JITET (Jurnal Informatika dan Teknik Elektro Terapan)
Vol. 13 No. 3S1, pISSN: 2303-0577 eISSN: 2830-7062 http://dx.doi.org/10.23960/jitet.v13i3S1.7888

PENDETEKSIAN DINI STUNTING PADA BALITA MENGGUNAKAN VISION TRANSFORMER (ViT) BERBASIS CITRA TUBUH

Reyvan Revolusioner Ar[1*], Agusriyati[2], Sumiarni Moka[3]

[1]Jurusan Teknik Informatika, Universitas Haluoleo; Jalan H.E.A. Mokodompit, Kota Kendari, Sulawesi Tenggara 93232; 0401-3194108

Keywords:
Stunting, Vision Transformer (ViT), Klasifikasi Citra.

Corespondent Email:
reyvanrevolusionerar@gmail.com

Abstrak. Penelitian ini bertujuan untuk mengembangkan sistem pendeteksian dini stunting berbasis citra tubuh balita menggunakan arsitektur Vision Transformer (ViT). Dataset terdiri atas 2.156 citra tubuh balita yang terbagi dalam tiga subset: pelatihan, validasi, dan pengujian. Citra dipraproses melalui konversi ke RGB, pengubahan ukuran menjadi 224×224 piksel, serta normalisasi menggunakan ViTImageProcessor. Model ViT-base dilatih selama lima epoch menggunakan optimizer AdamW dan batch size 8. Evaluasi dilakukan menggunakan confusion matrix dan classification report. Hasil evaluasi menunjukkan akurasi model sebesar 98%, dengan precision dan recall rata-rata masing-masing sebesar 0,98. Visualisasi attention map ditampilkan melalui antarmuka Gradio untuk menunjukkan area fokus model dalam proses klasifikasi. Sistem ini memberikan solusi alternatif pendeteksian stunting yang efisien, interpretatif, dan aplikatif, terutama untuk wilayah yang minim akses terhadap alat ukur dan tenaga medis. Hasil penelitian menunjukkan bahwa ViT memiliki performa unggul dalam klasifikasi citra tubuh balita dan berpotensi untuk diterapkan dalam deteksi status gizi secara otomatis dan adaptif.

Abstract. This study aims to develop an early detection system for stunting based on toddler body images using the Vision Transformer (ViT) architecture. The dataset consists of 2,156 toddler body images divided into three subsets: training, validation, and testing. The images were preprocessed by converting to RGB, resizing to 224×224 pixels, and normalizing using ViTImageProcessor. The ViT-base model was trained for five epochs using the AdamW optimizer and batch size 8. Evaluation was carried out using a confusion matrix and classification report. The evaluation results showed a model accuracy of 98%, with an average precision and recall of 0.98 each. The attention map visualization is displayed through the Gradio interface to show the model's focus area in the classification process. This system provides an alternative solution for stunting detection that is efficient, interpretative, and applicable, especially for areas with minimal access to measuring instruments and medical personnel. The results of the study show that ViT has superior performance in classifying toddler body images and has the potential to be applied in automatic and adaptive nutritional status detection.

1. PENDAHULUAN

Stunting merupakan salah satu permasalahan kesehatan global yang masih menjadi tantangan utama di negara berkembang, termasuk Indonesia. Kondisi ini diakibatkan oleh kekurangan gizi kronis yang terjadi sejak masa kehamilan hingga dua tahun pertama kehidupan anak, atau yang dikenal sebagai periode 1.000 Hari Pertama Kehidupan (HPK). Anak yang mengalami stunting umumnya menunjukkan tinggi badan di bawah standar usianya serta mengalami keterlambatan dalam perkembangan kognitif dan motorik.[1]

Menurut data Kementerian Kesehatan Republik Indonesia, prevalensi stunting nasional menurun dari 21,5% pada tahun 2023 menjadi 19,8% pada tahun 2024[2]. Namun demikian, angka tersebut masih berada di atas ambang batas yang ditetapkan oleh *World Health Organization* (WHO), yaitu 20%, untuk dikategorikan sebagai masalah kesehatan masyarakat yang seriu[3].Pemerintah Indonesia menargetkan penurunan angka stunting menjadi 14,2% pada tahun 2029 melalui program nasional yang sejalan dengan target *Sustainable Development Goals* (SDGs) 2030.

Berbagai penelitian sebelumnya telah berupaya mengembangkan sistem pendeteksian stunting berbasis machine learning dengan menggunakan data antropometri seperti tinggi badan, berat badan, umur, dan jenis kelamin. Ratnasari et al. (2024) berhasil membangun model deteksi stunting menggunakan algoritma *Random Forest*, *K-Nearest Neighbor* (KNN), *Naive Bayes*, dan SVM, di mana akurasi tertinggi mencapai 92,70% dengan *Random Forest*[4], Sementara itu, Febriyanti et al. (2025) membandingkan performa SVM, *Random Forest*, dan *Logistic Regression* dengan menggunakan dataset sebanyak 2.231 balita dan menunjukkan bahwa SVM memiliki akurasi 92%, *precision* 91%, recall 99%, f1-score 95%, dan AUC 99%[5]. Meskipun pendekatan ini menunjukkan hasil yang baik, seluruhnya masih bergantung pada input numerik dari pengukuran antropometri yang memerlukan alat ukur dan keterampilan teknis, serta rawan terhadap kesalahan manusia terutama di daerah dengan sumber daya terbatas.

Kemajuan teknologi dalam bidang kecerdasan buatan, khususnya pada pengolahan citra digital dan deep learning, memberikan peluang baru dalam pendeteksian stunting yang lebih adaptif dan efisien. *Vision Transformer* (ViT), sebagai arsitektur berbasis self-attention dalam domain visi komputer, telah menunjukkan performa unggul dalam klasifikasi citra medis. ViT mampu mengenali pola visual secara global dengan lebih baik dibanding arsitektur CNN tradisional[6]. Studi oleh Putri dan Harahap (2023) melaporkan bahwa ViT-B16 yang dilatih menggunakan citra retina berukuran 224×224 piksel menghasilkan akurasi sebesar 92,86%, precision 100%, recall 85,72%, dan F1-score 92,31% [7]. Penelitian lain juga menunjukkan bahwa ViT mampu mencapai akurasi 97,61% dalam klasifikasi citra X-ray dengan sensitivitas 95% dan spesifisitas 98%, menjadikannya sebagai arsitektur yang kompetitif dalam diagnosis berbasis citra[8] Selain itu, penelitian oleh Hanif et al. (2023) membuktikan keberhasilan model vision-based dalam tugas ekstraksi karakter plat nomor kendaraan menggunakan pendekatan OCR, memperkuat urgensi pengembangan sistem berbasis citra dalam domain teknis maupun medis[9].

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk mengembangkan metode pendeteksian dini stunting pada balita menggunakan arsitektur *Vision Transformer* (ViT) berbasis citra tubuh. Penelitian ini juga bertujuan untuk mengevaluasi akurasi dan efektivitas model ViT sebagai alternatif dari metode konvensional berbasis antropometri. Diharapkan pendekatan ini dapat memberikan solusi pendeteksian stunting yang lebih efisien, adaptif, dan aplikatif, khususnya pada daerah yang memiliki keterbatasan akses terhadap alat ukur dan tenaga medis.

2. TINJAUAN PUSTAKA

2.1 Stunting

Stunting adalah kondisi gangguan pertumbuhan yang ditandai dengan tinggi badan anak yang tidak sesuai dengan usianya akibat kekurangan gizi kronis yang berlangsung dalam jangka panjang. Kondisi ini umumnya terjadi sejak masa kehamilan hingga usia dua tahun dan dapat berdampak pada perkembangan fisik, kognitif, serta produktivitas anak di masa mendatang. Selain faktor gizi, sanitasi yang buruk, infeksi berulang, dan kurangnya edukasi tentang pola

asuh juga turut berkontribusi terhadap terjadinya stunting. Deteksi dini menjadi penting untuk mencegah dampak jangka panjang, namun pendekatan manual yang bergantung pada pengukuran antropometri sering kali tidak efisien di lapangan.[10]

2.2 Vision Transformer (ViT)

Vision Transformer (ViT) merupakan model deep learning yang mengadaptasi arsitektur Transformer dari bidang pemrosesan bahasa alami untuk tugas pengolahan citra. Berbeda dengan model konvolusional (CNN) yang mengandalkan filter lokal, ViT memproses gambar dengan membaginya menjadi patch kecil berukuran tetap, kemudian setiap patch diperlakukan seperti token dalam teks. ViT menggunakan mekanisme self-attention untuk memahami hubungan antar bagian gambar secara menyeluruh, sehingga mampu mengenali pola visual global yang kompleks. Model ini terbukti efektif dalam berbagai tugas klasifikasi gambar, termasuk citra medis, dengan performa yang kompetitif bahkan melebihi CNN pada dataset skala besar[11].

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan eksperimen dengan metode klasifikasi citra berbasis deep learning. Model yang digunakan adalah *Vision Transformer* (ViT), yang bertujuan untuk mengklasifikasikan gambar tubuh balita ke dalam dua kategori: *stunting* dan *normal*. Adapun Tahaoan metode penelitian sebagai berikut:

3.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini berupa citra tubuh balita yang terbagi ke dalam dua kategori, yaitu *normal* dan *stunting*. Data diperoleh dari sumber internal dan dikelompokkan ke dalam dua direktori, yakni direktori pelatihan (*train*) dan direktori validasi (*valid*). Penamaan file gambar disesuaikan untuk memungkinkan pelabelan otomatis berdasarkan isi nama file. Pendekatan ini memudahkan proses klasifikasi awal menggunakan skrip pemrograman berbasis Python[12].

3.2. Prapemrosesan Citra

Setelah data dikumpulkan, di lakukan pramprosesan citra yang dikonversi ke format RGB, kemudian diubah ukurannya menjadi 224×224 piksel dan dinormalisasi menggunakan *ViTImageProcessor*. Proses praproses ini bertujuan untuk memastikan keseragaman ukuran dan skala nilai piksel, sehingga gambar dapat diproses secara optimal oleh model ViT. Teknik serupa banyak digunakan dalam penelitian klasifikasi citra pada domain medis maupun umum[6].

3.3. Pelatihan Model

Model yang digunakan adalah *Vision Transformer ViT-base-patch* 16-224 yang di-fine-tune menggunakan data citra tubuh balita. Pelatihan dilakukan selama lima epoch menggunakan *optimizer* Adam W dan batch size 8 dengan bantuan pustaka transformers dan Trainer dari *Hugging Face*. Arsitektur ViT dipilih karena kemampuannya mengenali pola global pada citra dengan mekanisme *self-attention*

3.4. Evaluasi Model

Setelah proses pelatihan model selesai dilakukan, tahap selanjutnya adalah mengevaluasi kinerjanya. Tujuan dari evaluasi ini adalah untuk menilai sejauh mana model mampu bekerja secara optimal. Dalam penelitian ini, evaluasi dilakukan terhadap data uji menggunakan metode confusion matrix. Melalui metode ini diperoleh nilai metrik seperti akurasi, presisi, recall, dan f1-score. Nilai-nilai tersebut menjadi acuan dalam menentukan apakah model perlu dilatih ulang apabila performanya belum memenuhi harapan[13].

Adapun rumus untuk precision, recall, dan f1-score dijelaskan sebagai berikut:

$$\text{Precision} = \frac{Tp}{Tp + Fp} \quad (1)$$

$$\text{Recall} = \frac{Tp}{Tp + Tn} \quad (2)$$

$$\text{Recall} = \frac{precision\ .recall}{Precision + recall} \quad (3)$$

3.5. Implementasi

Model yang telah dilatih kemudian diimplementasikan dalam bentuk sistem klasifikasi berbasis citra. Sistem ini dirancang untuk menerima input berupa gambar tubuh

balita dan menghasilkan prediksi status stunting secara otomatis. Implementasi dilakukan dalam lingkungan gradio dengan antarmuka yang memungkinkan pengujian model secara real-time terhadap data uji.

4. HASIL DAN PEMBAHASAN

4.1. Pengumpulan Data

Dataset yang digunakan dalam penelitian ini terdiri atas 2.156 citra tubuh balita yang terbagi ke dalam dua kategori, yaitu stunting dan normal. Seluruh data dibagi ke dalam tiga subset, yakni data pelatihan sebanyak 1.298 gambar (615 stunting dan 683 normal), data validasi sebanyak 372 gambar (176 stunting dan 196 normal), serta data pengujian sebanyak 186 gambar (88 stunting dan 98 normal). Strategi pembagian ini bertujuan untuk menghindari overfitting dan memberikan evaluasi objektif terhadap kemampuan generalisasi model.

4.2. *Prapemrosesan Citra*

Tahap prapemrosesan citra dilakukan untuk menyiapkan data agar sesuai dengan format input yang dibutuhkan oleh model *Vision Transformer* (ViT). Setiap citra tubuh balita yang telah dikumpulkan terlebih dahulu dikonversi ke dalam format RGB, kemudian diubah ukurannya menjadi 224×224 piksel agar konsisten dengan dimensi *patch* pada arsitektur ViT. Setelah itu, citra dinormalisasi menggunakan *ViT Image Processor* dari pustaka *Hugging Face* untuk memastikan nilai piksel berada dalam skala yang sesuai bagi proses inferensi dan pelatihan. Tahapan ini penting agar model dapat mengenali pola visual secara optimal dan menghindari bias yang disebabkan oleh perbedaan resolusi atau distribusi warna antar gambar

4.3. Pelatihan Model

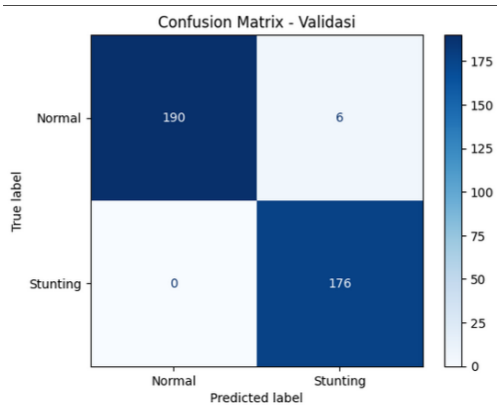
Penurunan nilai training loss selama proses pelatihan menjadi indikator penting dalam menilai sejauh mana model berhasil mempelajari pola dari data pelatihan.

Step	Training Loss
100	0.140200
200	0.077100
300	0.006800
400	0.005500
500	0.000100
600	0.000100
700	0.000100
800	0.000100

Gambar 1. Training Loss

Berdasarkan gambar di atas, terlihat bahwa proses pelatihan model Vision Transformer (ViT) menunjukkan penurunan *training loss* yang signifikan dan stabil. Nilai *loss* tercatat sebesar 0,1402 pada step ke-100, dan terus menurun secara konsisten hingga mencapai 0,0001 sejak step ke-500 hingga akhir pelatihan. Hal ini mengindikasikan bahwa model belajar dengan baik dari data pelatihan tanpa mengalami overfitting. Berdasarkan output terminal, diketahui bahwa model menyelesaikan 815 langkah pelatihan (global_step) dengan *training loss* rata-rata sebesar 0,0282. Proses pelatihan berlangsung selama 5 epoch dalam waktu sekitar 9.609 detik, dengan kecepatan pemrosesan sebesar 0,675 sampel per detik. Penurunan *loss* yang halus dan mendatar di akhir pelatihan pada gambar di atas juga mengindikasikan bahwa model telah mencapai konvergensi yang optimal, dan siap untuk dievaluasi pada data validasi dan pengujian.

4.4. Evaluasi Model



Gambar 2. Confusion Matrix

Berdasarkan hasil confusion matrix yang ditampilkan pada Gambar 4.2 model Vision Transformer (ViT) menunjukkan performa

klasifikasi yang sangat baik dalam membedakan citra tubuh balita dengan status gizi normal dan stunting. Dari total 372 data validasi, model mampu mengklasifikasikan 190 data balita normal secara tepat, sementara 6 data salah diklasifikasikan sebagai stunting. Sementara itu, seluruh 176 data balita dengan kondisi stunting berhasil diklasifikasikan dengan benar tanpa kesalahan (false negative = 0). Hal ini mencerminkan kemampuan model yang sangat baik dalam mendeteksi kasus stunting secara akurat, sekaligus mempertahankan tingkat kesalahan minimum dalam prediksi terhadap kelas normal.

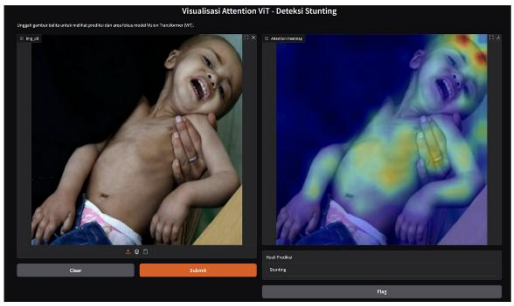
```
=== Classification Report ===
              precision    recall  f1-score   support

      Normal       1.00      0.97      0.98       196
    Stunting       0.97      1.00      0.98       176

    accuracy                           0.98       372
   macro avg       0.98      0.98      0.98       372
weighted avg       0.98      0.98      0.98       372
```

Gambar 3. Classification Report

Berdasarkan hasil evaluasi melalui *classification report*, model Vision Transformer (ViT) menunjukkan performa klasifikasi yang sangat tinggi dengan nilai akurasi keseluruhan sebesar 98%. Pada kelas "Normal", diperoleh precision sebesar 1.00 dan recall sebesar 0.97, yang menunjukkan bahwa hampir seluruh prediksi untuk kelas ini benar, meskipun terdapat sedikit kesalahan dalam mendeteksi beberapa data aktual. Sementara itu, pada kelas "Stunting", precision tercatat sebesar 0.97 dan recall mencapai 1.00, menandakan bahwa model berhasil mengidentifikasi seluruh balita stunting secara tepat tanpa ada yang terlewat. Nilai f1-score yang tinggi pada kedua kelas (0.98) mengindikasikan keseimbangan antara precision dan recall. Selain itu, nilai rata-rata makro dan rata-rata tertimbang untuk precision, recall, dan f1-score juga mencapai 0.98, memperkuat bukti bahwa model memiliki kinerja yang konsisten dan andal dalam mengklasifikasikan kedua kategori citra tubuh balita secara akurat.

4.5. Implementasi Sistem

Sistem klasifikasi citra untuk pendeteksian dini stunting pada balita diimplementasikan menggunakan antarmuka berbasis Gradio.



Gambar 4. Implementasi Sistem

Berdasarkan gambar di atas, sistem menampilkan visualisasi attention map untuk membantu pengguna memahami bagian gambar mana yang menjadi fokus perhatian model saat melakukan klasifikasi. Visualisasi ini ditampilkan dalam bentuk *heatmap* dengan skema warna jet, di mana setiap warna menunjukkan tingkat perhatian model terhadap area gambar. Warna merah menunjukkan area dengan perhatian tertinggi (*high attention*), diikuti oleh kuning sebagai perhatian sedang, sementara biru dan ungu menunjukkan perhatian rendah (*low attention*). Biasanya, area dengan warna merah atau kuning berada pada bagian tubuh yang relevan seperti perut, dada, atau kepala balita. Sebaliknya, bagian latar belakang atau area non-fisik seperti ujung gambar cenderung berwarna biru atau ungu. sistem menampilkan tidak hanya hasil klasifikasi, tetapi juga visualisasi attention map yang dihasilkan dari mekanisme *self-attention* ViT. Warna merah dan kuning pada *heatmap* menunjukkan area dengan tingkat perhatian tinggi (*high attention*), seperti bagian tubuh yang menjadi indikator visual penting, sedangkan biru atau ungu menunjukkan perhatian rendah (*low attention*). Visualisasi ini meningkatkan interpretabilitas model dan memberikan transparansi terhadap proses klasifikasi.

5. KESIMPULAN

1. Penelitian ini berhasil mengembangkan sistem klasifikasi citra tubuh balita menggunakan arsitektur Vision Transformer (ViT) untuk mendeteksi status gizi (stunting atau normal) secara otomatis, dengan hasil evaluasi yang menunjukkan performa tinggi, yaitu

akurasi sebesar 98% serta nilai precision dan recall rata-rata sebesar 0,98.

2. Visualisasi attention map yang dihasilkan memberikan transparansi terhadap proses klasifikasi, membantu pengguna memahami area tubuh yang menjadi fokus utama dalam pengambilan keputusan oleh model.
3. Kelebihan dari pendekatan ini adalah kemampuannya untuk mendeteksi stunting tanpa memerlukan alat ukur antropometri, sehingga sangat bermanfaat di wilayah yang memiliki keterbatasan infrastruktur medis.
4. Namun, kekurangan sistem ini terletak pada ketergantungannya terhadap kualitas dan resolusi citra yang baik agar klasifikasi dapat dilakukan secara akurat.
5. Pengembangan lebih lanjut dapat dilakukan dengan memperluas jumlah dataset, mengintegrasikan teknologi edge computing untuk implementasi offline, serta membandingkan performa ViT dengan model vision lain seperti EfficientNet atau ConvNeXt pada domain yang sama.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada:

1. Allah SWT atas segala rahmat, petunjuk, dan kekuatan yang diberikan sehingga penelitian ini dapat diselesaikan dengan baik.
2. Bapak Rizal Adi Saputra, S.Kom., M.Kom. selaku dosen pembimbing yang telah memberikan bimbingan, arahan, serta motivasi selama proses penelitian ini berlangsung.
3. Rekan-rekan dan teman seperjuangan yang turut memberikan dukungan, semangat, serta bantuan dalam berbagai bentuk selama pelaksanaan penelitian.

DAFTAR PUSTAKA

[1] s. Nanda, s. Iswahyudi, and r. E. Putra, "sistem deteksi stunting pada balita berbasis web menggunakan metode random forest," *journal of informatics and computer science*, vol. 06, 2024.

[2] f. A. Sany, "penerapan sistem pakar untuk deteksi stunting," *jurnal ilmiah ecosystem*, vol. 23, no. 3, pp. 602–609, dec. 2023, doi: 10.35965/eco.v23i3.3774.

[3] i. G. Wiryawan, k. Yuwita, and a. A. Kurniasari, "penerapan algoritma certainty factor pada metode case-based reasoning untuk sistem pakar deteksi stunting," *jurnal pekommas*, vol. 9, no. 1, pp. 67–79, jun. 2024, doi: 10.56873/jpkm.v9i1.5279.

[4] r. Ratnasari, a. J. Wahidin, and t. H. Andika, "deteksi dini stunting pada anak berdasarkan indikator antropometri dengan menggunakan algoritma machine learning," *jurnal algoritma*, vol. 21, no. 2, pp. 378–387, dec. 2024, doi: 10.33364/algoritma/v.21-2.2122.

[5] n. R. Febriyanti, k. Kusrini, and a. D. Hartanto, "analisis perbandingan algoritma svm, random forest dan logistic regression untuk prediksi stunting balita," *edumatic: jurnal pendidikan informatika*, vol. 9, no. 1, pp. 149–158, apr. 2025, doi: 10.29408/edumatic.v9i1.29407.

[6] w. Bismi and h. Harafani, "perbandingan metode deep learning dalam mengklasifikasi citra scan mri penyakit otak parkinson," *incomtech : jurnal telekomunikasi dan komputer*, vol. 12, no. 3, p. 177, dec. 2022, doi: 10.22441/incomtech.v12i3.15068.

[7] s. O. Purba, "klasifikasi penyakit mata pada manusia dengan menggunakan model arsitektur vision transformers," nov. 2024. Accessed: jul. 02, 2025. [online]. Available: https://repositori.uma.ac.id/jspui/bitstream/123456789/24588/2/198160066%20-%20sentia%20ovania%20purba%20-%20fulltext.pdf

[8] s. Singh, m. Kumar, a. Kumar, b. K. Verma, k. Abhishek, and s. Selvarajan, "efficient pneumonia detection using vision transformers on chest x-rays," *sci rep*, vol. 14, no. 1, dec. 2024, doi: 10.1038/s41598-024-52703-2.

[9] a. R. Hanif, e. Nasrullah, and f. X. A. Setyawan, "deteksi karakter plat nomor kendaraan dengan menggunakan metode optical character recognition (ocr)," *jurnal informatika dan teknik elektro terapan*, vol. 11, no. 1, jan. 2023, doi: 10.23960/jitet.v11i1.2897.

[10] h. Hen lukmana, m. Al-husaini, i. Hoeronis, and l. Desi puspareni, "pengembangan sistem informasi deteksi dini stunting

berbasis sistem pakar menggunakan metode forward chaining".

[11] a. Dosovitskiy *et al.*, "an image is worth 16x16 words: transformers for image recognition at scale," jun. 2021, [online]. Available: http://arxiv.org/abs/2010.11929

[12] a. Putra tupu djoru and s. Yulianto, "pendekatan machine learning untuk deteksi stunting pada balita menggunakan k-nearest neighbors," 2025.

[13] k. Mochammad *et al.*, "implementasi arsitektur alexnet dan resnet34 pada klasifikasi citra penyakit daun kentang menggunkan transfer learning," 2023.