



Analysis of GAN Effectiveness in Improving Hearing Impaired Detection Accuracy Using Audio and Visual Data

Analisis Efektivitas GAN dalam Meningkatkan Akurasi Deteksi Tuna Rungu dengan Menggunakan Data Audio dan Visual

Dadang Iskandar Mulyana^{1*}, Awaludin Taufiq Nurrohman²,

^{1,2}Program Studi Teknik Informatika, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, Indonesia

E-Mail: ¹mahvin2012@gmail.com, ²awaludintaufiq04@gmail.com

Received Aug 29th 2024; Revised Oct 11th 2024; Accepted Nov 12th 2024; Available Online Dec 5th 2024

Corresponding Author: Dadang Iskandar Mulyana

Copyright © 2025 by Authors, Published by Institut Riset dan Publikasi Indonesia (IRPI)

Abstract

This research analyzes the effectiveness of Generative Adversarial Networks (GAN) in improving deaf detection accuracy using audio and visual data. The data used in this research includes video recordings of non-deaf individuals obtained from GRID Corpus and deaf individuals collected from YouTube and SLB Islam As-Syafi'iyah. The research stages involved several steps, namely data collection, preprocessing, data augmentation, model building, model training, and model evaluation. The results showed that the GAN model was able to significantly improve the accuracy of deaf detection using visual and audio visual data. The model achieved an Area Under Curve (AUC) value of 0.97, which is much higher than the conventional visual model that only achieved an AUC of 0.52. This finding confirms the great potential of using GAN in the development of a more reliable and effective deafblind detection system. As such, this research provides a solid foundation for further exploration of the use of deep learning methods in hearing loss detection, as well as opening up opportunities for innovation in the field of health technology. The results obtained also show that the integration of audio and visual data can provide more comprehensive information in the detection system.

Keyword: Audio-Visual Data, Data Augmentation, Deep Learning, Generative Adversarial Networks, Hearing Loss Detection.

Abstrak

Penelitian ini menganalisis efektivitas Generative Adversarial Networks (GAN) dalam meningkatkan akurasi deteksi tuna rungu dengan menggunakan data audio dan visual. Data yang digunakan dalam penelitian ini meliputi rekaman video dari individu non-tunarungu yang diperoleh dari GRID Corpus dan individu tunarungu yang dikumpulkan dari YouTube dan SLB Islam As-Syafi'iyah. Tahapan penelitian melibatkan beberapa langkah, yaitu pengumpulan data, preprocessing, augmentasi data, pembangunan model, pelatihan model, dan evaluasi model. Hasil penelitian menunjukkan bahwa model GAN mampu meningkatkan akurasi deteksi tunarungu secara signifikan dengan menggunakan data visual dan audio visual. Model ini berhasil mencapai nilai Area Under Curve (AUC) sebesar 0,97, yang jauh lebih tinggi dibandingkan dengan model visual konvensional yang hanya mencapai AUC sebesar 0,52. Temuan ini menegaskan potensi besar penggunaan GAN dalam pengembangan sistem pendeteksian tuna rungu yang lebih andal dan efektif. Dengan demikian, penelitian ini memberikan dasar yang kuat untuk eksplorasi lebih lanjut tentang penggunaan metode pembelajaran mendalam dalam deteksi gangguan pendengaran, serta membuka peluang untuk inovasi di bidang teknologi kesehatan. Hasil yang diperoleh juga menunjukkan bahwa integrasi data audio dan visual dapat memberikan informasi yang lebih komprehensif dalam sistem deteksi.

Kata Kunci: Augmentasi Data, Data Audio-Visual, Deteksi Gangguan Pendengaran, Generative Adversarial Networks

1. PENDAHULUAN

Gangguan pendengaran, atau tunarungu, adalah kondisi medis yang signifikan dan mempengaruhi kemampuan seseorang untuk mendengar secara normal. Berdasarkan data terbaru dari *World Health Organization* (WHO) tahun 2024, sekitar 5% populasi dunia atau 430 juta orang mengalami beberapa tingkat gangguan pendengaran, dengan prevalensi yang terus meningkat seiring bertambahnya usia [1]. Gangguan pendengaran dapat memiliki dampak yang luas dan mendalam pada kehidupan individu yang mengalaminya,

termasuk penurunan kualitas hidup, isolasi sosial, kesulitan dalam komunikasi, dan penurunan kemampuan kognitif. Pada anak-anak, gangguan pendengaran dapat menghambat perkembangan bahasa dan komunikasi, yang pada gilirannya mempengaruhi prestasi akademik dan perkembangan sosial mereka [2]. Di Indonesia, data dari Kementerian Kesehatan menunjukkan bahwa sekitar 12,3% dari total populasi mengalami gangguan pendengaran, dengan prevalensi yang lebih tinggi di kalangan lansia. Gangguan pendengaran memiliki dampak yang signifikan terhadap kualitas hidup dan kemampuan komunikasi, serta dapat menghambat perkembangan sosial dan pendidikan, terutama pada anak-anak [3], [4].

Pendeteksian dini gangguan pendengaran sangat penting untuk mengurangi dampak negatif yang ditimbulkan. Teknologi yang digunakan untuk mendeteksi gangguan pendengaran telah mengalami perkembangan yang signifikan, termasuk penggunaan model pembelajaran mesin dan jaringan saraf tiruan. Salah satu teknologi yang menjanjikan adalah *Generative Adversarial Networks* (GAN) [5]. GAN adalah jenis jaringan saraf tiruan yang terdiri dari dua model, generator dan discriminator, yang dilatih secara bersamaan untuk menghasilkan data yang menyerupai data asli dan membedakan antara data asli dan data yang dihasilkan [5]–[8].

Selama beberapa dekade terakhir, model pembelajaran mesin, khususnya *Convolutional Neural Networks* (CNN), telah digunakan secara luas dalam pengolahan citra dan sinyal audio untuk berbagai aplikasi medis, termasuk deteksi gangguan pendengaran. Namun, metode konvensional seperti CNN memiliki keterbatasan dalam menangkap kompleksitas dan variasi data yang diperlukan untuk deteksi gangguan pendengaran yang akurat. Dalam konteks ini, GAN menawarkan pendekatan yang lebih fleksibel dan adaptif dengan kemampuan untuk mempelajari representasi data yang lebih mendalam dan kompleks [9].

Penelitian ini bertujuan untuk menganalisis efektivitas GAN dalam meningkatkan akurasi deteksi tunarungu dengan menggunakan data audio dan visual. Model klasifikasi yang dihasilkan akan dibandingkan dengan model yang menggunakan GAN untuk augmentasi yang hanya menggunakan data visual, model augmentasi GAN yang menggunakan data audio visual, serta model augmentasi konvensional yang kemudian dilanjutkan ke tahap klasifikasi menggunakan CNN yang menggunakan data visual. Penggunaan data audio dan visual diharapkan dapat memberikan informasi yang lebih kaya dan akurat dalam mendeteksi gangguan pendengaran. Audio memberikan informasi mengenai kualitas suara dan pola bicara, sedangkan data visual memberikan informasi mengenai gerakan bibir dan ekspresi wajah yang dapat menjadi indikator gangguan pendengaran.

Teknologi deteksi gangguan pendengaran berbasis data audio-visual memiliki potensi besar untuk diterapkan dalam berbagai konteks, termasuk skrining di sekolah, klinik, dan rumah sakit. Dengan semakin berkembangnya teknologi perangkat keras dan perangkat lunak, implementasi model deteksi yang canggih ini menjadi lebih fleksibel dan efisien. Penggunaan GAN dalam deteksi gangguan pendengaran juga membuka peluang untuk pengembangan aplikasi yang dapat digunakan secara luas oleh masyarakat, sehingga meningkatkan aksesibilitas dan deteksi dini gangguan pendengaran [3], [4].

Beberapa penelitian sebelumnya yang relevan dengan penggunaan GAN dalam berbagai domain telah menunjukkan efektivitas teknologi ini dalam pemrosesan data visual dan audio. Penelitian oleh Anugrah Akbar Pramadhan dan Guntur Eka Saputra (2021), berjudul *Cycle Generative Adversarial Networks Algorithm with Style Transfer for Image Generation*, membahas penggunaan *Cycle GAN* untuk transfer gaya pada citra, yang menghasilkan citra lebih realistis dengan mempertahankan gaya tertentu. Meskipun penelitian ini berfokus pada kemampuan GAN dalam mempelajari fitur visual dari data citra, penelitian ini terbatas pada data visual saja dan tidak menggabungkan data audio, yang menjadi fokus dalam penelitian ini untuk deteksi gangguan pendengaran [5]. Penelitian kedua oleh Andre Satriawan, Bahtiar Imran, dan Surni Erniwati (2023), berjudul *Identifikasi Kemiripan Foto Asli dan Sketsa Menggunakan Model Generative Adversarial Networks (GANs)*, meneliti penerapan GAN untuk mengidentifikasi kemiripan antara foto asli dan sketsa. Meskipun penelitian ini menunjukkan kemampuan GAN untuk mendeteksi kesamaan antara dua jenis representasi visual, ia hanya menggunakan data visual dan tidak memasukkan data audio, yang merupakan komponen penting dalam penelitian ini [6].

Penelitian lain yang dilakukan oleh Aram You, et al (2022) berjudul *Application of Generative Adversarial Networks (GAN) for Ophthalmology Image Domains*, meninjau berbagai aplikasi GAN dalam domain oftalmologi, terutama untuk meningkatkan kualitas citra medis dan mendeteksi penyakit mata. Studi ini menunjukkan potensi GAN dalam domain medis, tetapi terbatas pada pemrosesan citra tanpa mempertimbangkan integrasi data audio visual seperti yang diusulkan dalam penelitian ini [7]. Selain itu, penelitian yang dilakukan oleh Alankrita Aggarwal, Mamta Mittal, dan Gopi Battinen (2021), berjudul *Generative Adversarial Network: An Overview of Theory and Applications*, memberikan gambaran umum mengenai teori dan berbagai aplikasi GAN di berbagai bidang, termasuk augmentasi data dalam visi komputer. Namun, penelitian ini tidak spesifik pada penggunaan data audio-visual secara bersamaan, seperti yang diimplementasikan dalam penelitian ini untuk mendeteksi gangguan pendengaran [8].

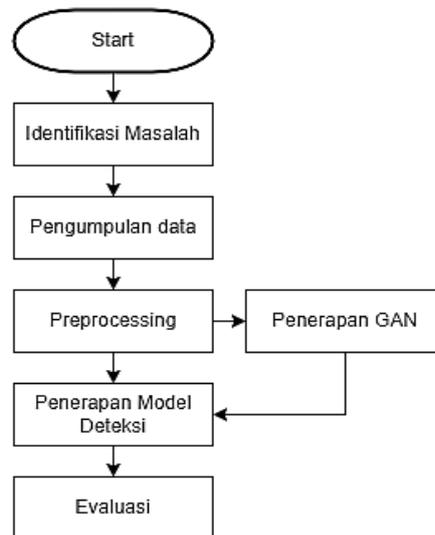
Perbedaan utama antara penelitian ini dengan penelitian terdahulu adalah penggunaan data audio dan visual secara bersamaan untuk meningkatkan akurasi deteksi gangguan pendengaran. Penelitian ini tidak hanya mengandalkan data visual, tetapi juga menggabungkan informasi audio seperti pola bicara dan kualitas

suara. Dalam penelitian ini mengembangkan model augmentasi GAN yang dapat mendeteksi tunarungu dengan menggunakan kombinasi data audio dan visual [9]. Penelitian ini juga membandingkan kinerja model ini dengan model GAN dan CNN yang hanya menggunakan data visual serta model augmentasi konvensional dan CNN yang menggunakan data visual [10]. Hasil dari penelitian ini diharapkan dapat memberikan kontribusi yang signifikan dalam pengembangan teknologi pendeteksian gangguan pendengaran dan meningkatkan akurasi deteksi sehingga dapat memberikan intervensi yang lebih cepat dan tepat bagi individu yang mengalami gangguan pendengaran.

Dengan demikian, penelitian ini tidak hanya bertujuan untuk mengembangkan model deteksi yang lebih akurat, tetapi juga untuk memberikan pemahaman yang lebih mendalam mengenai manfaat integrasi data audio dan visual dalam deteksi gangguan pendengaran. Harapannya, temuan dari penelitian ini dapat diadopsi dan dikembangkan lebih lanjut dalam praktek klinis dan aplikasi sehari-hari, sehingga dapat memberikan dampak positif bagi kualitas hidup individu dengan gangguan pendengaran.

2. BAHAN DAN METODE

Dalam penelitian ini, penerapan metodologi dilakukan melalui serangkaian langkah yang dirancang untuk mengembangkan model deteksi gangguan pendengaran berbasis GAN, melibatkan data audio dan visual. Metodologi penelitian ini terdiri dari beberapa tahap yang mencakup identifikasi masalah, pengumpulan data, pre-processing, penerapan metode GAN, penerapan model deteksi, dan evaluasi hasil.



Gambar 1. Metodologi Penelitian

2.1. Identifikasi Masalah

Fokus utama dalam penelitian ini adalah untuk meningkatkan akurasi deteksi gangguan pendengaran dengan memanfaatkan Generative Adversarial Network (GAN) yang mengintegrasikan data audio dan visual. Gangguan pendengaran merupakan masalah kesehatan penting dengan dampak yang signifikan terhadap kualitas hidup individu. Tantangan utama dari metode deteksi konvensional, yang cenderung terbatas dalam akurasi dan kemampuan integrasi data, seperti sensitivitas yang rendah terhadap lingkungan yang bising untuk metode audio dan resolusi yang terbatas untuk analisis visual, mendorong penggunaan GAN sebagai solusi potensial. Diharapkan bahwa penggunaan GAN dapat tidak hanya meningkatkan akurasi deteksi, tetapi juga memperluas kemampuan model untuk menginterpretasikan dan mengintegrasikan informasi dari kedua jenis data tersebut secara lebih efektif dalam konteks klinis deteksi gangguan pendengaran.

2.2. Pengumpulan Data

Langkah-langkah pengumpulan data melibatkan identifikasi sumber data yang terpercaya, validasi kualitas data, serta klasifikasi data sesuai dengan kebutuhan penelitian. Pengumpulan data yang efektif adalah kunci untuk memastikan bahwa model yang dikembangkan dapat memberikan hasil yang akurat dan andal. Pada tahap pengumpulan data, data yang digunakan adalah data video, yang kemudian dilakukan ekstraksi untuk memisahkan audio dari video tersebut. Tahap pengumpulan data mencakup langkah-langkah berikut :

2.2.1. Dataset Publik

Dataset publik yang digunakan untuk penelitian ini seperti GRID Corpus, dan Youtube. Dataset dari GRID Corpus menyediakan rekaman video gerakan bibir dan ekspresi wajah dari pembicara yang

mengucapkan kalimat sederhana dalam kondisi yang terkendali. GRID Corpus sangat berharga untuk penelitian deteksi gangguan pendengaran karena menyediakan data dengan kualitas tinggi dan konsisten, yang memungkinkan analisis detail gerakan bibir dan audio yang dihasilkan. Selain itu, dataset yang berasal dari Youtube mencakup konten dari individu dengan gangguan pendengaran yang sedang melakukan aktivitas berbicara sederhana, dan mempraktekan bahasa isyarat. Pemilihan video dari YouTube dilakukan dengan mempertimbangkan kualitas rekaman dan relevansi dengan tujuan penelitian, sehingga data yang dikumpulkan dapat merepresentasikan situasi nyata dengan lebih baik.

2.2.2. Dataset Private

Dataset publik yang digunakan untuk penelitian ini diperoleh dari SLB Islam As-Syafiiyah Jakarta. Dataset ini mencakup rekaman video dari individu dengan kondisi pendengaran yang telah diverifikasi. Video-video ini mencakup rekaman siswa yang berbicara dan mengeja huruf vokal serta menunjukkan ekspresi mereka dalam lingkungan belajar sehari-hari. Rekaman ini sangat berharga karena menyediakan data yang otentik dan relevan untuk melatih dan menguji model dalam kondisi yang sangat mirip dengan penggunaan nyata. Pengumpulan data dari SLB Islam As-Syafiiyah dilakukan dengan memperhatikan etika penelitian, termasuk mendapatkan izin dari pihak terkait dan memastikan kerahasiaan data pribadi.

2.3. Preprocessing

Tahap preprocessing sangat penting untuk memastikan bahwa data yang digunakan dalam penelitian ini bersih dan siap untuk dianalisis. Preprocessing mencakup beberapa langkah yang bertujuan untuk meningkatkan kualitas data dan mempersiapkannya untuk tahap analisis selanjutnya. Sebelum masuk ke dalam preprocessing dengan program, peneliti terlebih dahulu mengevaluasi data-data yang layak digunakan, dan melakukan pembersihan noise secara manual pada aplikasi editing, dan selanjutnya melakukan preprocessing dengan menggunakan program. Proses ini melibatkan berbagai teknik untuk membersihkan, mengolah, dan menormalisasi data, sehingga data yang dihasilkan siap untuk digunakan dalam model GAN. Tahap preprocessing ini dirancang untuk mengatasi berbagai masalah yang mungkin ada dalam data mentah, seperti noise, inkonsistensi, dan data yang hilang, sehingga memastikan bahwa data yang digunakan dalam analisis adalah data yang berkualitas tinggi dan representatif.

Langkah-langkah yang dilakukan dalam preprocessing meliputi pembersihan data dan augmentasi data. Pembersihan data dilakukan pada data audio dan visual. Untuk data audio, proses pembersihan mencakup penghilangan noise yang bisa mengganggu kualitas rekaman. Noise adalah gangguan yang bisa berasal dari berbagai sumber seperti lingkungan perekaman yang berisik. Untuk mengatasi masalah ini, digunakan teknik pengurangan noise dengan filter adaptif. Filter ini mampu mengidentifikasi dan menghilangkan noise tanpa mengorbankan kualitas suara asli.

Data audio diekstraksi dari video menggunakan algoritma pemisahan audio, seperti yang tersedia dalam pustaka seperti pydub atau librosa. Setelah itu, audio kemudian di ekstraksi menggunakan MFCC (Mel-Frequency Cepstral Coefficients) untuk mengidentifikasi ciri khas suara yang relevan dengan kondisi pendengaran. Rumus untuk MFCC adalah $C_m = \sum \log Sk \cos[\pi m(k - 0.5)/K]Kk - 1$ dimana sk adalah spektrum daya pada bin ke- k dan K adalah jumlah total bin frekuensi.

Untuk data visual, proses pembersihan melibatkan memastikan kualitas gambar video dengan menghilangkan artefak visual dan meningkatkan resolusi gambar. Proses ini melibatkan konversi video menjadi frame gambar, resize, dan normalisasi piksel untuk konsistensi data.

Augmentasi data adalah langkah krusial dalam metodologi penelitian ini untuk meningkatkan jumlah dan keragaman data yang tersedia. Teknik augmentasi data digunakan untuk mengatasi keterbatasan dataset, mengurangi risiko overfitting, dan meningkatkan kemampuan generalisasi model. Dalam konteks deteksi gangguan pendengaran berbasis GAN, augmentasi data dilakukan pada data audio dan visual.

Untuk augmentasi data audio, teknik-teknik yang digunakan meliputi:

1. Additive Noise: Menambahkan noise acak pada sinyal audio untuk meningkatkan robus model terhadap variasi data.

$$Y_{noisy}(t) = y(t) + \alpha \cdot N(0,1) \quad (1)$$

2. Pitch Shifting: Mengubah tinggi nada suara tanpa mengubah durasi rekaman.

$$Y_{shifted}(t) = librosa.effects.pitch_shift(, sr, n_steps) \quad (2)$$

Untuk augmentasi data visual, teknik-teknik yang digunakan meliputi:

1. Rotasi Frame: Memutar gambar dalam rentang sudut tertentu untuk menciptakan variasi dalam orientasi wajah. Rumus yang digunakan adalah

$$M = \begin{bmatrix} \cos \theta & -\sin \theta & (1 - \cos \theta) \cdot \frac{W}{2} + \sin(\theta) \cdot \frac{H}{2} \\ \sin \theta & \cos \theta & (1 - \cos \theta) \cdot \frac{H}{2} - \sin(\theta) \cdot \frac{W}{2} \end{bmatrix} \quad (3)$$

$$\text{Rotated Frame}(x^1, y^1) = \text{Gray}(M_{11} \cdot x + M_{12} \cdot y + M_{13}, M_{21} \cdot x + M_{22} \cdot y + M_{23}) \quad (4)$$

2. Flip Frame: Membalik gambar secara horizontal atau vertikal untuk menciptakan variasi orientasi.

$$\text{Flipped Frame}(x, y) = \text{Gray}(W - x - 1, y) \quad (5)$$

2.4. Convolutional Neural Network (CNN)

Convolutional Neural Networks (CNN) merupakan salah satu metode yang sangat efektif dalam klasifikasi citra, terutama dalam berbagai aplikasi di bidang pengolahan gambar. CNN bekerja dengan menggabungkan lapisan-lapisan konvolusi yang memungkinkan model untuk menangkap pola dan fitur penting dari citra, seperti tepi, tekstur, dan bentuk, yang kemudian digunakan untuk melakukan klasifikasi. Penelitian oleh Pangestu et al. (2020) menunjukkan bahwa CNN berhasil diterapkan untuk klasifikasi citra lahan dengan akurasi yang tinggi, menunjukkan keandalannya dalam pengenalan pola citra kompleks [11].

CNN telah diterapkan dalam berbagai konteks, seperti pengenalan pola pada motif batik yang menghasilkan klasifikasi yang tepat dan efisien untuk motif batik yang lebih beragam [12]. Kemampuan CNN untuk mengklasifikasikan gambar secara akurat melalui ekstraksi fitur yang lebih mendalam dan representatif dibandingkan metode konvensional juga digunakan untuk klasifikasi objek, dengan hasil yang menunjukkan tingkat akurasi yang lebih tinggi [13].

2.4.1 Struktur dan Fungsi CNN

CNN terdiri dari beberapa lapisan utama, yaitu lapisan konvolusi, lapisan aktivasi, lapisan pooling, dan lapisan fully connected. Proses utama dalam CNN adalah konvolusi, di mana filter (atau kernel) bergerak di atas citra untuk menghasilkan peta fitur (feature map). Proses ini dapat dijelaskan dengan persamaan matematis berikut:

$$F(i, j) = \sum_{m=-k}^k \sum_{n=-k}^n I(i + m, j + n) \cdot K(m, n) \quad (6)$$

di mana: $F(i, j)$ adalah nilai piksel pada posisi (i, j) pada peta fitur yang dihasilkan; I adalah citra input; K adalah filter konvolusi; dan k adalah ukuran kernel yang digunakan.

Setelah konvolusi, fungsi aktivasi, biasanya ReLU (Rectified Linear Unit), diterapkan untuk memperkenalkan non-linearitas ke dalam model:

$$f(x) = \max(0, x) \quad (7)$$

Setelah itu, lapisan pooling (misalnya, max pooling) digunakan untuk mengurangi dimensi peta fitur, yang membantu dalam mengurangi jumlah parameter dan komputasi dalam jaringan:

$$P(i, j) = \max_{(m,n) \in R} F(i + m, j + n) \quad (8)$$

Fungsi lain dari CNN adalah kemampuannya untuk melakukan deteksi dan klasifikasi secara real-time, seperti yang diungkapkan oleh Charli et al. (2020) dalam penelitian mereka mengenai pengenalan jenis burung lovebird menggunakan Faster R-CNN, sebuah pengembangan dari CNN yang memungkinkan deteksi objek lebih cepat dan efisien [14]. Terakhir, penelitian oleh Hossain dan Sajib (2019) juga menggarisbawahi kemudahan dalam melatih model CNN dan konsistensi hasil yang diperoleh dalam berbagai dataset gambar, menekankan fleksibilitas CNN dalam menangani variasi data [15].

2.5. Generative Adversarial Networks (GAN)

Generative Adversarial Networks (GAN) pertama kali diperkenalkan oleh Ian Goodfellow pada tahun 2014 sebagai model jaringan saraf tiruan yang dirancang untuk menghasilkan data baru dengan distribusi yang mirip dengan data asli. GAN terdiri dari dua komponen utama yang berkompetisi, yaitu generator dan discriminator. Generator bertugas menghasilkan data palsu yang menyerupai data asli, sedangkan discriminator bertugas membedakan antara data asli dan data yang dihasilkan oleh generator.

Penggunaan GAN telah meluas ke berbagai aplikasi, seperti pembuatan gambar, deteksi wajah, pengenalan suara, dan pemrosesan sinyal medis. Salah satu penerapannya dijelaskan dalam penelitian oleh Vika Vitaloka Pramansah et al. (2022), yang membahas penciptaan karakter anime otomatis menggunakan

GAN [16]. Dalam studi tersebut, generator berhasil menghasilkan karakter anime realistis dengan gaya yang konsisten melalui transfer fitur visual yang dipelajari dari data citra animeitu, GAN juga telah diimplementasikan dalam sistem presensi berbasis deteksi wajah. GAN berperan dalam menghasilkan citra wajah yang lebih akurat untuk memperkaya data pelatihan dalam sistem deteksi wajah otomatis [17].

Penggunaan GAN dalam pemrosesan data suara, di mana GAN mampu meningkatkan kualitas data suara melalui teknik augmentasi [18]. Hal ini memungkinkan model untuk mempelajari lebih banyak variasi pola suara, yang pada akhirnya meningkatkan akurasi deteksi. Di sisi lain, Idriss (2021) menerapkan GAN dalam mengurangi efek motion blur pada citra bergerak dengan menggunakan Canny Edge Detector, menunjukkan kemampuan GAN dalam menangani masalah pada domain visual yang kompleks [19].

GAN dengan pendekatan feature reconstruction loss dapat digunakan untuk melakukan inpainting pada citra wajah yang tidak selaras. Penelitian ini mengatasi permasalahan ketidaksejajaran dalam data visual dan menekankan pentingnya rekonstruksi fitur visual dalam GAN [20]. Teknik ini menjadi kunci dalam menghasilkan data citra yang realistis. Dalam konteks deteksi gangguan pe teknik ini dapat diterapkan untuk memperbaiki data visual dan audio yang hilang atau rusak, memperkaya informasi yang digunakan dalam proses klasifikasi.

Penggabungan data audio dan visual menggunakan GAN memberikan manfaat yang signifikan dalam menangkap pola yang lebih kompleks. Dalam penelitian ini, GAN digunakan untuk meningkatkan akurasi deteksi tunarungu dengan menggabungkan data audio yang berisi informasi suara dan pola bicara, serta data visual seperti ekspresi wajah dan gerakan bibir. Dengan memanfaatkan GAN, model dapat mempelajari fitur yang lebih mendalam dari data tersebut, menghasilkan sistem deteksi gangguan pendengaran yang lebih akurat dan efisien.

2.6. Penerapan Teknologi GAN dan CNN

Setelah data siap digunakan, langkah berikutnya adalah menerapkan teknologi GAN dan CNN untuk melakukan pelatihan model deteksi. GAN merupakan teknologi yang sangat kuat untuk menghasilkan data baru yang realistis dan telah digunakan dalam berbagai aplikasi medis dan pengolahan citra. Implementasi GAN dalam penelitian ini bertujuan untuk memanfaatkan keunggulan teknologi ini dalam menghasilkan data yang akurat dan realistis, yang dapat meningkatkan akurasi deteksi gangguan pendengaran. GAN digunakan untuk mengintegrasikan data audio dan visual dengan rumus $\min_D \max_G V(D, G) = E_{x \sim p_{data}(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log(1 - D(G(z)))]$ dimana D adalah Discriminator dan G adalah Generator.

Generator berfungsi untuk menghasilkan data audio dan visual baru yang realistis. Generator dilatih untuk menghasilkan data yang menyerupai data asli. Arsitektur generator melibatkan beberapa lapisan konvolusi dan dekonvolusi untuk membangun data baru dari noise. Fungsi loss untuk generator adalah $L_G = -E_{z \sim p_z(z)} \log D(G(z))$ dimana $G(z)$ adalah output Generator dan D adalah Discriminator.

Discriminator bertugas membedakan antara data asli dan data yang dihasilkan oleh generator. Discriminator dilatih untuk mendeteksi data palsu yang dihasilkan oleh generator. Fungsi loss untuk discriminator adalah $L_D = -E_{x \sim p_{data}(x)} [\log D(x)] - E_{z \sim p_z(z)} [\log(1 - D(G(z)))]$. Di mana D_x adalah output discriminator untuk data asli x dan $D(G(z))$ untuk data yang dihasilkan oleh generator.

Selain itu, peneliti juga menggunakan model CNN untuk melakukan perbandingan hasil model deteksi. Arsitektur CNN yang digunakan dalam penelitian ini melibatkan beberapa lapisan, yaitu :

1. Dense Layer : lapisan dense atau fully connected berfungsi untuk menghubungkan semua neuron di lapisan sebelumnya dengan semua neuron di lapisan berikutnya, memungkinkan model untuk mempelajari hubungan yang kompleks dalam data.
2. Dropout Layer : dropout digunakan untuk mencegah overfitting dengan secara acak menonaktifkan sejumlah neuron selama pelatihan. Hal ini membantu model untuk lebih generalisasi terhadap data baru.
3. Lapisan output : berfungsi untuk menghasilkan prediksi akhir dari model. Untuk tugas klasifikasi, lapisan ini menggunakan fungsi aktivasi softmax untuk menghasilkan probabilitas kelas.

2.7. Evaluasi

Hasil Model deteksi yang telah dibuat akan menghasilkan klasifikasi tingkat gangguan pendengaran dari data uji yang diberikan. Dengan membandingkan dengan metode konvensional maka akan di dapatkan hasil apakah metode GAN ini efektif untuk mendeteksi tunarungu atau tidak. Evaluasi model dilakukan untuk mengukur performa model dengan menggunakan data uji. Model yang telah dilatih dievaluasi dengan menggunakan metrik evaluasi seperti akurasi, presisi, recall, dan F1-score. Hasil evaluasi menunjukkan seberapa baik model mampu mengklasifikasikan data uji dengan benar.

3. HASIL DAN PEMBAHASAN

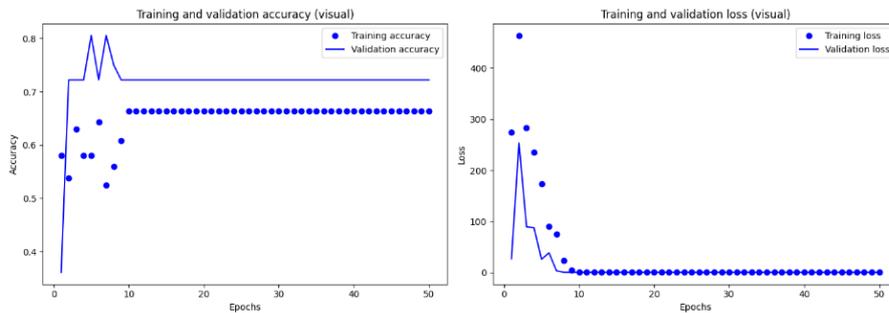
Bab ini menyajikan hasil penelitian tentang efektivitas Generative Adversarial Networks (GAN) dalam meningkatkan akurasi deteksi tuna rungu menggunakan data audio dan visual. Analisis mencakup

evaluasi kinerja model menggunakan berbagai metrik, perbandingan antara model konvensional dan model berbasis GAN, serta interpretasi hasil yang diperoleh.

3.1. Evaluasi Model

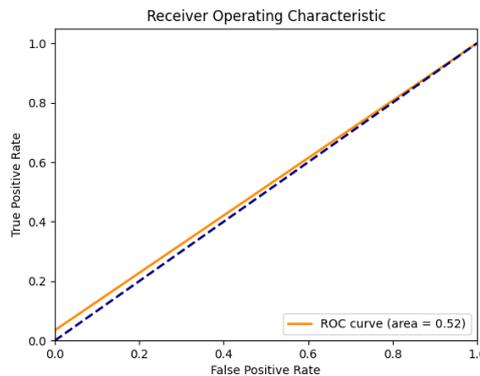
Penelitian ini mengevaluasi efektivitas Generative Adversarial Networks (GAN) dalam meningkatkan akurasi deteksi tuna rungu menggunakan data audio dan visual. Evaluasi dilakukan dengan menggunakan beberapa metrik, termasuk kurva ROC (Receiver Operating Characteristic), area di bawah kurva (AUC), presisi, recall, dan F1-score.

3.1.1. Model Visual



Gambar 2. Grafik Akurasi Model Visual

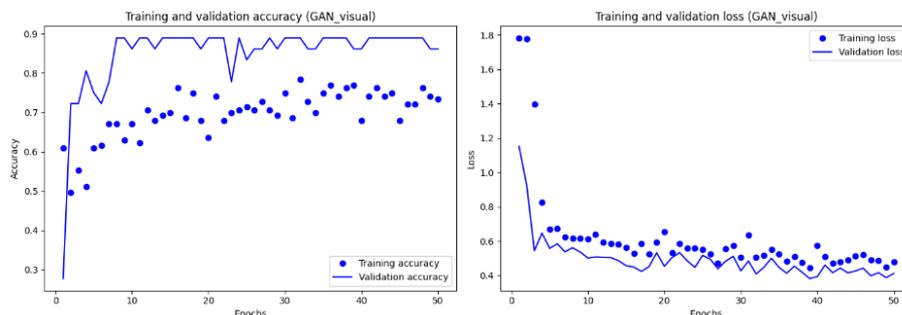
Grafik akurasi pelatihan dan validasi serta loss pelatihan dan validasi menunjukkan performa model selama proses pelatihan. Dari grafik terlihat bahwa akurasi validasi stabil namun akurasi pelatihan meningkat, yang mengindikasikan kemungkinan terjadinya overfitting.



Gambar 3. Grafik ROC Model Visual

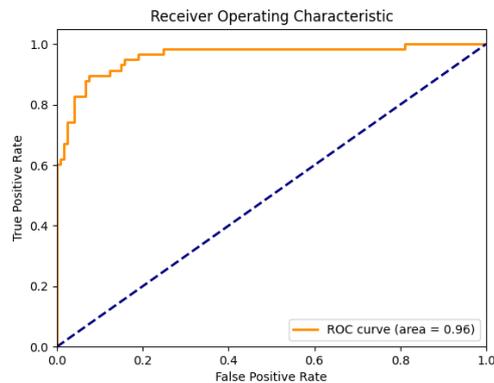
Grafik ROC menunjukkan kemampuan model dalam membedakan antara kelas positif (tuna rungu) dan kelas negatif (non-tuna rungu). AUC untuk model visual adalah 0.52, menunjukkan bahwa model ini memiliki kemampuan yang mendekati acak dalam membedakan kedua kelas.

3.1.2. Model GAN Visual



Gambar 4. Grafik Akurasi Model GAN Visual

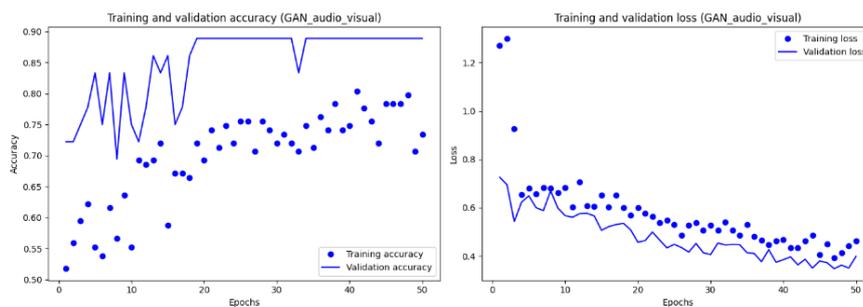
Grafik menunjukkan bahwa akurasi pelatihan dan validasi model GAN visual cenderung meningkat dan stabil. Loss pada pelatihan dan validasi menunjukkan penurunan yang konsisten, menandakan bahwa model ini belajar dengan baik dari data yang diberikan.



Gambar 5. Grafik ROC Model GAN Visual

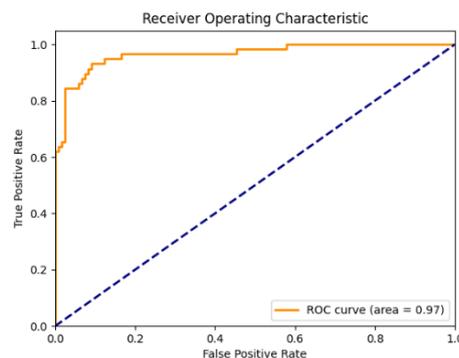
AUC untuk model GAN visual adalah 0.96, menunjukkan bahwa model ini memiliki kemampuan yang sangat baik dalam membedakan antara kelas tuna rungu dan non-tuna rungu.

3.1.3. Model GAN Audio-visual



Gambar 6. Grafik Akurasi Model GAN Audio-visual

Grafik akurasi pelatihan dan validasi serta loss pelatihan dan validasi menunjukkan performa yang baik dari model GAN audio-visual. Akurasi validasi mencapai nilai yang tinggi, dan loss menunjukkan tren penurunan yang konsisten selama pelatihan.



Gambar 7. Grafik ROC Model GAN Audio-visual

AUC untuk model GAN audio-visual adalah 0.97, menunjukkan performa yang sangat baik dalam membedakan antara kelas tuna rungu dan non-tuna rungu.

3.2. Perbandingan Kinerja Model

Laporan klasifikasi memberikan metrik seperti precision, recall, dan F1-score untuk mengevaluasi kinerja masing-masing model. Berikut adalah tabel yang menunjukkan hasil evaluasi dari ketiga model:

Tabel 1. Matrik Klasifikasi

Model	Precision	Recall	F1Score
CNN Visual	0.65	0.70	0.67
GAN Visual	0.96	0.96	0.96
GAN AudioVisual	0.97	0.97	0.97

3.3. Diskusi

Hasil penelitian menunjukkan bahwa penerapan GAN secara signifikan meningkatkan kinerja model dalam deteksi tuna rungu. Model GAN, baik yang menggunakan data visual maupun kombinasi audio-visual, menunjukkan peningkatan substansial dalam semua metrik evaluasi dibandingkan dengan model visual konvensional. Peningkatan AUC dari 0,52 (model visual) menjadi 0,97 (model GAN) menggambarkan perbaikan dramatis dalam kemampuan model untuk membedakan antara individu tuna rungu dan non-tuna rungu. Hal ini mengindikasikan bahwa GAN berhasil menghasilkan data sintesis yang memperkaya dataset pelatihan, memungkinkan model untuk belajar fitur-fitur yang lebih diskriminatif.

Menariknya, model GAN visual dan GAN audio-visual menunjukkan kinerja yang setara, dengan nilai presisi, recall, dan F1-score hampir sama (96 dan 97). Hal ini menimbulkan pertanyaan tentang kontribusi relatif dari data audio dalam meningkatkan akurasi deteksi. Penelitian lebih lanjut mungkin diperlukan untuk menyelidiki apakah penambahan data audio memberikan manfaat signifikan di luar apa yang dapat dicapai dengan data visual saja ketika menggunakan GAN. Stabilitas dan konsistensi yang ditunjukkan dalam grafik akurasi dan loss untuk model-model GAN juga mengindikasikan bahwa teknik ini membantu mengatasi masalah overfitting yang terlihat pada model visual konvensional.

4. KESIMPULAN

Kesimpulan dari penelitian ini menunjukkan bahwa penggunaan Generative Adversarial Networks (GAN) secara signifikan meningkatkan akurasi dalam mendeteksi tunarungu dibandingkan dengan model visual konvensional. Model konvensional menghasilkan AUC sebesar 0,52, yang hampir setara dengan klasifikasi acak, menunjukkan kinerja yang kurang memadai. Sebaliknya, model GAN visual menunjukkan peningkatan dramatis dengan AUC sebesar 0,96, sementara model GAN audio-visual mencapai AUC 0,97, menunjukkan bahwa kombinasi data audio dan visual memberikan informasi yang lebih kaya untuk deteksi tunarungu. Selain itu, penerapan GAN berhasil mengatasi masalah overfitting yang terlihat pada model visual konvensional, sebagaimana ditunjukkan oleh stabilitas dalam grafik akurasi dan loss. Namun, hasil yang hampir setara antara model GAN visual dan GAN audio-visual menimbulkan pertanyaan tentang sejauh mana kontribusi data audio dalam meningkatkan akurasi deteksi.

Adapun kelemahan penelitian ini adalah terbatasnya variasi data yang digunakan, khususnya dalam jumlah sampel audio dan visual, yang dapat mempengaruhi generalisasi model. Selain itu, kompleksitas GAN memerlukan komputasi yang tinggi, yang mungkin menjadi tantangan dalam aplikasi skala besar. Untuk penelitian lanjutan, disarankan untuk menggunakan dataset yang lebih bervariasi dan mencakup kondisi lingkungan yang berbeda, serta mengeksplorasi teknik fine-tuning model GAN agar lebih efisien secara komputasi. Selain itu, pengujian lebih mendalam terhadap kontribusi masing-masing data (audio dan visual) terhadap akurasi model akan memberikan wawasan yang lebih baik tentang peran masing-masing modalitas dalam deteksi tunarungu.

REFERENSI

- [1] W. H. O. (WHO), "Global Prevalence of Hearing Loss," *World Health Organization (WHO)*, 2024. <https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss>
- [2] "Hearing Loss in Children: Prevalence and Impact," *American Speech-Language-Hearing Association (ASHA)*, 2023. <https://www.asha.org/>
- [3] L. Cai, Z. Zhang, Q. Li, and L. Zhang, "Data augmentation based on conditional generative adversarial networks for lesion classification in ultrasound images," *J. Biotech Res.*, pp. 112–125, 2024.
- [4] H. Phan *et al.*, "Improving GANs for Speech Enhancement," *IEEE Signal Process. Lett.*, vol. 27, pp. 1700–1704, 2020, doi: 10.1109/LSP.2020.3025020.
- [5] A. A. Praramadhan dan G. E. Saputra, "Cycle Generative Adversarial Networks Algorithm with Style Transfer for Image Generation," *Proceedings of the International Conference on Computer Science and Information Technology*, 2021.
- [6] A. Satriawan, B. Imran, dan S. Erniwati, "Identifikasi Kemiripan Foto Asli dan Sketsa Menggunakan Model Generative Adversarial Networks (GANs)," *Jurnal Teknologi Informasi dan Komunikasi*, vol. 8, no. 2, pp. 55-63, 2023.
- [7] A. You, J. K. Kim, I. H. Ryu, dan T. K. Yoo, "Application of Generative Adversarial Networks (GAN) for Ophthalmology Image Domains: A Survey," *IEEE Access*, vol. 10, pp. 48065–48078, 2022.

-
- [8] A. Aggarwal, M. Mittal, dan G. Battinen, "Generative Adversarial Network: An Overview of Theory and Applications," *Journal of Artificial Intelligence and Data Mining*, vol. 9, no. 4, pp. 237-248, 2021.
- [9] G. W. Lindsay, "Convolutional Neural Networks as a Model of the Visual System: Past, Present, and Future," *J. Cogn. Neurosci.*, vol. 33, no. 10, pp. 2017–2031, 2021, doi: 10.1162/jocn_a_01544.
- [10] J. Qin, W. Pan, X. Xiang, Y. Tan, and G. Hou, "A biological image classification method based on improved CNN," *Ecol. Inform.*, vol. 58, p. 101093, 2020, doi: <https://doi.org/10.1016/j.ecoinf.2020.101093>.
- [11] R. A. Pangestu, B. Rahmat, dan F. T. Anggraeny, "Implementasi Algoritma CNN untuk Klasifikasi Citra Lahan dan Perhitungan Luas," *Jurnal Informatika dan Sistem Informasi (JIFoSI)*, vol. 1, no. 1, pp. 1-6, Maret 2020.
- [12] T. A. Bowo, H. Syaputra, dan M. Akbar, "Penerapan Algoritma Convolutional Neural Network Untuk Klasifikasi Motif Citra Batik Solo," *Journal of Software Engineering Ampera*, vol. 1, no. 2, pp. 47-55, Juni 2020.
- [13] H. Herdianto dan D. Nasution, "Klasifikasi Objek Menggunakan Metode Convolutional Neural Network (CNN)," *SNASTIKOM Ke 9*, pp. 12-18, Oktober 2022.
- [14] F. Charli, H. Syaputra, M. Akbar, S. Sauda, dan F. Panjaitan, "Implementasi Metode Faster Region Convolutional Neural Network (Faster R-CNN) Untuk Pengenalan Jenis Burung Lovebird," *Journal of Information Technology Ampera*, vol. 1, no. 3, pp. 19-25, Desember 2020.
- [15] Md. A. Hossain dan Md. S. A. Sajib, "Classification of Image using Convolutional Neural Network (CNN)," *Global Journal of Computer Science and Technology: D Neural & Artificial Intelligence*, vol. 19, no. 2, 2019.
- [16] Vika Vitaloka Pramansah, Dadang Iskandar Mulyana, Titi Silfia, Rudi Tri Jaya, (2022). "Penciptaan Karakter Anime Otomatis Dengan Menggunakan Generative Adversarial Networks," *Jurnal Teknik Elektro dan Komputasi (ELKOM) Volume 4 Nomor 1*, Maret 2022.
- [17] Tati Suprpti, Dian Ade Kurnia, Doni Anggara, Rananda Deva Rian, Aldi Setiawan, (2022). "Implementasi Generative Adversarial Networks (GANs) Sistem Presensi Berbasis Deteksi Wajah (SIDEWA)," *Jurnal Teknologi Informasi Komunikasi Vol. 9 No. 2*.
- [18] Heskyel Pranata Tarigan, (2023). "Design And Implementation Of Adversarial Neural Network For Voice Data Processing," *Jurnal Komputer Indonesia*.
- [19] Idriss Moussa Idriss, (2021). "Blind Motion Image Deblurring Using Canny Edge Detector with Generative Adversarial Networks," *Research Paper*, 2021.
- [20] Avin Maulana, Chastine Fatichah, Dan Nanik Suciati, (2020). "Facial Inpainting Pada Citra Wajah Unaligned Menggunakan Generative Adversarial Network Dengan Feature Reconstruction Loss," *Juti: Jurnal Ilmiah Teknologi Informasi*, Vol. 18 No. 2, Juli 2020.