = \sum_j \sum_k h [j, k] f [m - j, n - k] \quad 1$$

Selanjutnya, pada *feature map* digunakan fungsi aktivasi non-linear, yaitu *Rectified Linear Unit (ReLU)*. Fungsi ini dikatakan lebih efisien karena semua *neuron* tidak diaktifkan secara bersamaan, namun hanya sejumlah kecil saja pada satu waktu. *Output* yang dihasilkan dari ReLU adalah nilai negatif pada *feature map* akan diganti menjadi nol sedangkan nilai positif tidak diubah. Hal ini dapat didefinisikan secara matematis sebagai berikut.

$$\text{ReLU}(x) = \max(0, x) \quad 2$$

*Layer* berikutnya yaitu *pooling layer* yang berfungsi untuk melakukan *downsampling* pada *feature map* yang sebelumnya terjadi perubahan terhadap posisi fitur sehingga dapat membuat representasi yang tetap dan tidak berubah terhadap perubahan tersebut. Terdapat dua jenis fungsi paling umum, yaitu *average pooling* dan *maximum pooling*. *Average pooling* sendiri digunakan ketika nilai rata-rata pada tiap *patch* dibutuhkan pada *feature map*. Dalam penelitian ini, digunakan *maximum pooling* di mana nilai maksimum pada setiap *patch* dibutuhkan.

*Layer* terakhir *fully connected layer* yang terhubung ke *node* terakhir pada setiap jaringan untuk meratakan atau *flatten output* pada *layer* sebelumnya. Selanjutnya, *layer* ini menghasilkan probabilitas prediksi kelas menggunakan fungsi non-linear. Pada penelitian ini, fungsi yang digunakan pada *fully connected layer*

adalah *Softmax* dalam *layer* terakhir yaitu *layer* ke- $i + 1$ [13]. Sesuai dengan persamaan

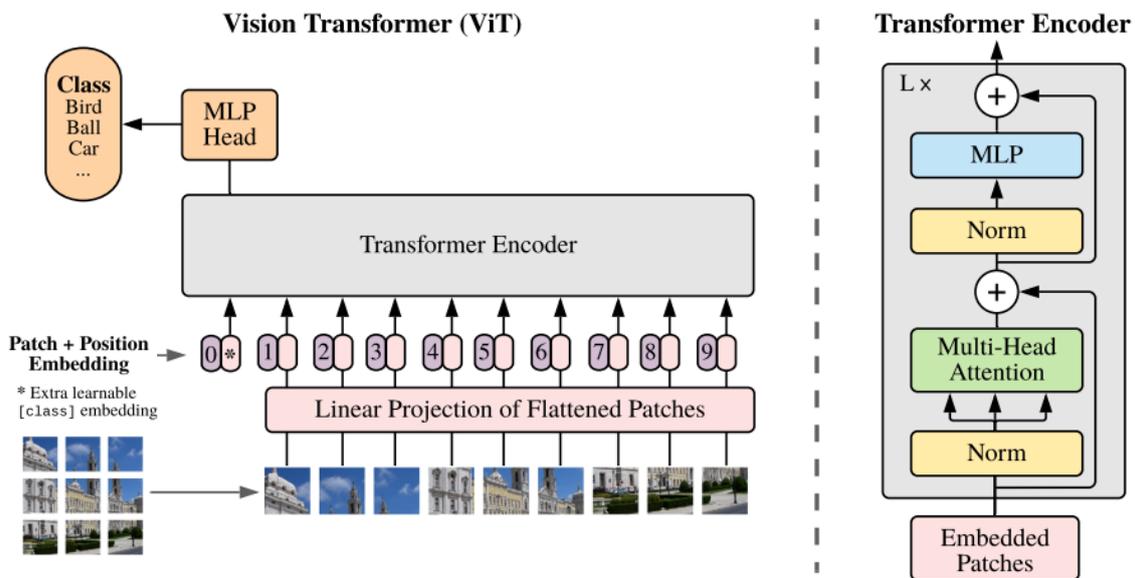
$$f_j(z) = \frac{e^{z_j}}{\sum_{k=1}^n e^{z_k}} \text{ dengan } j = 1, \dots, n \quad 3$$

di mana  $z$  adalah vektor sebarang dengan nilai vektor  $z_j$ , yang dihasilkan pada *layer* ke- $i$  dari CNN dan  $n$  merupakan ukuran vektor[13].

Pada klasifikasi *multiclass*, fungsi *softmax* menghasilkan probabilitas pada setiap data dari semua kelas. Oleh karena itu, dalam *deep neural network* pada kasus *multiclass*, jumlah neuron pada *output layer* akan bersesuaian pada jumlah target kelas.

### **Vision Transformer**

*Vision Transformer* adalah penerapan *transformer* yang berbasis *self-attention* pada data citra[14]. Struktur awal dari ViT dimulai dengan mengubah ukuran citra menjadi ukuran tetap, lalu membagi citra menjadi potongan dan meratakan atau melakukan *flatten* pada potongan-potongan tersebut lalu diubah menjadi urutan *linear embedding*. Hasil dari proses tersebut dinamakan *patch embeddings*. Selanjutnya, *patch embeddings* diproses dalam *transformer encoder*, yang terdiri atas beberapa *layer*, yaitu *Multiheaded Self-Attention* (MSA) dan *Multi-Layer Perceptron* (MLP). ViT menggunakan token kelas untuk ditambahkan di awal urutan *patch embedded* dan *head classification* untuk tugas *pre-training* dan *fine tuning*. Untuk mempertahankan informasi posisi, digunakan *position embeddings* 1D yang dapat di-*train* secara umum.



**Gambar 2.** Arsitektur *Vision Transformer* (ViT)[14]

Dalam penelitian ini, ViT *pre-trained* model yang digunakan adalah ViT *pre-trained* pada Image-Net21k dengan dataset sebanyak 14 juta citra dan 21.843 kelas pada resolusi 224×224.

### **Weighted Cross Entropy**

Ketika suatu dataset memiliki proporsi kelas yang tidak seimbang, model klasifikasi yang digunakan cenderung lebih fokus pada kelas yang memiliki jumlah sampel terbanyak, atau lebih dikenal sebagai *class imbalance*. Untuk mengatasi masalah ini, digunakan *Weighted Cross Entropy* (WCE) dengan  $L_{nWCE}$  sebagai *loss function*[15]. Perbedaan WCE dengan *cross entropy* yang umum digunakan adalah penambahan parameter bobot kelas, yang berbanding terbalik dengan jumlah *pixel* pada tiap kelas untuk menyeimbangkan pengaruh pada setiap kelas tersebut. Hal ini dapat didefinisikan sebagai persamaan berikut dengan  $\omega_i$  parameter bobot kelas yang berbanding terbalik dengan jumlah *pixel* di setiap kelas.

$$L_{nWCE} = - \sum_i^c \omega_i y_{i,n} \log(z_{i,n}) \quad 4$$

### **Performance Metric**

Dalam penelitian ini, kinerja model yang digunakan dapat dinilai dengan beberapa *performance metric* yang ditunjukkan pada *confusion matrix*, yaitu *accuracy*, *recall*, *precision*, dan *F1-Score*[5]. Untuk menghitung *performance metric*, diperlukan variabel *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN).

*Accuracy* merupakan nilai dari semua data yang terprediksi positif berbanding dengan jumlah semua data

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad 5$$

*Recall* adalah nilai perbandingan antara jumlah data terprediksi positif yang benar dengan semua data yang benar positif

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad 6$$

Selanjutnya, *precision* adalah rasio antara jumlah data terprediksi positif yang benar dengan semua data yang terprediksi positif

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad 7$$

Kombinasi antara *precision* dan *recall* sehingga membentuk menjadi satu parameter disebut *F1-Score*, seperti persamaan berikut.

$$F1_{\text{Score}} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \quad 8$$

### ***Transfer Learning***

*Transfer learning* adalah teknik yang memanfaatkan suatu model yang telah dilatih pada dataset lain. *Transfer learning* bertujuan untuk meningkatkan kemampuan belajar pada data yang baru dengan memanfaatkan pengetahuan dari sumber dan *learning task* sebelumnya. Perbedaan penggunaan *transfer learning* didefinisikan berdasarkan jenis *task* dan sifat data yang di domain sumber dan target[16].

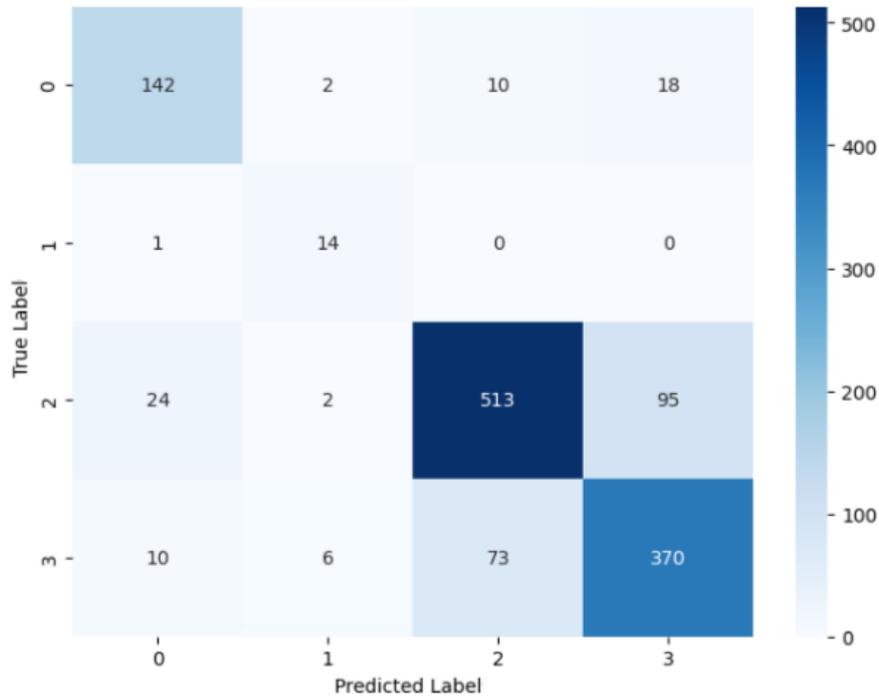
Dalam bidang *machine learning*, *transfer learning* berhasil mengungguli metode yang sudah ada. Kinerja yang meningkat, khususnya pada *computer vision*, mendorong penggunaan deep learning dalam diagnosis dan prediksi berbasis komputer. *Transfer learning* dapat digunakan dengan sedikit atau tanpa informasi dalam fase *training*, hal ini yang membuat *transfer learning* menjadi lebih efisien. Dengan kata lain, data yang sudah teruji dapat disesuaikan melalui *transfer learning* dari suatu domain ke domain lain.

### **C. Hasil dan Pembahasan**

Pada penelitian ini, jumlah dataset yang digunakan menunjukkan bahwa data tersebut merupakan data yang *imbalance*. Oleh karena itu, dilakukan penambahan bobot pada masing-masing kelas sebagai parameter *cross entropy*.

Hasil yang diperoleh dari CNN dengan bobot menghasilkan *accuracy* sebesar 81,1%, *recall* 81,1%, *precision* 81,5%, dan *F1-Score* 81,2%. Dengan nilai *recall* 81,1%, maka hal ini menyatakan bahwa kinerja model CNN untuk mendeteksi suatu jenis Alzheimer jika kelas tersebut ada sebesar 81,1%, sedangkan nilai *precision* 81,5% diartikan sebagai probabilitas model CNN memprediksi jenis Alzheimer dengan benar.

Berikut merupakan *confusion matrix* yang diperoleh dari model CNN dengan bobot.



**Gambar 3.** *Confusion Matrix* CNN dengan bobot

Pada Gambar 3, ditunjukkan seberapa banyak data pada masing-masing kelas yang terdeteksi benar dan tidak. Jumlah data yang terdeteksi benar dapat dilihat pada nilai diagonal dari *matrix*, sedangkan nilai yang lain menunjukkan jumlah data yang terdeteksi salah. Dari *confusion matrix* tersebut, untuk melihat kemampuan model CNN dalam mendeteksi kelas Alzheimer, dapat dibuat *performance metric* pada masing-masing kelas seperti berikut

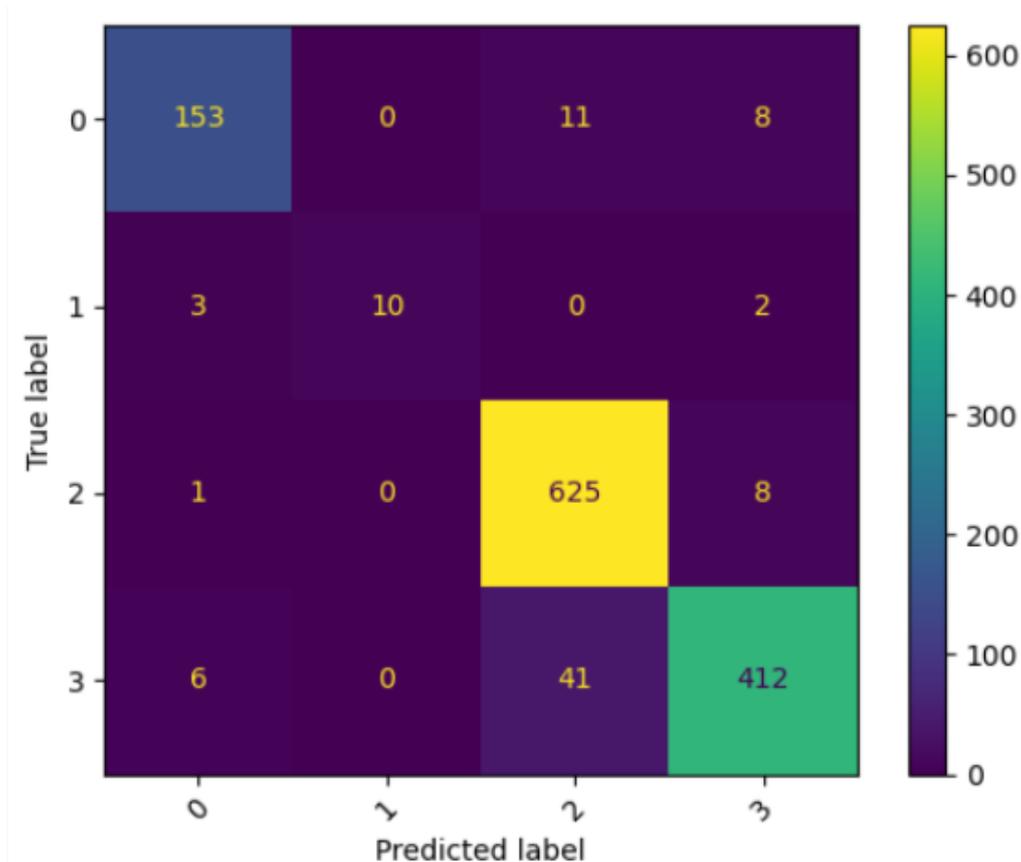
**Tabel 1.** *Performance Metric* dengan CNN

No	Kelas	Accuracy	Recall	Precision	F1-Score
1	<i>Mild Demented</i>	0,949	0,825	0,802	0,813
2	<i>Moderate Demented</i>	0,991	0,933	0,583	0,717
3	<i>Non Demented</i>	0,840	0,809	0,860	0,834
4	<i>Very Mild Demented</i>	0,842	0,806	0,766	0,785

Berdasarkan Tabel 1, terlihat nilai *precision* pada kelas *Moderate Demented* lebih rendah secara signifikan dibandingkan dengan nilai kelas lain yaitu 58.3%. Hal ini menjelaskan bahwa penderita Alzheimer *Moderate Demented* akan cenderung tidak mendapatkan perawatan intensif yang seharusnya mereka dapatkan dikarenakan adanya potensi kekeliruan yang lebih besar dalam mendeteksi Alzheimer kelas *Moderate Demented*.

Berbeda dari CNN dengan bobot, hasil dari menggunakan ViT menunjukkan *accuracy* 93,8%, *precision* 93,9%, *recall* 93,8%, dan *F1-Score* 93,8%. Hasil *recall* tersebut menunjukkan persentase suatu kelas yang terprediksi dengan benar dari seluruh data kelas tersebut adalah 93,8% dan model mampu memprediksi kelas Alzheimer dengan benar sebesar 93,9% berdasarkan nilai *precision*.

Berikut merupakan *confusion matrix* yang diperoleh dari ViT tanpa bobot.



**Gambar 4.** Confusion Matrix ViT tanpa bobot

Pada Gambar 4, kelas *Moderate Demented* mengalami penurunan jumlah data yang terdeteksi benar sedangkan kelas lain mengalami kenaikan. Hal ini dapat dilihat lebih lanjut pada *performance metric* berikut.

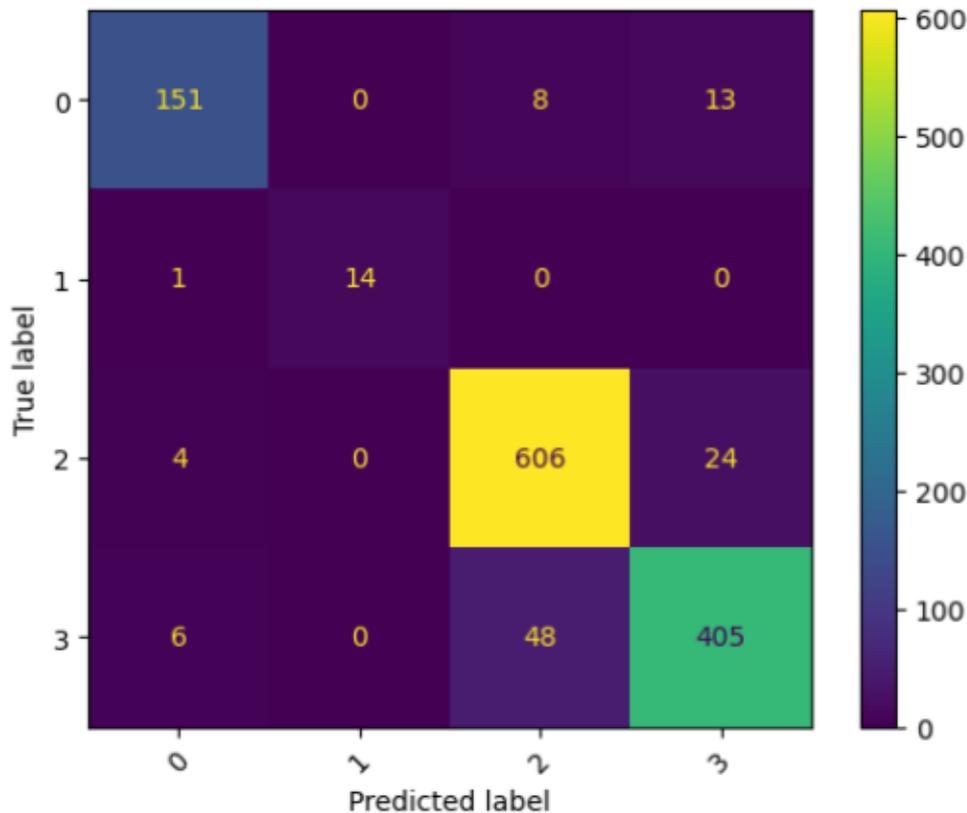
**Tabel 2.** Performance Metric dengan ViT

No	Kelas	Accuracy	Recall	Precision	F1-Score
1	<i>Mild Demented</i>	0,977	0,889	0,938	0,913
2	<i>Moderate Demented</i>	0,996	0,667	1,000	0,800
3	<i>Non Demented</i>	0,952	0,985	0,923	0,953
4	<i>Very Mild Demented</i>	0,949	0,897	0,958	0,926

Berdasarkan penurunan jumlah data yang terdeteksi benar pada kelas *Moderate Demented* pada Tabel 2, hal ini menyebabkan menurunnya pada nilai *recall*, yang menunjukkan persentase kelas yang terprediksi benar dari seluruh data *Moderate Demented* sebesar 66,7%. Seperti hasil pada model CNN, potensi kekeliruan dalam mendeteksi Alzheimer kelas *Moderate Demented* menjadi lebih besar yang akan mengakibatkan tidak tersalurkannya perawatan intensif yang dibutuhkan oleh penderita.

Pada model ViT dengan bobot, hasil *performance metric* mengalami penurunan dibandingkan ViT tanpa bobot, yaitu dihasilkan *accuracy* 92%, *precision* 92,1%, *recall* 92%, dan *F1-Score* 91,9%. Pernyataan tersebut dapat diartikan kinerja model dalam memprediksi suatu kelas dengan benar dari keseluruhan data kelas

tersebut sebesar 92% dan mendeteksi kelas Alzheimer dengan benar sebanyak 92,1%.



**Gambar 5.** Confusion Matrix ViT dengan bobot

Meskipun mengalami penurunan pada *performance metric*, terlihat pada Gambar 5 bahwa data kelas *Moderate Demented* cenderung lebih banyak terprediksi dengan benar pada ViT dengan bobot dibandingkan dengan ViT tanpa bobot.

**Tabel 3.** Performance Metric ViT dengan bobot

No	Kelas	Accuracy	Recall	Precision	F1-Score
1	<i>Mild Demented</i>	0,975	0,877	0,932	0,904
2	<i>Moderate Demented</i>	0,999	0,933	1,000	0,965
3	<i>Non Demented</i>	0,934	0,955	0,915	0,935
4	<i>Very Mild Demented</i>	0,928	0,882	0,916	0,899

Berdasarkan Tabel 3, dihasilkan nilai yang tidak berbeda secara signifikan dibanding dua metode sebelumnya. Hal ini menunjukkan kemampuan model ViT dalam mendeteksi Alzheimer pada masing-masing kelas cukup akurat. Hal ini menjadi poin penting dalam mendeteksi kelas Alzheimer dengan tepat dikarenakan adanya pengaruh terhadap penanganan penderita pada masing-masing kelas Alzheimer. Semakin tinggi tingkatan kelasnya, semakin intensif perawatan dan penanganan yang diperlukan dalam melakukan aktivitas sehari-hari.

#### D. Simpulan

Pada penelitian ini, dilakukan perbandingan antara CNN dengan bobot, ViT tanpa bobot, dan ViT dengan bobot dalam mendeteksi penyakit Alzheimer dalam empat kelas, yaitu *Mild Demented*, *Moderate Demented*, *Non Demented*, dan *Very Mild Demented*. Pada pengaplikasian ViT, digunakan *pre-trained* model dengan menggunakan metode *transfer learning* pada data yang diolah. Dalam analisis performa masing-masing model, digunakan *confusion matrix* serta *performance metric* yang terdiri atas *recall*, *precision*, *accuracy*, dan *F1-Score*.

Hasil dari penelitian ini menunjukkan bahwa dengan menggunakan ViT menghasilkan *performance metric* tertinggi yaitu *accuracy* 93,8%, *recall* 93,8%, *precision* 93,9%, dan *F1-Score* 93,7%. Urutan kedua yaitu ViT dengan bobot dengan *accuracy* 92%, *recall* 92%, *precision* 92,1%, dan *F1-Score* 91,9%. Terakhir adalah CNN dengan bobot dengan *accuracy* 81%, *recall* 81,1%, *precision* 81,5% dan *f1-score* 81,2%

Meskipun *performance metric* pada ViT lebih tinggi, namun pada kelas *Moderate Demented* *recall* yang dihasilkan adalah 66,7% yang cenderung lebih rendah dibanding ViT dengan bobot pada kelas yang sama yaitu 93,3%. Hal ini menunjukkan bahwa ViT dengan bobot mampu mendeteksi Alzheimer kelas *Moderate Demented* dengan persentase sebesar 93,3% dari seluruh kelas tersebut, sedangkan ViT memiliki persentase yang lebih rendah yaitu sebanyak 66,7%.

Berdasarkan ketiga metode ini, kemampuan ViT dengan bobot dalam mendeteksi keempat kelas tersebut cukup akurat sehingga penanganan dan perawatan terhadap penderita Alzheimer mampu didapatkan sesuai dengan kelas Alzheimer yang diderita.

Saran untuk penelitian selanjutnya adalah dapat menerapkan augmentasi citra pada data, khususnya pada data *imbalance*, dan melakukan melakukan eksplorasi pada model maupun *pre-trained* model terbaru guna meningkatkan kinerja dalam memprediksi kelas Alzheimer.

#### E. Ucapan Terima Kasih

Pada kesempatan ini, penulis ingin mengucapkan terima kasih sebanyak-banyaknya kepada Prodi Matematika Universitas Islam Negeri Syarif Hidayatullah Jakarta yang telah memberikan kesempatan dan fasilitas dalam terbitnya artikel ini.

#### F. Referensi

- [1] "2022 Alzheimer's disease facts and figures," *Alzheimer's & Dementia*, vol. 18, no. 4, pp. 700–789, Apr. 2022, doi: 10.1002/alz.12638.
- [2] N. Farina *et al.*, "Comprehensive measurement of the prevalence of dementia in low- and middle-income countries: STRiDE methodology and its application in Indonesia and South Africa," *BJPsych Open*, vol. 9, no. 4, p. e102, Jul. 2023, doi: 10.1192/bjo.2023.76.
- [3] F. K. Kok, S. L. van Leerdam, and E. C. M. de Lange, "Potential Mechanisms Underlying Resistance to Dementia in Non-Demented Individuals with Alzheimer's Disease Neuropathology," *Journal of Alzheimer's Disease*, vol. 87, no. 1, pp. 51–81, May 2022, doi: 10.3233/JAD-210607.

- [4] K. L. Lanctôt *et al.*, "Association between clinical dementia rating and clinical outcomes in Alzheimer's disease," *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, vol. 16, no. 1, Jan. 2024, doi: 10.1002/dad2.12522.
- [5] Y. AbdulAzeem, W. M. Bahgat, and M. Badawy, "A CNN based framework for classification of Alzheimer's disease," *Neural Comput Appl*, vol. 33, no. 16, pp. 10415–10428, Aug. 2021, doi: 10.1007/s00521-021-05799-w.
- [6] V. Patil, M. Madgi, and A. Kiran, "Early prediction of Alzheimer's disease using convolutional neural network: a review," *Egyptian Journal of Neurology, Psychiatry and Neurosurgery*, vol. 58, no. 1. Springer Science and Business Media Deutschland GmbH, Dec. 01, 2022. doi: 10.1186/s41983-022-00571-w.
- [7] S. Sarraf, A. Sarraf, D. D. DeSouza, J. A. E. Anderson, and M. Kabia, "OViTAD: Optimized Vision Transformer to Predict Various Stages of Alzheimer's Disease Using Resting-State fMRI and Structural MRI Data," *Brain Sci*, vol. 13, no. 2, Feb. 2023, doi: 10.3390/brainsci13020260.
- [8] M. F. Almufareh, S. Tehsin, M. Humayun, and S. Kausar, "Artificial Cognition for Detection of Mental Disability: A Vision Transformer Approach for Alzheimer's Disease," *Healthcare*, vol. 11, no. 20, p. 2763, Oct. 2023, doi: 10.3390/healthcare11202763.
- [9] H. Shin, S. Jeon, Y. Seol, S. Kim, and D. Kang, "Vision Transformer Approach for Classification of Alzheimer's Disease Using 18F-Florbetaben Brain Images," *Applied Sciences (Switzerland)*, vol. 13, no. 6, Mar. 2023, doi: 10.3390/app13063453.
- [10] A. N. Salsabila, "Alzheimer Classification Repository," GitHub, 2024. [Online]. Available: [https://github.com/aisyahbil/Alzheimer\\_classification](https://github.com/aisyahbil/Alzheimer_classification)
- [11] T. C. Harris, R. de Rooij, and E. Kuhl, "The Shrinking Brain: Cerebral Atrophy Following Traumatic Brain Injury," *Ann Biomed Eng*, vol. 47, no. 9, pp. 1941–1959, Sep. 2019, doi: 10.1007/s10439-018-02148-2.
- [12] A. W. Salehi *et al.*, "A Study of CNN and Transfer Learning in Medical Imaging: Advantages, Challenges, Future Scope," *Sustainability*, vol. 15, no. 7, p. 5930, Mar. 2023, doi: 10.3390/su15075930.
- [13] I. Kouretas and V. Paliouras, "Hardware Implementation of a Softmax-Like Function for Deep Learning," *Technologies (Basel)*, vol. 8, no. 3, p. 46, Aug. 2020, doi: 10.3390/technologies8030046.
- [14] A. Dosovitskiy *et al.*, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," Oct. 2020.
- [15] M. Sazzad Hossain, J. M. Betts, and A. P. Paplinski, "Corrigendum to 'Dual Focal Loss to address class imbalance in semantic segmentation' [Neurocomputing 462 (2021) 69-87]," *Neurocomputing*, vol. 477, p. 61, Mar. 2022, doi: 10.1016/j.neucom.2022.01.009.
- [16] J. A. L. Marques, F. N. B. Gois, J. P. do V. Madeiro, T. Li, and S. J. Fong, "Artificial neural network-based approaches for computer-aided disease diagnosis and treatment," in *Cognitive and Soft Computing Techniques for the Analysis of Healthcare Data*, Elsevier, 2022, pp. 79–99. doi: 10.1016/B978-0-323-85751-2.00008-6.