

**PENGOLAHAN CITRA DIGITAL UNTUK EKSTRAKSI
FITUR PADA HYBRID MODEL *CNN-SVM* DALAM
DETEKSI PENYAKIT PNEUMONIA
BERBASIS X-RAY DADA**

SKRIPSI

DISUSUN OLEH

MUHAMMAD AULIA IRZA

NPM. 2109020171



UMSU
Unggul | Cerdas | Terpercaya

**PROGRAM STUDI TEKNOLOGI INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA**

MEDAN

2025

**PENGOLAHAN CITRA DIGITAL UNTUK EKSTRAKSI FITUR PADA HYBRID
MODEL CNN-SVM DALAM DETEKSI PENYAKIT PNEUMONIA BERBASIS X-
RAY DADA**

SKRIPSI

**Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer
(S.Kom) dalam Program Studi Teknologi Informasi pada Fakultas Ilmu Komputer dan
Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara**

MUHAMMAD AULIA IRZA

NPM. 2109020171

**PROGRAM STUDI TEKNOLOGI INFORMASI FAKULTAS ILMU KOMPUTER
DAN TEKNOLOGI INFORMASI**

UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA MEDAN

2025

LEMBAR PENGESAHAN

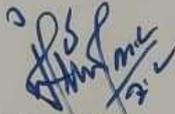
Judul Skripsi : PENGOLAHAN CITRA DIGITAL UNTUK
EKSTRAKSI FITUR PADA HYBRID MODEL CNN-
SVM DALAM DETEKSI PENYAKIT PNEUMONIA
BERBASIS X-RAY DADA

Nama Mahasiswa : Muhammad Aulia Irza

NPM : 2109020171

Program Studi : Teknologi Informasi

Menyetujui Komisi Pembimbing



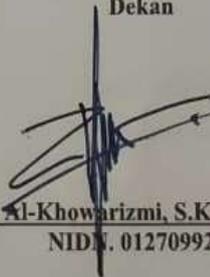
Dr. Irvan, M.Si
NIDN. 0117117301

Ketua Program Studi



Fatma Sari Hutagalung, S.Kom., M.Kom
NIDN. 0117019301

Dekan



Dr. Al-Khowarizmi, S.Kom., M.Kom.
NIDN. 0127099201

PERNYATAAN ORISINALITAS

PENGOLAHAN CITRA DIGITAL UNTUK EKSTRAKSI FITUR PADA
HYBRID MODEL *CNN-SVM* DALAM DETEKSI PENYAKIT
PNEUMONIA BERBASIS X-RAY DADA

SKRIPSI

Saya menyatakan bahwa karya tulis ini adalah hasil karya sendiri, kecuali beberapa kutipan dan ringkasan yang masing-masing disebutkan sumbernya.

Binjai, 28 Februari 2025

Yang membuat pernyataan



Muhammad Aulia Irza

NPM. 2109020171

**PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN
AKADEMIS**

Sebagai sivitas akademika Universitas Muhammadiyah Sumatera Utara, saya bertanda tangan dibawah ini:

Nama : Muhammad Aulia Irza

NPM : 2109020171

Program Studi : Teknologi Informasi

Karya Ilmiah : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Muhammadiyah Sumatera Utara Hak Bedas Royalti Non-Eksekutif (*Non-Exclusive Royalty free Right*) atas penelitian skripsi saya yang berjudul:

Pengolahan Citra Digital Untuk Ekstraksi Fitur Pada Hybrid Model *CNN-SVM* Dalam Deteksi Penyakit Pneumonia Berbasis X-Ray Dada

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non-Eksekutif ini, Universitas Muhammadiyah Sumatera Utara berhak menyimpan, mengalih media, memformat, mengelola dalam bentuk database, merawat dan mempublikasikan Skripsi saya ini tanpa meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemegang dan atau sebagai pemilik hak cipta.

Demikian pernyataan ini dibuat dengan sebenarnya.

Binjai, 28 Februari 2025
Yang membuat pernyataan



Muhammad Aulia Irza
NPM. 2109020171

RIWAYAT HIDUP

DATA PRIBADI

Nama Lengkap : Muhammad Aulia Irza Tempat dan
Tanggal Lahir : Binjai, 05 Juni 2003
Alamat Rumah : Jln. Gunung Jayawijaya Lk x
Telepon/Faks/HP : 0895621044067
E-mail : aulairza566@gmail.com
Instansi Tempat Kerja : -
Alamat Kantor : -

DATA PENDIDIKAN

SD : SD Negeri 020252 Binjai TAMAT: 2016
SMP : SMP Negeri 1 Binjai TAMAT: 2019
SMA : SMA Negeri 1 Binjai TAMAT: 2021

KATA PENGANTAR



Segala puji dan syukur Penulis panjatkan ke hadirat Allah SWT atas limpahan rahmat dan karunia-Nya sehingga Penulis berhasil menyelesaikan penulisan skripsi ini. Karya ilmiah ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer di Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara.

Penulis menyampaikan rasa terima kasih yang mendalam kepada kedua orang tua atas segala bentuk dukungan, baik secara moril maupun materil, yang telah diberikan selama ini. Penulis juga menyadari bahwa penyusunan skripsi ini tidak akan dapat diselesaikan tanpa adanya bantuan, arahan, serta dukungan dari berbagai pihak sejak awal perkuliahan hingga tahap akhir penulisan. Dengan penuh rasa hormat dan rendah hati, Penulis mengucapkan terima kasih kepada semua pihak yang telah berkontribusi.

Penulis ingin menyampaikan terima kasih dan rasa hormat kepada:

1. Bapak Prof. Dr. Agussani, M.AP., selaku Rektor Universitas Muhammadiyah Sumatera Utara (UMSU).
2. Bapak Dr. Al-Khowarizmi, S.Kom., M.Kom. selaku Dekan Fakultas Ilmu Komputer dan Teknologi Informasi (FIKTI) UMSU.
3. Bapak Halim Maulana, S.T., M.Kom, selaku Wakil Dekan I Fakultas Ilmu Komputer dan Teknologi Informasi.

4. Bapak Dr. Lutfi Basit, S.Sos., M.I.Kom, selaku Wakil Dekan III Fakultas Ilmu Komputer dan Teknologi Informasi.
5. Ibu Fatma Sari Hutagalung, S.Kom., M.Kom, selaku Ketua Program Studi Teknologi Informasi, atas segala arahan, motivasi, dan dukungan selama masa studi hingga penyusunan skripsi ini.
6. Bapak Mhd. Basri, S.Si., M.Kom, selaku Sekretaris Program Studi Teknologi Informasi.
7. Bapak Dr. Irvan, M.Si., selaku dosen pembimbing yang telah membimbing dan memberikan masukan berharga dalam penyusunan skripsi ini.
8. Bapak Halim Maulana, S.T., M. Kom selaku Dosen Penguji I, dan Bapak Rizaldy Khair, S.Kom., M.Kom., selaku Dosen Penguji II, atas waktu, saran, serta penilaiannya terhadap karya ilmiah ini.
9. Segenap pihak yang telah berkontribusi, baik secara langsung maupun tidak langsung, yang tidak dapat disebutkan satu per satu, namun perannya sangat berarti dalam menyelesaikan skripsi ini.

Binjai, 28 Februari 2025

Penulis



Muhammad Aulia Irza

**PENGOLAHAN CITRA DIGITAL UNTUK EKSTRAKSI FITUR PADA
HYBRID MODEL CNN-SVM DALAM DETEKSI PENYAKIT
PNEUNOMIA BERBASIS X-RAY DADA**

ABSTRAK

Pneumonia merupakan salah satu penyakit infeksi paru-paru yang masih menjadi penyebab utama kematian, terutama pada anak-anak dan lansia. Deteksi dini dan akurat terhadap pneumonia sangat penting untuk mendukung pengambilan keputusan medis yang cepat dan tepat. Penelitian ini bertujuan untuk mengembangkan sistem deteksi pneumonia berbasis citra X-ray dada dengan pendekatan hybrid yang menggabungkan *Convolutional Neural Network* (CNN) sebagai ekstraktor fitur dan *Support Vector Machine* (SVM) sebagai algoritma klasifikasi. Dataset yang digunakan berasal dari Guangzhou Women and Children's Medical Center, terdiri dari citra X-ray dada yang telah dikategorikan ke dalam dua kelas, yaitu "Normal" dan "Pneumonia". Proses pra-pemrosesan dilakukan dengan normalisasi dan augmentasi citra. Ekstraksi fitur dilakukan menggunakan arsitektur CNN dengan lima blok konvolusional dan satu lapisan GlobalAveragePooling. Fitur yang dihasilkan dari proses ini kemudian diklasifikasikan menggunakan SVM dengan parameter terbaik yang diperoleh melalui GridSearch dan validasi silang Stratified K-Fold. Hasil pengujian menunjukkan bahwa model hybrid CNN-SVM mencapai akurasi sebesar 92.75% pada data pelatihan dan 81.57% pada data pengujian. Meskipun akurasi pada data validasi hanya 50%, hal ini disebabkan oleh jumlah data validasi yang sangat terbatas. Model menunjukkan sensitivitas tinggi terhadap kasus pneumonia, yang merupakan metrik penting dalam deteksi penyakit. Implementasi akhir dilakukan dengan mengintegrasikan model ke dalam aplikasi antarmuka sederhana berbasis Gradio untuk memudahkan pengguna melakukan deteksi secara mandiri. Penelitian ini menunjukkan bahwa pendekatan hybrid CNN-SVM efektif digunakan untuk deteksi pneumonia berbasis citra X-ray dada dan berpotensi dikembangkan lebih lanjut sebagai alat bantu diagnosis di bidang medis.

Kata Kunci: Pneumonia, Citra X-ray Dada, CNN, SVM, Ekstraksi Fitur, Klasifikasi Citra, Deteksi Penyakit, Machine Learning

DIGITAL IMAGE PROCESSING FOR FEATURE EXTRACTION IN CNN-SVM HYBRID MODEL FOR CHEST X-RAY BASED PNEUNOMIA DISEASE DETECTION

ABSTRACT

Pneumonia is one of the leading causes of death, especially among children and the elderly, due to lung infections. Early and accurate detection of pneumonia is crucial to support prompt and effective medical decision-making. This research aims to develop a pneumonia detection system based on chest X-ray images using a hybrid approach that combines Convolutional Neural Network (CNN) as a feature extractor and Support Vector Machine (SVM) as the classification algorithm. The dataset used in this study originates from the Guangzhou Women and Children's Medical Center, consisting of chest X-ray images categorized into two classes: "Normal" and "Pneumonia." Preprocessing steps involved normalization and image augmentation. Feature extraction was performed using a CNN architecture composed of five convolutional blocks followed by a Global Average Pooling layer. The extracted features were then classified using an SVM model optimized through Grid Search and Stratified K-Fold cross-validation. Experimental results showed that the hybrid CNN-SVM model achieved an accuracy of 92.75% on the training data and 81.57% on the test data. Although the validation accuracy was only 50%, this was primarily due to the limited number of validation images. The model demonstrated high sensitivity in identifying pneumonia cases, which is a critical metric in medical diagnosis. The final implementation involved integrating the model into a Gradio-based application interface to enable user-friendly, real-time detection. This study demonstrates that the hybrid CNN-SVM approach is effective for pneumonia detection based on chest X-ray images and has promising potential as a diagnostic aid in the medical field.

Keywords: Pneumonia, Chest X-ray Image, CNN, SVM, Feature Extraction, Image Classification, Disease Detection, Machine Learning

DAFTAR ISI

KATA PENGANTAR	i
ABSTRAK	iii
ABSTRACT	iv
DAFTAR ISI	v
DAFTAR TABEL	vii
DAFTAR GAMBAR	viii
BAB I PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah.....	7
1.3 Batasan Masalah	7
1.4 Tujuan Penelitian.....	8
1.5 Manfaat Penelitian.....	8
BAB II LANDASAN TEORI	10
2.1 Pengolahan Citra Digital	10
2.2 Anatomi dan Fisiologi Paru-Paru	11
2.3 Pneumonia	13
2.4 Machine Learning.....	14
2.5 Convolutional Neural Network (CNN)	15
2.6 SVM (Support Vector Machine)	18
2.7 Python.....	19
2.8 Confusion Matrix	19
2.9 Penelitian Terkait.....	21
BAB III METODOLOGI PENELITIAN	27
3.1 Pendekatan Penelitian	27
3.2 Teknik Pengumpulan Data.....	28
3.3 Tahapan Penelitian	30
BAB IV HASIL DAN PEMBAHASAN	37
4.1 Gambaran Umum Dataset	37

4.2 Hasil Preprocessing Data.....	38
4.3 Arsitektur dan Hasil Ekstraksi Fitur CNN	40
4.4 Pelatihan dan Optimasi Model SVM.....	44
4.5 Evaluasi Performa Model	46
4.6 Penyesuaian Threshold dan Analisis	49
4.7 Visualisasi Prediksi Model	50
4.8 Implementasi Model ke Aplikasi.....	52
4.9 Pembahasan dan Analisis	54
BAB V PENUTUP	57
5.1 Kesimpulan.....	57
5.2 Saran.....	58
DAFTAR PUSTAKA	59
LAMPIRAN.....	63

DAFTAR TABEL

Tabel 2.1.....	24
Tabel 4.1.....	41

DAFTAR GAMBAR

Gambar 2.1.....	12
Gambar 2.2.....	16
Gambar 2.3.....	20
Gambar 3.1.....	28
Gambar 3.2.....	30
Gambar 3.3.....	34
Gambar 3.4.....	34
Gambar 3.5.....	35
Gambar 3.6.....	35
Gambar 4.1.....	39
Gambar 4.2.....	40
Gambar 4.3.....	43
Gambar 4.4.....	45
Gambar 4.5.....	46
Gambar 4.6.....	47
Gambar 4.7.....	48
Gambar 4.8.....	51
Gambar 4.9.....	51
Gambar 4.10.....	53

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Perkembangan jaman saat ini kita sekarang menghadapi masalah yang semakin rumit, yang menuntut kreativitas di samping kemajuan pengetahuan dan teknologi yang cepat. Kemajuan metode *image processing* saat ini semakin pesat, khususnya di bidang pengolahan gambar, yang merupakan teknologi digital. Berbagai bidang telah terkena dampak dari kemajuan teknologi digital. Sektor kesehatan adalah salah satu industri di mana teknologi telah menjadi lebih terdigitalisasi. Dalam hal ini, image processing melayani tujuan teknologi, seperti dalam bidang kedokteran dan medis, di mana hal ini disebut sebagai pengolahan citra medis.

Pencitraan medis merupakan proses visualisasi pada bagian-bagian dan organ tubuh agar dokter dapat mendiagnosis, menganalisis, dan mengobati penyakit pada pasien. Pencitraan X atau sinar-x (X-ray) adalah salah satu jenis pencitraan medis. Dengan menggunakan teknologi ini, ahli radiologi dapat dengan cepat memastikan status internal pasien mereka. Sebagai contoh, rontgen dada dapat menunjukkan kondisi paru-paru, jantung, dan trakea, serta membantu mendeteksi infeksi paru-paru. Sinar-X dapat digunakan untuk mengklasifikasikan gangguan paru-paru secara tepat dan efisien, yang mempercepat proses diagnostik bagi ahli radiologi dan dokter. (All Allied Health Schools, n.d., 2024)

Deteksi citra untuk pneumonia adalah bidang utama yang menarik dalam pemrosesan citra medis. Pneumonia adalah salah satu penyakit yang dapat berakibat fatal bahkan dapat menyebabkan kematian. Banyak mikroorganisme

menular dan tidak menular, seperti bakteri, virus, atau jamur, dapat menyebabkan pneumonia, penyakit radang parenkim paru-paru (alveoli) yang mengganggu fungsi paru-paru. (Rahman dkk., 2020) *Streptococcus pneumoniae* dan *Haemophilus influenzae* tipe B adalah penyebab pneumonia yang paling umum terjadi. Keduanya merupakan contoh flora bakteri normal di dalam tubuh, yang mungkin muncul ketika sistem kekebalan tubuh terganggu. Mikroorganisme lain penyebab pneumonia adalah *Staphylococcus aureus*, pada golongan virus yaitu *influenza*, *adenovirus*, *respiratory syncytial virus*, *virus Hantala*, *rhinovirus*, *virus herpes simpleks*, *sitomegalovirus*, *mikoplasma*, *pneumokokus*, dan *streptokokus*. Sedangkan *Aspergifosis*, *histoplasmosis*, dan *kandidiasis* adalah bagian dari kelompok jamur. (Fida, 2020)

Berdasarkan laporan dari UNICEF tahun 2018 penyakit pneumonia masih menjadi penyebab utama dari kematian balita di dunia dengan banyak data yang tercatat sekitar 16% dari 5,6 juta kematian balita. Pada tahun 2016. Pneumonia tercatat menewaskan sekitar 2.400 anak setiap hari dan pada tahun tersebut total angka kematian anak karena pneumonia mencapai 880.000 anak. (WHO, 2023) Pneumonia juga dapat menyerang orang dewasa khususnya lansia yang sudah berumur diatas 65 tahun ataupun orang yang sudah memiliki masalah kesehatan sebelumnya. Seseorang yang tidak terkena pneumonia, udara akan mengisi kantung-kantung kecil yang membentuk paru-paru. *Alveoli* yang mengembang dengan cairan dan nanah adalah tanda ketika seseorang terkena pneumonia, yang membuatnya sulit bernapas dan menurunkan kapasitasnya untuk menyerap oksigen. Sehingga oksigen yang dihirup menjadi lebih sedikit dan terasa nyeri ketika melakukan pernapasan.

Menurut data Riset Kesehatan Dasar (Riskesdas) 2018 di Indonesia, jumlah penderita pneumonia meningkat. Angka ini mencapai 2,5 persen untuk penduduk berusia 55 hingga 64 tahun, 3,0% untuk penduduk berusia 65 hingga 74 tahun, dan 2,9% untuk penduduk berusia 75 tahun ke atas. (World Pneumonia Day 2022 – Kementerian Kesehatan, 2022)

Beberapa tahun terakhir di beberapa kota terbesar di Indonesia, polusi udara telah menjadi perhatian serius. Kualitas udara yang sangat buruk dan mengancam kesehatan merupakan hasil dari tingginya polusi udara. Emisi polutan, seperti partikel halus, knalpot kendaraan bermotor, dan limbah industri, telah meningkat karena pertumbuhan penduduk yang cepat, urbanisasi yang tidak terkendali, dan industri yang berkembang pesat. Kota-kota besar di Indonesia seperti Jakarta, Tangerang Selatan, Bandung, khususnya Medan memiliki polusi dan kualitas udara yang sangat berbahaya. Kualitas udara yang tidak sehat dapat menyebabkan berbagai penyakit pernapasan dengan tingkat kematian yang tinggi, seperti pneumonia, dan konsekuensi negatif dari polusi udara cukup nyata, seperti yang terlihat dari peningkatan kejadian alergi, gangguan pernapasan, dan penyakit kronis lainnya. Menurut statistik MerceMarshBenefit, Indonesia mengalami lonjakan kasus pneumonia sebesar 56,9% pada orang dewasa dan 88,1% pada anak-anak selama Semester I tahun 2023 dibandingkan dengan Semester I tahun 2022.

Dari banyaknya kasus pneumonia dari data di atas dilakukannya tes untuk mengidentifikasi pneumonia sangat diperlukan karena banyaknya kasus penyakit yang telah dibahas sebelumnya. Rontgen dada, CT scan paru, USG dada, biopsi jarum paru, dan MRI dada adalah tes yang dilakukan untuk mendiagnosis

pneumonia. Karena rontgen dada saat ini merupakan alat pencitraan diagnostik yang paling populer dan mudah diakses, maka rontgen dada merupakan salah satu cara terbaik untuk mengidentifikasi pneumonia.

Gambar rontgen dada yang khas untuk penyakit pneumonia ditandai dengan gambar radiopak atau segmen yang tampak padat pada bagian paru yang terkena akibat pembentukan eksudat inflamasi yang menggantikan udara dalam *alveoli*. Gambar radiologi, seperti gambar rontgen dada, dapat memberikan deteksi dini dan tindakan yang cepat. Pneumonia harus diidentifikasi secara dini dan akurat untuk mencegah konsekuensi yang parah, terutama pada pasien usia muda dan lanjut usia. (Alsharif dkk., 2021)

Diagnosis yang cepat dan akurat sangat penting dalam menentukan campuran medis yang efektif. Namun, metode diagnostik konvensional seperti pemeriksaan X-ray dada masih menghadapi berbagai keterbatasan. Pemeriksaan pneumonia secara klinis memerlukan sumber daya yang besar, memakan waktu, serta sangat bergantung pada keahlian radiologi dalam menafsirkan hasil citra. Proses mendapatkan hasil foto rontgen dada bervariasi tergantung pada kebijakan dan prosedur masing-masing dari fasilitas kesehatan. Dalam kasus darurat, hasilnya dapat tersedia dalam beberapa jam, tetapi untuk situasi non-darurat, analisis dan pelaporannya bisa memakan waktu hingga beberapa hari.

Meskipun pneumonia dapat dideteksi melalui foto rontgen, kualitas citra yang dihasilkan tidak selalu optimal. Beberapa faktor seperti noise, kontras rendah, atau kemiripan visual dengan penyakit paru lainnya, seperti tuberkulosis, pneumothorax, infiltrasi, nodul, dan kanker paru-paru, dapat menyebabkan kesulitan dalam diagnosis. Hal ini menjadi tantangan bagi dokter radiologi untuk

membedakan jenis penyakit secara akurat.

Untuk mengatasi kendala tersebut, saat ini telah dikembangkan berbagai sistem berbasis *Computer-Aided Diagnosis (CAD)* guna membantu dokter dalam menganalisis citra medis. Dengan kemajuan teknologi kecerdasan buatan, khususnya dalam bidang *Deep Learning*, berbagai metode telah diterapkan untuk meningkatkan akurasi deteksi penyakit paru. Salah satu pendekatan yang banyak digunakan dalam dunia medis adalah *Convolutional Neural Network (CNN)*, yang mampu mengekstraksi fitur penting dari citra X-ray dada secara otomatis. Namun, CNN masih memiliki keterbatasan dalam hal memprediksi dan sensitivitas terhadap variasi citra. Oleh karena itu, pendekatan *Hybrid Model*, seperti kombinasi CNN dengan *Support Vector Machine (SVM)*, menjadi solusi yang potensial untuk meningkatkan akurasi deteksi pneumonia.

Beberapa studi literatur mengungkapkan bahwa banyak upaya yang dilakukan untuk menggunakan teknik *Artificial Intelligence (AI)* dan *Deep Learning* untuk mendeteksi ada tidaknya penyakit pneumonia.

Penelitian yang dilakukan oleh Ilham Jawas dan rekan-rekannya pada tahun 2024 berkaitan erat dengan pemanfaatan pengolahan citra digital dalam sistem machine learning untuk mendiagnosis penyakit pada paru-paru. Studi tersebut mengembangkan sistem berbasis Python yang memanfaatkan berbagai framework untuk melakukan tahapan seperti peningkatan kualitas gambar, segmentasi area paru, hingga proses ekstraksi fitur. Selanjutnya, algoritma klasifikasi digunakan untuk membedakan antara gambar paru-paru normal dan yang menunjukkan gejala pneumonia. Temuan dalam penelitian ini memperlihatkan bahwa sebelum dilakukan proses pengolahan citra, akurasi sistem berada di angka 75%. Namun

setelah penerapan teknik pemrosesan citra, akurasi meningkat menjadi 78%. Fakta ini menunjukkan bahwa penerapan pengolahan citra digital dapat meningkatkan ketepatan diagnosis berbasis machine learning, meskipun tingkat akurasi yang dicapai masih tergolong rendah. (Jawas dkk., 2024)

Dalam penelitian lain yang dilakukan oleh Ami Rahmawati dkk. 2025, yang membangun model klasifikasi citra menggunakan dataset radiografi thorax dari pasien kelompok balita dengan menerapkan arsitektur *ResNet-50*, sebuah varian dari *Convolutional Neural Network (CNN)*. Dalam penelitian tersebut, metode *data augmentation* diterapkan untuk meningkatkan keragaman data pelatihan, sementara arsitektur CNN digunakan untuk mengidentifikasi pola pada citra X-ray dada secara otomatis. Setelah melalui proses pelatihan dan evaluasi, model yang dikembangkan berhasil mencapai tingkat akurasi sebesar 85%, yang menunjukkan bahwa pendekatan ini cukup efektif dalam mengklasifikasikan pneumonia secara otomatis. (Rahmawati dkk., 2025)

Karena dapat secara otomatis mengekstrak fitur-fitur penting dari gambar, CNN adalah salah satu metode *deep learning* yang sering digunakan dalam pemrosesan citra medis. Terlepas dari kelebihan dalam ekstraksi fitur, kinerja klasifikasi CNN terkadang masih memiliki kekurangan. Oleh karena itu, dalam penelitian ini SVM dianggap sebagai salah satu metode klasifikasi terbaik untuk memisahkan data non-linear dengan margin yang optimal, hasil ekstraksi fitur CNN akan digunakan sebagai masukan (*input feature*) untuk SVM.

Dengan menggabungkan keunggulan SVM dalam klasifikasi dan CNN dalam ekstraksi fitur, pendekatan *hybrid model* CNN-SVM ini diharapkan dapat untuk meningkatkan akurasi deteksi pneumonia. Dibandingkan dengan metode

konvensional. Dengan metode ini memungkinkan sistem diagnosis berbasis citra X-ray dada dapat berfungsi lebih efisien dalam mendeteksi keberadaan pneumonia dengan tingkat akurasi yang lebih tinggi.

Penelitian ini bertujuan untuk menilai kemampuan atau tingkat akurasi model CNN-SVM dalam mengidentifikasi pneumonia secara akurat dan menganalisis seberapa baik performanya ketika diuji pada berbagai dataset. Oleh karena itu, penelitian ini diharapkan mampu memberikan sumbangsih dalam mengembangkan sistem pendeteksian dini yang lebih efisien dan akurat, sehingga dapat membantu mempercepat diagnosis serta mendukung penanganan yang lebih efektif dalam dunia medis.

1.2 Rumusan Masalah

1. Mengembangkan sistem deteksi pneumonia berbasis citra X-ray dada menggunakan pendekatan hybrid CNN-SVM agar mampu menghasilkan klasifikasi yang akurat dan efisien.
2. Menganalisis efektivitas kombinasi CNN sebagai *feature extractor* dan SVM sebagai algoritma klasifikasi dalam mendeteksi pneumonia.
3. Menganalisis performa model hybrid CNN-SVM dalam mendeteksi pneumonia ditinjau dari metrik evaluasi seperti akurasi, sensitivitas, spesifisitas, presisi.
4. Mengidentifikasi tantangan - tantangan dalam penerapan model hybrid CNN-SVM untuk deteksi pneumonia.

1.3 Batasan Masalah

1. Penelitian ini hanya berfokus pada klasifikasi biner, yaitu normal dan pneumonia. Tidak dilakukan klasifikasi lebih lanjut antara *bacterial*

pneumonia dan *viral pneumonia*.

2. Dalam penelitian ini metode yang digunakan adalah *Hybrid CNN-SVM*, dimana CNN digunakan sebagai ekstraktor fitur dan SVM sebagai algoritma klasifikasi. Penelitian tidak membandingkan metode lain seperti *Random Forest*, k-NN, dan metode lainnya.
3. Kinerja model pada penelitian ini dievaluasi menggunakan metrik akurasi, sensitivitas (*recall*), spesifisitas, dan presisi.
4. Data yang digunakan adalah citra X-ray dada dalam format JPEG yang diperoleh dari dataset Guangzhou Women and Children's Medical Center yang tersedia di situs Kaggle.
5. Pengidentifikasian pada penelitian ini dirancang menggunakan Google Colaboratory.

1.4 Tujuan Penelitian

1. Mengembangkan sistem deteksi pneumonia berbasis citra X-ray dada dengan pendekatan *Hybrid CNN-SVM* untuk meningkatkan akurasi klasifikasi.
2. Meneliti efektivitas kombinasi CNN sebagai *feature extractor* dan SVM sebagai algoritma klasifikasi dalam mendeteksi pneumonia.
3. Menyediakan wawasan mengenai penerapan teknologi pengolahan citra digital dalam bidang medis, khususnya dalam mendukung deteksi penyakit berbasis pencitraan radiologi.

1.5 Manfaat Penelitian

1. Penelitian ini diharapkan dapat menambah wawasan dalam bidang *computer vision* dan pengolahan citra medis, khususnya dalam penerapan metode *Hybrid CNN-SVM* untuk deteksi pneumonia. Selain itu, hasil penelitian ini

dapat dijadikan referensi dalam pengembangan teknologi berbasis kecerdasan buatan untuk diagnosis penyakit berbasis pencitraan radiologi menggunakan *machine learning*.

2. Dengan memanfaatkan kombinasi CNN sebagai ekstraktor fitur dan SVM sebagai algoritma klasifikasi, dengan penelitian ini diharapkan dapat meningkatkan efektivitas analisis citra X-ray dada dalam mendeteksi pneumonia. Metode yang dikembangkan diharapkan dapat menjadi alternatif atau referensi bagi tenaga medis dalam melakukan diagnosis secara lebih cepat dan akurat.
3. Hasil penelitian ini berpotensi mendukung peningkatan aksesibilitas teknologi deteksi pneumonia, terutama di fasilitas kesehatan dengan keterbatasan tenaga ahli radiologi. Selain itu, penelitian ini juga dapat mendorong pemanfaatan kecerdasan buatan dalam dunia medis guna mendukung proses diagnosis penyakit yang lebih efisien dan presisi.

BAB II

LANDASAN TEORI

2.1 Pengolahan Citra Digital

Pengolahan citra digital merupakan teknik yang dimanfaatkan untuk memproses dan menganalisis gambar dalam format digital, sehingga memungkinkan gambar tersebut dianalisis atau diubah untuk tujuan tertentu. Di dalam bidang medis, pengolahan citra digital berkontribusi secara signifikan dalam meningkatkan ketepatan diagnostik dengan mengambil data penting dari citra medis, termasuk X-ray dada. Metode ini mencakup beberapa tahap, seperti meningkatkan kualitas citra, segmentasi, dan mengekstraksi fitur sebelum analisis selanjutnya. (Gonzalez dan Woods, 2018 dalam Sharma & Parveen., 2021).

Dengan menggunakan teknik segmentasi dan peningkatan kontras untuk memperjelas area yang mengalami kelainan pada X-ray dada, pengolahan citra digital dapat meningkatkan kualitas gambar, telah dibuktikan bahwa metode ini mengungguli analisis manual dalam meningkatkan deteksi pneumonia. (Sari & Tampubolon., 2022) Selain itu, penelitian Rahman dkk. dari tahun 2021 menunjukkan bahwa menggabungkan metode *deep learning* dan pengolahan citra dapat meningkatkan ketepatan diagnosis penyakit berdasarkan pencitraan radiologi. (Rahman dkk., 2021)

Pentingnya proses *preprocessing* dalam pengolahan citra digital sebelum dilakukan klasifikasi. Dalam penelitian tersebut, digunakan teknik normalisasi RGB pada citra untuk meningkatkan kualitas gambar dan membantu sistem dalam mengenali pola lebih akurat saat menggunakan algoritma *Convolutional Neural Network* (CNN). Hasil yang diperoleh menunjukkan bahwa proses *preprocessing*

seperti normalisasi citra mampu meningkatkan performa deteksi citra digital secara signifikan. (Khairullah dkk., 2020)

Dengan perkembangan teknologi pengolahan citra digital, khususnya dalam analisis X-ray dada, berbagai metode terus dikembangkan untuk meningkatkan akurasi deteksi penyakit, termasuk pneumonia. Oleh karena itu, penggunaan teknik pengolahan citra yang tepat menjadi kunci dalam mendukung sistem diagnosis berbasis pencitraan medis.

2.2 Anatomi dan Fisiologi Paru-Paru

Organ utama dalam sistem pernapasan manusia yang terlibat dalam proses transfer gas antara sistem pernapasan dan aliran darah adalah paru-paru. Organ ini disebut paru-paru kiri dan kanan, yang terletak di dada rongga. Karena adanya jantung rongga, paru-paru kiri memiliki dua lobus, sedangkan paru-paru kanan memiliki tiga lobus. Struktur ini memungkinkan fungsi pernapasan yang optimal. (Hall & Guyton, 2020).

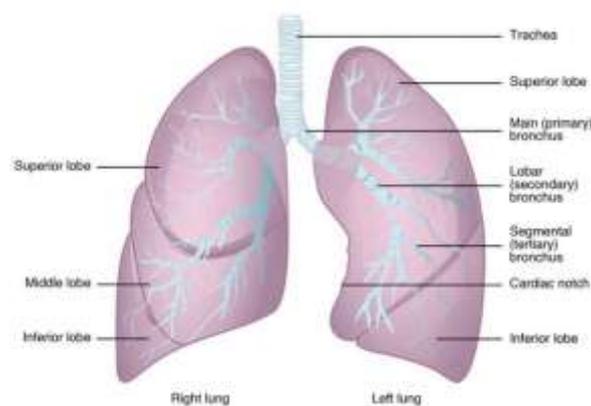
Menurut Zhang dkk, Struktur paru-paru mencakup bronkus, bronkiolus, serta alveoli sebagai komponen utamanya. adalah konstituen utama dari struktur paru-paru yang rumit, Alveoli adalah kantung udara kecil yang sangat penting untuk mengalirkan karbondioksida dan oksigen. Dinding alveoli yang tipis memungkinkan karbon dioksida dikeluarkan dari tubuh saat menghembuskan napas dan oksigen masuk ke dalam sirkulasi. (Zhang dkk., 2024)

Selain itu, sebuah studi oleh Ou dkk, menunjukkan bahwa mekanisme ventilasi, yang meliputi proses inspirasi dan ekspirasi, sangat penting bagi fisiologi paru-paru. Proses inspirasi melibatkan otot diafragma yang berkontraksi, yang meningkatkan volume rongga dada dan menurunkan tekanan intrapulmoner,

sehingga memungkinkan udara masuk ke dalam paru-paru, proses ekspirasi melibatkan relaksasi diafragma, yang meningkatkan tekanan di dalam paru-paru dan mengeluarkan udara dari dalam tubuh. (Oudijk, 2025)

Elastisitas jaringan paru-paru memainkan peran krusial dalam fungsi pernapasan. Penurunan elastisitas, yang dapat disebabkan oleh kondisi seperti pneumonia, mengarah pada penurunan kapasitas paru-paru dan kesulitan dalam pertukaran gas. Oleh karena itu, pemahaman mendalam tentang struktur dan fungsi paru-paru sangat penting untuk deteksi dini penyakit paru-paru, terutama melalui analisis pencitraan medis seperti X-ray dada. (Hader dkk., 2023)

Dengan memahami struktur dan mekanisme kerja paru-paru, pengolahan citra medis dapat dikembangkan lebih lanjut untuk membantu mendeteksi penyakit paru-paru secara lebih akurat. Seiring dengan kemajuan teknologi, pendekatan berbasis pencitraan medis seperti X-ray dan teknik *deep learning* menjadi semakin penting dalam diagnosis penyakit paru-paru.



Gambar 2.1. Anatomi Paru-Paru

Sumber: www.kompas.com

2.3 Pneumonia

Pneumonia merupakan suatu kondisi infeksi akut yang memengaruhi jaringan paru-paru, khususnya alveoli, dan dapat dipicu oleh mikroorganisme seperti bakteri, virus, jamur, maupun parasit. Infeksi tersebut memicu terjadinya inflamasi pada alveoli, yang mengakibatkan gangguan pertukaran gas dalam paru-paru. Gejala umum yang sering muncul meliputi batuk, demam, sesak napas, dan peningkatan produksi dahak. pneumonia dapat diklasifikasikan berdasarkan penyebab infeksi, seperti *bacterial pneumonia*, *viral pneumonia*, dan *fungal pneumonia*, di mana setiap jenis memiliki karakteristik klinis yang berbeda. (Chendkk., 2021)

Salah satu pertanyaan terpenting tentang apakah penyebab penyakitnya adalah virus atau bakteri. Infeksi biasanya dimulai di trakea dan berkembang ke parenkim paru-paru. Sebaliknya, tanda pertama infeksi pada pneumonia yang disebabkan oleh virus adalah nyeri di area pernapasal yang disebabkan oleh lesi pada epitel pernapasan. Jika terjadi infeksi yang disebabkan oleh virus atau bakteri, hewan akan mengalami ketidaknyamanan atau kematian akibat infeksi tersebut. Karena saluran pernapasan anatomis yang lebih kecil, anak-anak, terutama bayi, mengalami infeksi yang lebih parah. Hipoksemia, atau keadaan di mana kadar oksigen terdapat dalam darah rendah, akan diakibatkan oleh penyumbatan saluran napas. (Kelly & Sandora, 2016, dalam Fauziyah., 2020)

Tujuan mendiagnosis pneumonia adalah untuk memandu pemberian perawatan atau terapi. Hal ini melibatkan penentuan tingkat keparahan, bentuk, dan luasnya penyakit serta perkiraan jenis mikroba yang menyebabkan infeksi. Secara klinis, adanya gambaran konsolidasi pada X-ray dada dan adanya kelainan

fisis merupakan dasar untuk mendiagnosis pneumonia. Pada radiografi dada, konsolidasi tampak sebagai bintik-bintik kabur pada paru-paru. Namun, diagnosis anatomis dan etiologis diperlukan untuk diagnosis yang lengkap. Banyak teknik yang dapat digunakan untuk mendiagnosis pneumonia, yaitu:

a. Pemeriksaan laboratorium

Studi ini didasarkan pada sampel darah atau dahak pasien.

b. Radiografi dada

Selain mengidentifikasi adanya penyebaran, pemeriksaan sinar-X dapat mengungkapkan hasil penumpukan nanah (abses) multipel, stafilocokus, distribusi atau area masuknya bakteri ke dalam jaringan.

c. Analisis gas darah

Studi ini dilakukan dengan mengukur ambang hipoksemia darah.

2.4 Machine Learning

Machine Learning (ML) merupakan cabang dari kecerdasan buatan (AI) yang dirancang agar sistem dapat memperoleh pengetahuan melalui data dan meningkatkan kinerjanya secara otomatis tanpa pemrograman eksplisit. Algoritme ML digunakan untuk mengidentifikasi mengenali struktur data untuk menghasilkan prediksi atau estimasi berdasarkan temuan tersebut. Beberapa pendekatan, seperti *supervised learning*, dan *reinforcement learning*, dapat digunakan untuk menjalankan proses pembelajaran dalam *machine learning*. (Mishra., 2024)

Dalam bidang kesehatan untuk membantu mendiagnosis penyakit, *machine learning* sering digunakan untuk menganalisis data medis, termasuk citra X-ray. (Harahap dkk., 2022) Telah dibuktikan bahwa pendekatan *machine learning* (ML)

seperti *Convolutional Neural Network* (CNN) dan *Support Vector Machine* (SVM) berguna untuk mengidentifikasi kelainan, termasuk pneumonia. (Mardianto dkk., 2024)

Dengan berbagai pendekatan ini, *Machine Learning* telah menjadi teknologi yang berkontribusi besar dalam era revolusi industri 4.0, terutama dalam sektor layanan kesehatan dan pengolahan citra medis.

2.5 Convolutional Neural Network (CNN)

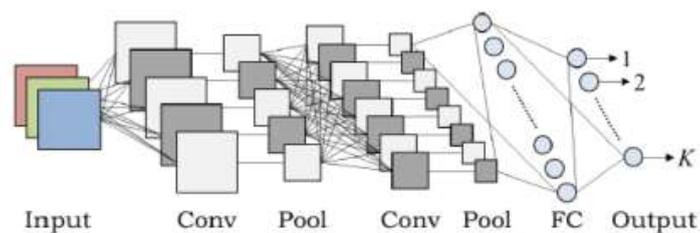
Convolutional Neural Network (CNN) merupakan jenis jaringan saraf tiruan (*Artificial Neural Network* – ANN) dirancang khusus untuk mengekstrak data dari gambar dan data spasial lainnya. CNN memiliki kemampuan untuk menganalisis karakteristik citra secara otomatis, mengurangi kebutuhan rekayasa fitur manual yang biasanya diperlukan dalam metode klasik (*feature engineering*). CNN adalah salah satu model pembelajaran mendalam yang paling umum diterapkan dalam analisis citra medis, termasuk rontgen dada, karena akurasi yang sangat baik dalam mengidentifikasi pola yang rumit. CNN mengidentifikasi karakteristik utama seperti tepi, sudut, dan tekstur dengan menggunakan filter konvolusi yang berjalan di atas gambar. (Zhang dkk., 2021)

Menurut Fauziah. (2020), Arsitektur CNN terdiri dari beberapa lapisan yang bekerja untuk menganalisis data citra.. Arsitektur umum CNN terdiri dari:

a. *Convolution Layer*

Lapisan konvolusi (*convolution layer*) merupakan salah satu bagian utama dalam arsitektur Convolutional Neural Network (CNN) yang berfungsi untuk mengekstraksi fitur dari data masukan. Biasanya, lapisan ini menggabungkan operasi linear dan non-linear untuk mengolah informasi.

Proses konvolusi dilakukan terhadap hasil keluaran dari lapisan sebelumnya, dengan tujuan menyoroti fitur penting dari data. Konvolusi sendiri adalah jenis operasi matematis satu dimensi yang digunakan secara khusus untuk proses ekstraksi fitur. Operasi ini melibatkan penerapan *kernel*—sebuah kumpulan nilai numerik kecil—pada input yang berbentuk array multidimensi, dikenal juga sebagai *tensor*. Kompleksitas dari operasi konvolusi dipengaruhi oleh dua parameter utama, yaitu ukuran filter dan jumlah kernel yang digunakan. Persamaan matematis yang merepresentasikan operasi konvolusi ini dapat dijelaskan melalui Persamaan 1



Gambar 2.2 Arsitektur CNN

Sumber: (Fauziyah dkk., 2020)

$$s(t) = (x * w)(t) \quad (1)$$

Menurut Persamaan 1, x mewakili masukan, w adalah kernel, dan t adalah langkah waktu. Nilai output pada posisi tertentu dalam tensor keluaran diperoleh melalui hasil perkalian setiap elemen kernel dengan elemen yang bersesuaian pada tensor input di lokasi yang sama, lalu dijumlahkan ini dikenal sebagai fitur map, seperti yang ditunjukkan pada Gambar 2.2. Proses ini melibatkan penggunaan beberapa kernel untuk membuat sejumlah besar fitur map.

b. *Pooling Layer*

Proses penentuan ukuran suatu set data dikenal sebagai *pooling layer*. Dalam pemrosesan citra, *pooling* juga bertujuan untuk meningkatkan posisi invariansi fitur. Teknik *pooling* yang digunakan dalam CNN terbesar adalah *max pooling*. *Max pooling* mengubah keluaran lapisan konvolusi menjadi beberapa kisi yang lebih kecil, kemudian menghitung nilai minimum setiap kisi untuk menentukan matriks citra.citra.

c. *Dropout Layer*

Dropout merupakan salah satu teknik regulasi yang digunakan untuk mengurangi risiko *overfitting* serta meningkatkan efisiensi proses pelatihan model. *Overfitting* sendiri terjadi ketika model sangat akurat dalam mempelajari data latih, namun gagal menghasilkan performa yang baik saat diuji pada data baru atau data yang tidak dikenal, sehingga klasifikasinya menjadi tidak optimal.

d. *GlobalAveragePooling2D*

GlobalAveragePooling2D merupakan salah satu jenis lapisan *pooling* yang sering digunakan dalam arsitektur Convolutional Neural Network (CNN). Lapisan ini berfungsi dengan menghitung nilai rata-rata dari seluruh elemen yang terdapat dalam setiap *feature map*, sehingga menghasilkan representasi yang lebih ringkas dari fitur tersebut.

Secara umum, *GlobalAveragePooling2D* digunakan untuk mengurangi dimensi fitur pada *output* konvolusi sebelum masuk ke *fully connected layer* atau langsung ke layer klasifikasi. Lapisan ini menggantikan peran lapisan *fully connected* tradisional dengan cara yang lebih ringan dan efisien, serta

memiliki keuntungan dalam mengurangi jumlah parameter model, sehingga memperkecil risiko *overfitting*.

2.6 SVM (Support Vector Machine)

Support Vector Machine (SVM) merupakan satu dari banyak metode *machine learning* yang digunakan untuk klasifikasi dan regresi. SVM bekerja dengan mencari *hyperplane* yang optimal untuk membagi data ke dalam dua kelas yang berbeda dengan margin terbesar. Ketika dipasangkan dengan kernel yang sesuai, pendekatan ini memiliki manfaat untuk mencegah *overfitting* dan sangat berhasil untuk data berdimensi tinggi. SVM mampu memetakan data ke dalam ruang dengan dimensi yang lebih tinggi melalui penerapan fungsi kernel seperti linear, polinomial, dan *radial basis function* (RBF) untuk menemukan batas keputusan optimal yang memisahkan kelas. (Artiyati & Ahdika., 2024)

SVM juga memiliki peran khususnya dalam mengklasifikasikan sentimen publik terhadap perkembangan teknologi kecerdasan buatan di sektor pendidikan. SVM non-linear digunakan untuk memproses data teks dan berhasil memberikan hasil klasifikasi yang akurat. (Putri dkk., 2024) Kombinasi algoritma SVM dengan teknik N-gram dapat digunakan untuk membangun sistem deteksi berita palsu dalam bahasa Indonesia. Dalam penelitian tersebut, kernel linear memberikan performa terbaik dengan akurasi mencapai 97,4%, menunjukkan bahwa SVM mampu menangani masalah klasifikasi teks dengan baik. (Jollyta dkk., 2023)

Dengan kemampuannya dalam menangani berbagai jenis data dan fleksibilitas dalam pemilihan kernel, SVM menjadi algoritma yang sangat direkomendasikan digunakan dalam tugas klasifikasi, termasuk di bidang kesehatan.

2.7 Python

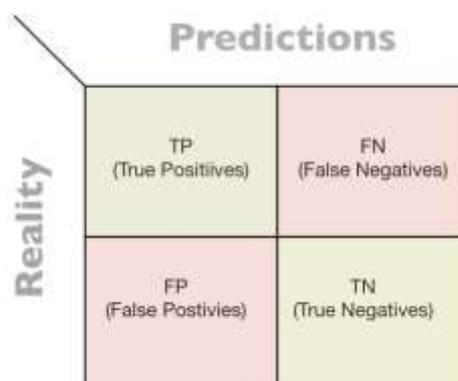
Python merupakan bahasa pemrograman berlevel tinggi yang pertama kali dikembangkan oleh Guido van Rossum dan dirilis pada tahun 1991. Keunggulan Python terletak pada kemampuannya dalam mendukung berbagai pustaka dan kerangka kerja yang memfasilitasi pengembangan aplikasi secara efisien dan efektif. Python adalah bahasa pemrograman multifungsi yang banyak digunakan, terutama dalam bidang *deep learning* dan *machine learning*. Python adalah alat yang sangat membantu untuk membuat model kecerdasan buatan karena kemampuannya untuk menangani data dan mendukung berbagai pustaka, termasuk *TensorFlow* dan *Keras*. (Alfarizi dkk., 2023)

Python adalah bahasa yang sangat populer untuk aplikasi machine learning karena sintaksisnya yang sederhana dan keberagaman pustaka yang tersedia, seperti *scikit-learn*, *TensorFlow*, *Keras*, dan *PyTorch*. Pustaka-pustaka ini memungkinkan pengembang untuk dengan cepat dan efisien membangun model yang kompleks (DigitalOcean, 2025). Dalam konteks analisis data, Python juga digunakan untuk menganalisis pola penyebaran COVID-19 selama pandemi. Python dapat dimanfaatkan untuk mengidentifikasi dinamika dan faktor-faktor yang mempengaruhi penyebaran COVID-19, serta mengeksplorasi kemampuan visualisasi data Python untuk menyajikan informasi yang jelas dan mudah dipahami. (Hermanto dkk., 2023)

2.8 Confusion Matrix

Confusion Matrix merupakan salah satu cara untuk menilai performa suatu model klasifikasi dengan cara membandingkan hasil prediksi model terhadap label yang sebenarnya dalam bentuk tabel. Terdapat empat elemen utama dalam

confusion matrix, yaitu *True Positive (TP)*, *False Positive (FP)*, *True Negative (TN)*, dan *False Negative (FN)*. Keempat komponen ini menjadi dasar perhitungan berbagai metrik evaluasi seperti *accuracy*, *precision*, *recall*, dan *F1-score*, yang secara keseluruhan memberikan penilaian menyeluruh terhadap kinerja model klasifikasi.



Gambar 2.3 Ilustrasi *Confusion Matrix*

Sumber: www.oreilly.com

Gambar 2.3 menampilkan ilustrasi dari matriks konfusi, yang terdiri atas empat komponen utama dalam proses evaluasi klasifikasi, yaitu *False Positive (FP)*, *False Negative (FN)*, *True Positive (TP)*, dan *True Negative (TN)* (Mohajon, 2020). Nilai *True Positive* menunjukkan jumlah prediksi yang benar ketika model berhasil mengidentifikasi kelas positif sesuai dengan label sebenarnya. Sebaliknya, *False Negative* terjadi ketika model gagal mengenali kelas positif yang seharusnya terdeteksi. Sementara itu, *True Negative* menggambarkan kondisi di mana model secara tepat mengklasifikasikan data sebagai negatif sesuai dengan kenyataan. Terakhir, *False Positive* muncul saat model secara keliru mengklasifikasikan data negatif sebagai positif (Mardhotillah, 2018, dalam Hidayat., 2021)

a. Akurasi

Akurasi (*accuracy*) adalah perbandingan antara jumlah prediksi yang akurat dengan jumlah total prediksi akurat dan salah. Berdasarkan tabel di atas maka akurasi dapat dirumuskan:

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (2)$$

b. Presisi

Presisi (*precision*) merujuk pada perbedaan antara prediksi positif sebenarnya dan prediksi positif yang diprediksi untuk label tertentu, atau rasio True Positive terhadap prediksi positif. Presisi dapat diwakili dengan rumus:

$$Presisi = \frac{TP}{(TP+FP)} \quad (3)$$

c. Recall

Recall yang juga dikenal sebagai sensitivitas, menyoroti beberapa model yang baik dalam mengidentifikasi False Negative atau rasio True Positive terhadap *acual positive*. *recall* dapat dihitung dengan rumus:

$$Recall = \frac{TP}{(TP+FN)} \quad (4)$$

d. F1-Score

F1-score adalah kombinasi rata-rata antara ingatan dan recall. *F1-score* bisa dihitung dengan menggunakan rumus:

$$f1 - Score = \frac{2TP}{(2TP+FP+FN)} \quad (5)$$

2.9 Penelitian Terkait

Penelitian oleh (Yopento dkk., 2021) membahas penerapan metode *Convolutional Neural Network* (CNN) dalam deteksi pneumonia pada citra X-ray paru-paru. Dalam penelitian tersebut, CNN digunakan untuk membedakan citra

X-ray menjadi dua kategori, yaitu pneumonia dan normal, dengan mengandalkan ekstraksi fitur *Sobel* untuk mendeteksi tepi-tepi yang relevan dalam gambar. Dataset yang digunakan terdiri dari citra X-ray paru-paru yang telah dikurasi dan diproses terlebih dahulu. Hasil penelitian menunjukkan bahwa model CNN yang dikombinasikan dengan fitur Sobel dapat memberikan akurasi yang tinggi yaitu 91% dalam mendeteksi pneumonia, yang menunjukkan efektivitas CNN dalam aplikasi medis. Penelitian ini juga menekankan pentingnya *preprocessing* citra serta pemilihan fitur yang tepat dalam meningkatkan kinerja model deteksi penyakit.

Penelitian oleh (Khoirunisa & Charibaldi., 2024) mengevaluasi penerapan metode Support Vector Machine (SVM) dalam klasifikasi penyakit menggunakan citra medis dengan ekstraksi fitur *Gray-Level Co-occurrence Matrix* (GLCM), *Local Binary Pattern* (LBP), dan *Histogram of Oriented Gradients* (HOG). Dalam penelitian ini, ketiga metode ekstraksi fitur ini digabungkan untuk meningkatkan akurasi dalam mengidentifikasi jenis penyakit, dengan SVM sebagai metode klasifikasi utama. Hasil pengujian menunjukkan bahwa kombinasi GLCM, LBP, dan HOG mampu memberikan akurasi klasifikasi 74% yang lebih baik dibandingkan dengan penggunaan satu metode ekstraksi fitur saja. Peneliti menyimpulkan bahwa kombinasi ekstraksi fitur yang tepat dengan SVM dapat menghasilkan kinerja yang optimal dalam klasifikasi penyakit pada citra medis, meskipun hasil ini masih tergantung pada kualitas citra yang digunakan.

Penelitian oleh (Al Reshan dkk., 2023) mengevaluasi penggunaan model *MobileNet* untuk deteksi pneumonia pada citra X-ray. Penelitian ini memanfaatkan kekuatan arsitektur *MobileNet* yang ringan dan efisien untuk

mendeteksi pola-pola terkait pneumonia dalam citra medis. Hasil pengujian menunjukkan bahwa model *MobileNet* mampu mencapai akurasi yang tinggi dalam membedakan pneumonia dengan klasifikasi normal. Peneliti menyimpulkan bahwa model *MobileNet* merupakan alternatif yang efektif untuk deteksi penyakit berbasis citra X-ray, dengan keunggulan dalam hal komputasi yang lebih efisien dibandingkan dengan model *deep learning* lainnya seperti CNN, khususnya dalam perangkat dengan keterbatasan sumber daya.

Penelitian oleh (Mabrouk dkk., 2023) mengusulkan pendekatan deteksi pneumonia menggunakan *ensemble* model dari beberapa arsitektur Convolutional Neural Network (CNN) pada citra X-ray dada. Metode ini menggabungkan kekuatan beberapa model CNN populer seperti VGG16, ResNet50, dan DenseNet121 untuk meningkatkan akurasi deteksi. Dataset yang digunakan terdiri dari gambar X-ray dada yang telah melalui proses augmentasi untuk menyeimbangkan jumlah data antara kelas normal dan pneumonia. Hasil penelitian menunjukkan bahwa pendekatan *ensemble* ini mampu mencapai akurasi lebih tinggi dibandingkan dengan penggunaan model tunggal. Penelitian ini menyimpulkan bahwa penggabungan beberapa model CNN dalam satu sistem dapat meningkatkan performa klasifikasi citra medis dan memberikan hasil yang lebih andal untuk diagnosis otomatis pneumonia.

Penelitian oleh (Maysanjaya, 2020) membahas penerapan *Convolutional Neural Network* (CNN) dalam klasifikasi pneumonia berdasarkan citra X-ray paru-paru. Dalam penelitian ini, CNN dibangun dengan arsitektur sederhana yang terdiri dari beberapa lapisan konvolusi, *max-pooling*, dan *fully connected layer*. Data yang digunakan merupakan citra X-ray dari pasien dengan kondisi normal

dan pneumonia yang diolah melalui proses augmentasi untuk meningkatkan variasi data latih. Hasil pengujian menunjukkan bahwa model CNN mampu mencapai akurasi tinggi dalam mengklasifikasikan citra X-ray yaitu 89.58%, menunjukkan bahwa CNN memiliki kemampuan representasi fitur yang kuat dalam mendeteksi kelainan pada paru-paru. Penelitian ini menyimpulkan bahwa CNN sangat efektif dalam tugas klasifikasi citra medis, terutama jika didukung dengan *preprocessing* dan augmentasi data yang tepat.

Tabel 2.1 Penelitian Terkait

No	Nama Peneliti (Tahun)	Penelitian	Hasil
1	Yopento dkk., (2021)	Penerapan metode <i>Convolutional Neural Network</i> (CNN) dalam deteksi pneumonia pada citra X-ray paru-paru.	Model CNN yang digunakan dalam penelitian ini mengandalkan ekstraksi fitur <i>Sobel</i> untuk mendeteksi tepi-tepi yang relevan pada citra X-ray. Hasil penelitian menunjukkan bahwa kombinasi CNN dengan ekstraksi fitur <i>Sobel</i> berhasil mencapai akurasi deteksi pneumonia sebesar 91%. Penelitian ini juga menekankan pentingnya tahap <i>preprocessing</i> citra dan pemilihan fitur yang tepat agar model dapat bekerja secara optimal, mengingat variasi kualitas citra pada dataset.
2	Khoirunisa & Charibaldi, (2024)	Evaluasi metode <i>Support Vector Machine</i> (SVM) dalam klasifikasi penyakit menggunakan citra medis dengan ekstraksi fitur GLCM, LBP, dan HOG.	Penelitian ini mengeksplorasi kombinasi metode ekstraksi fitur GLCM, LBP, dan HOG yang dipadukan dengan SVM untuk meningkatkan akurasi klasifikasi pada citra medis. Hasil pengujian menunjukkan bahwa dengan menggabungkan ketiga

			metode ekstraksi fitur tersebut, akurasi klasifikasi yang diperoleh mencapai 74%. Hasil ini lebih baik dibandingkan dengan penggunaan satu metode ekstraksi fitur saja, namun kualitas citra tetap mempengaruhi hasil akhir, yang menjadi pertimbangan utama dalam pengembangan sistem klasifikasi penyakit berbasis citra medis.
3	Al Reshan dkk., (2023)	Penggunaan model <i>MobileNet</i> untuk deteksi pneumonia pada citra X-ray.	Model <i>MobileNet</i> yang digunakan dalam penelitian ini menunjukkan hasil yang sangat baik dalam membedakan citra pneumonia dengan citra normal. Keunggulan utama dari model ini adalah efisiensi komputasi yang lebih baik dibandingkan dengan model CNN yang lebih besar, seperti VGG atau <i>ResNet</i> . Hasil pengujian menunjukkan bahwa <i>MobileNet</i> dapat mencapai akurasi deteksi yang tinggi pada dataset X-ray dengan sumber daya komputasi yang lebih rendah, sehingga cocok digunakan pada perangkat dengan keterbatasan sumber daya.
4	Mabrouk dkk., (2023)	Deteksi pneumonia menggunakan <i>ensemble</i> model dari beberapa arsitektur CNN (VGG16, <i>ResNet50</i> , <i>DenseNet121</i>) pada citra X-ray dada..	Pendekatan <i>ensemble</i> yang menggabungkan beberapa arsitektur CNN, seperti VGG16, <i>ResNet50</i> , dan <i>DenseNet121</i> , berhasil meningkatkan akurasi deteksi pneumonia dibandingkan dengan penggunaan model CNN tunggal. Hasil penelitian menunjukkan bahwa <i>ensemble</i> model ini mampu mencapai hasil yang lebih andal dan konsisten dalam

			mengklasifikasikan citra X-ray dada. Penggunaan ensemble model memberikan keuntungan dari kekuatan masing-masing arsitektur CNN yang digabungkan, menghasilkan performa deteksi yang lebih tinggi dan akurat.
5	Maysanjaya, (2020)	Penerapan <i>Convolutional Neural Network</i> (CNN) dalam klasifikasi pneumonia berdasarkan citra X-ray paru-paru.	Dalam penelitian ini, model CNN yang dibangun menggunakan arsitektur yang sederhana, terdiri dari beberapa lapisan konvolusi, <i>max-pooling</i> , dan <i>fully connected layer</i> , berhasil mencapai akurasi sebesar 89.58% dalam mengklasifikasikan citra X-ray. Hasil ini menunjukkan bahwa CNN memiliki kemampuan representasi fitur yang kuat dalam mendeteksi kelainan pada paru-paru, terutama dengan <i>preprocessing</i> dan augmentasi data yang optimal.

BAB III

METODOLOGI PENELITIAN

3.1 Pendekatan Penelitian

Penelitian ini menggunakan pendekatan prediktif dalam kerangka kuantitatif, di mana fokus utama penelitian adalah membangun model prediksi berbasis citra X-ray dada untuk mendeteksi pneumonia menggunakan metode *hybrid* CNN-SVM. Pendekatan ini relevan karena bertujuan untuk menghasilkan model klasifikasi yang mampu memprediksi kondisi pasien secara akurat berdasarkan data citra.

Penelitian ini berfokus pada pengolahan citra digital untuk mengekstraksi fitur dari gambar X-ray dada, yang kemudian digunakan sebagai input dalam klasifikasi antara kondisi normal dan pneumonia. Model CNN digunakan untuk mengekstraksi fitur otomatis dari citra, sementara SVM digunakan sebagai *classifier* untuk menghasilkan prediksi akhir. Pemilihan pendekatan *hybrid* CNN-SVM didasarkan pada landasan teori yang menunjukkan bahwa kombinasi ini mampu meningkatkan akurasi klasifikasi, terutama pada citra medis yang memiliki pola visual kompleks dan membutuhkan representasi fitur yang baik.

Selain itu, penelitian ini dirancang sebagai studi terapan dengan melibatkan tahapan perancangan, pelatihan, dan pengujian model menggunakan dataset terbuka citra X-ray dada yang didapat dari situs Kaggle. Untuk menilai seberapa efektif pendekatan yang digunakan, dilakukan evaluasi model menggunakan sejumlah metrik seperti *accuracy*, *precision*, *recall*, dan *F1-score*.

Pendekatan ini sejalan dengan rumusan masalah dan tujuan penelitian yang ingin mengevaluasi efektivitas *hybrid* model CNN-SVM dalam klasifikasi citra X-

ray dada, serta memberikan solusi terhadap rendahnya akurasi klasifikasi pada pendekatan konvensional. Melalui pendekatan eksperimental ini, peneliti dapat mengidentifikasi kombinasi arsitektur model yang optimal dan memberikan dasar ilmiah untuk penerapan teknologi kecerdasan buatan dalam bidang medis, khususnya dalam deteksi pneumonia.

3.2 Teknik Pengumpulan Data

Teknik pengumpulan data dalam penelitian ini dilakukan melalui pengumpulan dataset citra X-ray dada yang mengandung informasi mengenai kondisi paru-paru pasien, baik yang terdiagnosis pneumonia maupun yang sehat atau normal. Data ini bersifat sekunder dan diperoleh dari sumber terbuka yang telah divalidasi serta banyak digunakan dalam penelitian sejenis. Salah satu Penelitian ini memanfaatkan dataset *Chest X-ray Images (Pneumonia)* yang dapat diakses secara terbuka melalui situs Kaggle.



Gambar 3.1 Contoh Citra X-Ray Dada

Sumber: www.Kaggle.com

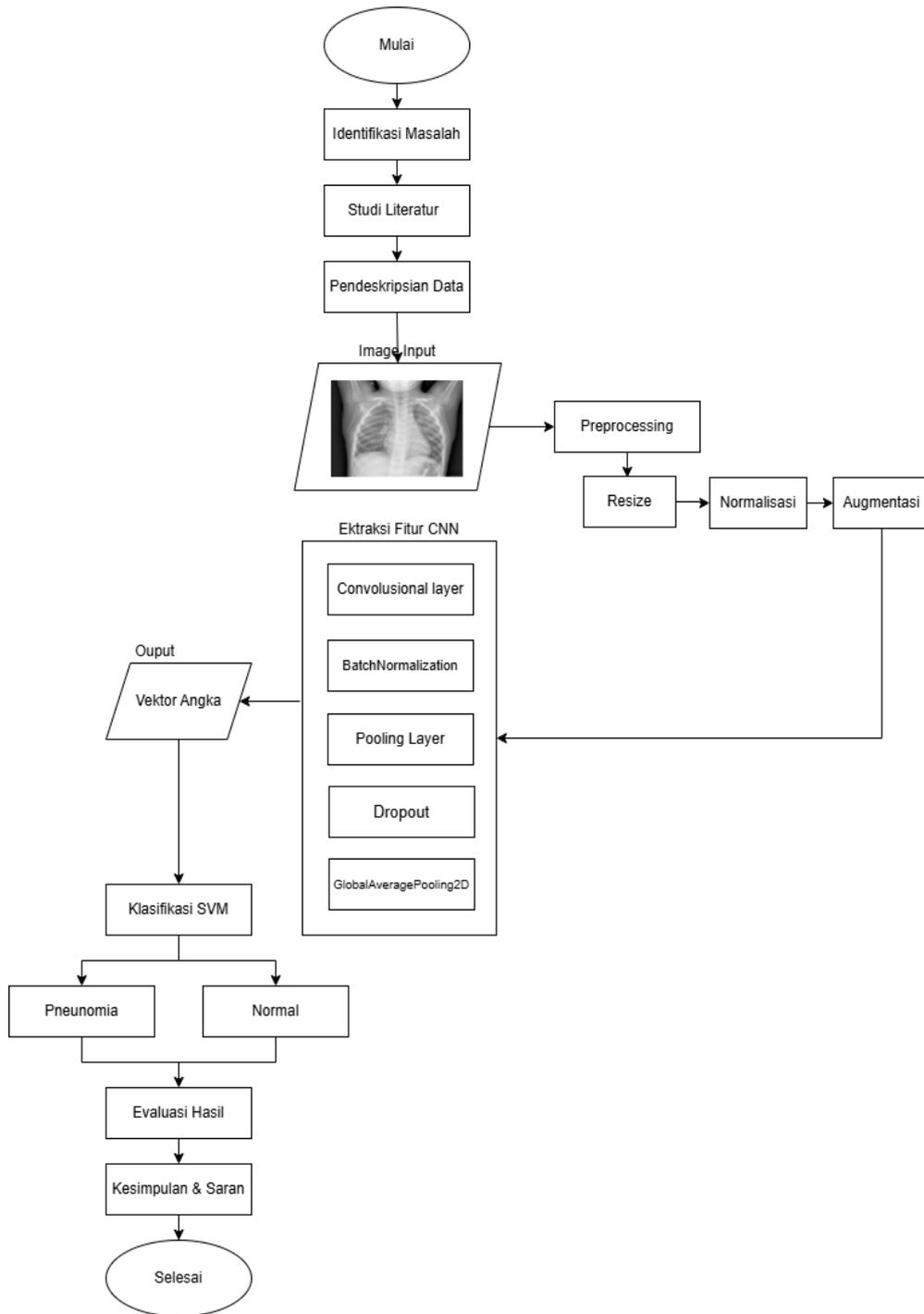
Selain pengumpulan data citra, dilakukan juga pembagian data menjadi data latih (*training data*), data validasi (*validation data*), dan data uji (*testing data*) guna mendukung proses pelatihan dan evaluasi model CNN-SVM secara akurat.

Seluruh data yang digunakan dalam penelitian ini telah melalui proses pra-pemrosesan untuk meningkatkan kualitas citra dan menyesuaikan ukuran input

sesuai kebutuhan model CNN. Proses ini meliputi perubahan resolusi gambar, konversi warna (*grayscale*), normalisasi piksel, dan augmentasi data jika diperlukan.

Teknik pengumpulan data ini dilakukan secara sistematis untuk mendukung pendekatan prediktif yang diimplementasikan dalam penelitian, dengan tujuan agar model yang dibangun mampu mengenali pola pada citra X-ray dada secara optimal dalam mendeteksi pneumonia.

3.3 Tahapan Penelitian



Gambar 3.2 Tahapan Penelitian

Sumber : (Penulis, 2025)

a. Identifikasi Masalah

Pada tahap awal, peneliti mengidentifikasi permasalahan yang menjadi dasar penelitian, yaitu perlunya sistem yang mampu mendeteksi penyakit pneumonia secara otomatis dan akurat berdasarkan citra X-ray dada. Identifikasi ini dilatarbelakangi oleh tingginya angka penderita pneumonia dan keterbatasan tenaga medis dalam menganalisis hasil X-ray secara cepat dan tepat, terutama di wilayah dengan fasilitas kesehatan terbatas.

b. Studi Literatur

Di tahap ini, peneliti menganalisis sejumlah literatur yang berkaitan dengan referensi ilmiah untuk mendalami teori-teori yang relevan, seperti pengolahan citra digital, pneumonia dan anatomi paru-paru, metode CNN dan SVM, serta pendekatan hybrid dalam klasifikasi citra medis. Studi literatur ini menjadi dasar dalam merancang sistem deteksi pneumonia yang sesuai dengan perkembangan teknologi terkini.

c. Pendeskripsian Data

Dataset yang digunakan adalah Chest X-ray dataset yang diperoleh dari Guangzhou Women and Children's Medical Center, berisi 5.863 gambar X-ray yang terbagi ke dalam kategori Pneumonia dan Normal. Dataset telah melewati proses quality control dan validasi oleh tim dokter ahli. Data dibagi dalam tiga folder utama: *train*, *validation*, dan *test*, masing-masing digunakan dalam proses pelatihan dan pengujian model. Tahap ini mencakup penjabaran mengenai dataset yang digunakan dalam penelitian, seperti sumber data, jumlah dan jenis citra X-ray (normal dan pneumonia), serta format data. Selain itu, dilakukan pula eksplorasi terhadap karakteristik data, termasuk

ukuran gambar, format penyimpanan, dan distribusi kelas, guna memastikan kelayakan data dalam proses pelatihan model.

d. Preprocessing

Tahap *preprocessing* dilakukan untuk mempersiapkan citra X-ray dada sebelum digunakan dalam pelatihan model. Proses ini mencakup beberapa langkah penting, yaitu *resize*, normalisasi, dan augmentasi. *Resize* dilakukan untuk menyeragamkan ukuran seluruh citra menjadi 224x224 piksel agar sesuai dengan kebutuhan input arsitektur CNN, sehingga model dapat memproses data secara konsisten. Selanjutnya, dilakukan proses normalisasi dengan mengubah nilai piksel pada citra ke dalam rentang antara 0 hingga 1. Langkah ini bertujuan untuk menstabilkan distribusi data dan mempercepat proses pelatihan. Agar model tidak mengalami *overfitting* dan mampu mengenali pola dari berbagai kondisi citra, dilakukan augmentasi data secara *artifisial*. Teknik augmentasi yang digunakan meliputi rotasi, *flipping* horizontal, *zoom*, serta *shifting*. Dengan tahapan *preprocessing* ini, citra X-ray menjadi lebih siap untuk diproses pada tahap ekstraksi fitur menggunakan CNN dan tahap klasifikasi dengan SVM.

d. Ekstraksi Fitur CNN

Setelah data dipersiapkan, citra X-ray diproses melalui arsitektur *Convolutional Neural Network* (CNN) untuk melakukan ekstraksi fitur. CNN digunakan untuk mengekstraksi pola visual penting dari citra, seperti struktur paru-paru dan indikasi pneumonia.

Proses ekstraksi fitur digunakan untuk menemukan properti pemetaan fitur asli ke fitur baru yang membedakannya dari objek lain. Arsitektur metode

feature extraction layer terdiri dari beberapa lapisan. Misalnya, Lapisan Ekstraksi Fitur menggunakan *Convolutional Layer*, *Pooling Layer*, *Dropout Layer*, dan *GlobalAveragePooling2D* untuk mengekstrak fitur dari gambar, dan Lapisan Klasifikasi biasanya menggunakan *Fully Connected Layer*.

Umumnya, proses klasifikasi dilakukan menggunakan lapisan *fully-connected*, namun dalam penelitian ini digunakan pendekatan alternatif dengan memanfaatkan metode SVM sebagai pengklasifikasi. Beberapa arsitektur CNN populer seperti AlexNet, GoogleNet, dan VGGNet sering dimanfaatkan dalam proses ekstraksi ciri. Akan tetapi, dalam penelitian ini, proses ekstraksi fitur dilakukan menggunakan arsitektur CNN yang terdiri atas tiga jenis lapisan, yaitu *Convolutional Layer*, *Max Pooling Layer*, dan *Dropout Layer*.

e. Klasifikasi SVM

Fitur yang telah diekstraksi oleh CNN kemudian digunakan sebagai input pada algoritma klasifikasi *Support Vector Machine* (SVM). SVM digunakan untuk membedakan antara citra yang menunjukkan indikasi pneumonia dan citra normal. Algoritma ini dipilih karena kemampuannya dalam bekerja pada data berdimensi tinggi dan menghasilkan keputusan klasifikasi yang optimal berdasarkan margin maksimum. Sebelum proses klasifikasi, perlu dilakukan analisis parameter, termasuk yang berikut ini.

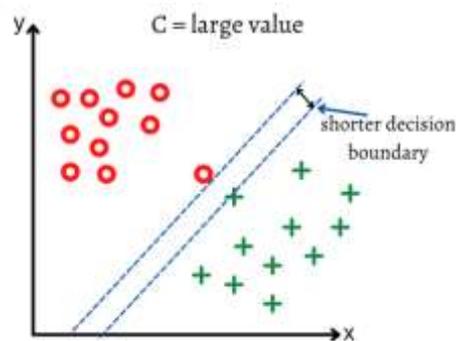
1) Kernel

Kernel digunakan untuk menentukan supervektor, sehingga pemahaman tentang fungsi non-linier tidaklah diperlukan. Beberapa fungsi kernel digunakan dalam studi ini. *Radial Basic Function* (RBF), *Polynomial*,

dan *Sigmoid*.

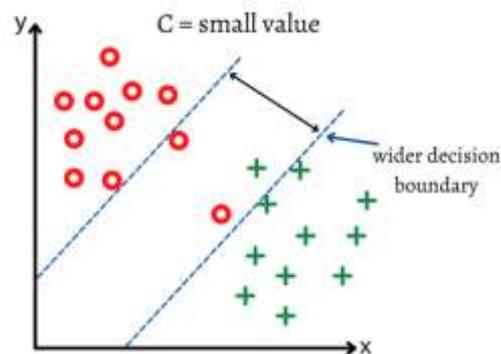
2) Cost (C)

Parameter *cost* adalah metrik yang digunakan dalam optimasi SVM untuk mengidentifikasi kesalahan klasifikasi dalam setiap sampel set data pelatihan. Jika nilai *cost* meningkat, *decision boundary* dalam proses klasifikasi juga akan menurun. Selain itu, seiring dengan meningkatnya presisi klasifikasi, *decision boundary* dalam proses klasifikasi menjadi lebih signifikan. Pemanfaatan *cost* dapat dilihat pada Gambar 3.2 dan 3.3.



Gambar 3.3 Nilai Cost Tinggi

Sumber : www.trivusi.web.id

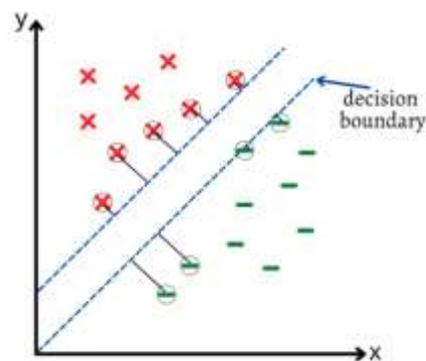


Gambar 3.4 Nilai Cost Rendah

Sumber : www.trivusi.web.id

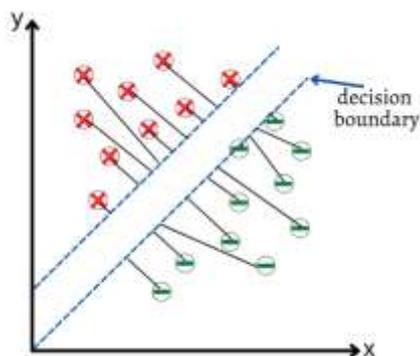
3) Gamma

Parameter gamma adalah ukuran yang menunjukkan beberapa temuan kunci dari set data pelatihan. Jika parameter gamma diatur ke nilai tinggi, titik-titik yang ada di sekitar garis batas keputusan akan terpengaruh dalam perhitungan. Selain itu, parameter gamma rendah akan digunakan untuk menentukan perhitungan titik yang terletak di ujung garis batas keputusan. Penggunaan gamma dapat dilihat pada Gambar 3.4 dan 3.5.



Gambar 3.5 Nilai Gamma Tinggi

Sumber : www.trivusi.web.id



Gambar 3.6 Nilai Gamma Rendah

Sumber : www.trivusi.web.id

4) Degree

Parameter *degree* (d) digunakan untuk menemukan nilai terbaik dalam

dataset.

f. Evaluasi Hasil

Setelah proses klasifikasi dengan kombinasi parameter yang telah dilakukan, tahap selanjutnya adalah melakukan evaluasi hasil terhadap data *test*. Tujuan dari evaluasi ini adalah untuk mengukur kualitas dan tingkat akurasi yang diperoleh dari metode yang digunakan. Dalam mengevaluasi metode, *deep learning* sering menunjukkan variasi performa meskipun arsitektur dan proses *training*-nya serupa. Pada penelitian ini, evaluasi dilakukan menggunakan *confusion matrix*, dengan mengacu pada nilai akurasi, *precision*, *recall*, dan *F1-score*.

g. Kesimpulan dan Saran

Tahap akhir dalam penelitian ini adalah menarik kesimpulan berdasarkan hasil evaluasi yang diperoleh. Selain itu, peneliti juga memberikan saran untuk pengembangan sistem di masa mendatang, baik dari sisi metode, data, maupun penerapannya dalam dunia medis.

BAB IV

HASIL DAN PEMBAHASAN

4.1 Gambaran Umum Dataset

Dataset yang digunakan dalam penelitian ini adalah Chest X-ray Images (Pneumonia) yang diperoleh dari Guangzhou Women and Children's Medical Center, dan tersedia secara publik melalui *platform* Kaggle. Dataset ini terdiri dari 5.863 citra X-ray dada, yang telah melalui proses validasi oleh dua hingga tiga dokter ahli untuk memastikan keakuratan label. Dataset dibagi ke dalam tiga subset, yaitu:

- 1) *Training set* yang digunakan untuk melatih model, terdiri dari 5.216 gambar yang dimana terdapat 1.341 normal dan 3.875 pneumonia.
- 2) *Validation set* yang digunakan untuk validasi selama pelatihan, terdiri dari 16 gambar yang dimana terdapat 8 normal dan 8 pneumonia.
- 3) *Testing set* yang digunakan untuk evaluasi hasil akhir model, terdiri dari 624 gambar yang dimana terdapat 234 normal dan 390 pneumonia.

Setiap gambar dikategorikan ke dalam dua kelas, yaitu Normal dan Pneumonia, yang direpresentasikan dalam format citra berformat JPEG. Dataset ini memiliki karakteristik variasi intensitas piksel, posisi paru-paru, serta kontras yang berbeda-beda, yang menjadi tantangan sekaligus potensi dalam proses pelatihan model deteksi berbasis citra.

Seluruh data digunakan dalam penelitian ini melalui proses *preprocessing*, termasuk *resize* ke dimensi 224x224 piksel, konversi warna menjadi RGB, normalisasi piksel ke rentang [0, 1], serta augmentasi data pada subset *training* untuk meningkatkan generalisasi model. Dengan struktur dan kualitas dataset

yang memadai, data ini sangat sesuai untuk digunakan dalam pelatihan dan evaluasi model hybrid CNN-SVM dalam mendeteksi pneumonia berdasarkan citra X-ray dada.

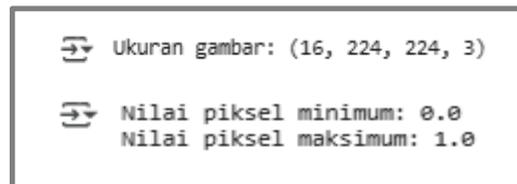
4.2 Hasil *Preprocessing* Data

Pada tahap awal, dilakukan proses *preprocessing* terhadap citra X-ray dada sebelum digunakan dalam pelatihan model. Tujuan dari *preprocessing* ini adalah untuk menyiapkan data agar sesuai dengan kebutuhan arsitektur *Convolutional Neural Network* (CNN), meningkatkan kualitas citra, serta membantu model dalam mengenali pola secara lebih akurat. Langkah-langkah *preprocessing* yang dilakukan antara lain:

- 1) *Rezise* gambar dimana seluruh citra diubah ukurannya menjadi 224×224 piksel agar konsisten dan sesuai dengan dimensi input CNN.
- 2) Normalisasi piksel dimana seluruh nilai piksel pada citra diubah ke dalam rentang $[0, 1]$ dengan cara membaginya dengan angka 255 untuk proses normalisasi. Hal ini bertujuan untuk mempercepat proses pelatihan dan menstabilkan nilai input.
- 3) Augmentasi data dimana augmentasi hanya diterapkan pada data latih (*training*) guna memperbaiki kualitas data serta meminimalkan kemungkinan terjadinya *overfitting*, dilakukan teknik augmentasi data seperti rotasi citra pergeseran horizontal dan vertikal, zoom, serta *flipping horizontal*.
- 4) Konversi warna dimana seluruh citra dikonversi menjadi RGB untuk mempertahankan tiga saluran warna yang diperlukan oleh model CNN.
- 5) *Batching* dan *shuffle* dimana data diorganisasi dalam bentuk *batch* dan

diacak (shuffled) untuk meningkatkan efisiensi pelatihan dan menghindari urutan data yang dapat mempengaruhi proses pembelajaran model.

Berikut adalah hasil pemeriksaan bentuk data setelah dilakukan *preprocessing*

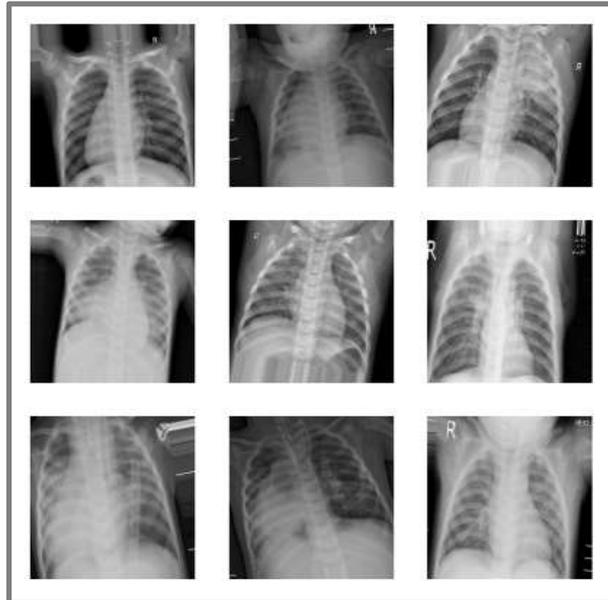


Gambar 4.1 Hasil Resize dan Normalisasi

Sumber : (Penulis, 2025)

Berdasarkan gambar 4.1 hasil ini menunjukkan bahwa satu *batch* data pelatihan terdiri atas 16 gambar (*batch size* = 16), masing-masing dengan resolusi 224×224 piksel dan memiliki 3 channel warna (RGB), meskipun data aslinya berupa citra grayscale. Hal ini terjadi karena proses augmentasi atau konversi otomatis oleh ImageDataGenerator, yang mengubah format citra menjadi tiga channel agar kompatibel dengan arsitektur model CNN seperti yang digunakan dalam penelitian ini.

Nilai piksel yang berada pada rentang 0.0 hingga 1.0 menunjukkan bahwa data citra telah melalui proses normalisasi, yaitu dengan membagi nilai piksel asli (0–255) menjadi skala 0 hingga 1. Proses normalisasi ini penting dilakukan agar mempercepat konvergensi selama pelatihan model dan mencegah dominasi nilai piksel yang besar terhadap fungsi *loss* dan pembaruan bobot pada jaringan saraf. Setelah di resize dan normalisasi dilakukan juga proses augmentasi pada citra



Gambar 4.2 Hasil Augmentasi

Sumber : (Penulis, 2025)

Gambar 4.2 menunjukkan beberapa contoh citra hasil dari proses augmentasi yang diterapkan pada dataset pelatihan. Augmentasi dilakukan untuk meningkatkan keragaman data dan membantu model dalam mengenali pola yang lebih general sehingga dapat mengurangi risiko *overfitting*.

Beberapa teknik augmentasi yang diterapkan meliputi rotasi acak, translasi (pergeseran), zoom, dan *horizontal flipping*. Hal ini dapat diamati dari variasi posisi dan orientasi dada dalam gambar-gambar tersebut. Meskipun berasal dari dataset yang sama, setiap citra hasil augmentasi memiliki sedikit perbedaan dari bentuk aslinya, namun tetap mempertahankan struktur utama organ paru-paru.

4.3 Arsitektur dan Hasil Ekstraksi Fitur CNN

Arsitektur *Convolutional Neural Network* (CNN) yang digunakan dalam penelitian ini dirancang sebagai *feature extractor*. Model ini dibangun untuk mengekstraksi fitur dari citra X-ray dada dan menghasilkan representasi numerik yang berguna untuk proses klasifikasi menggunakan *Support Vector Machine*

(SVM).

Model CNN menerima input berupa citra berwarna (RGB) dengan ukuran 224x224 piksel. Arsitektur terdiri dari lima blok konvolusi, dan setiap blok terdiri dari beberapa layer: *Conv2D*, *BatchNormalization*, *MaxPooling2D*, dan *Dropout*. Layer *Conv2D* bertugas mengekstraksi fitur spasial dari citra, sedangkan *BatchNormalization* berfungsi untuk menormalkan output layer sebelumnya sehingga proses pelatihan menjadi lebih stabil dan cepat. *MaxPooling2D* digunakan untuk melakukan *downsampling* guna mengurangi dimensi spasial fitur, dan *Dropout* digunakan sebagai teknik regularisasi untuk mencegah *overfitting* dengan menilai sejumlah besar neuron secara akurat selama pelatihan..

Setelah melewati lima blok konvolusi, hasil akhir diproses menggunakan layer *GlobalAveragePooling2D*. Layer ini berfungsi untuk mereduksi fitur dari dimensi tiga menjadi vektor satu dimensi dengan cara menghitung rata-rata dari setiap filter. Hasil dari layer ini adalah vektor fitur berdimensi 512, yang kemudian digunakan sebagai input ke dalam algoritma SVM untuk klasifikasi antara kondisi paru-paru normal dan pneumonia.

Adapun detail dari arsitektur CNN yang digunakan ditunjukkan pada tabel 4.1 berikut:

Tabel 4.1 Arsitektur CNN

No	Layer (Tipe)	Output Shape	Jumlah Parameter
1	Input Layer	(None, 224, 224, 3)	0
2	Conv2D	(None, 224, 224, 32)	896
3	BatchNormalization	(None, 224, 224, 32)	128
4	MaxPooling2D	(None, 224, 224, 32)	0

5	Dropout	(None, 112, 112, 32)	0
6	Conv2D_2	(None, 112, 112, 64)	18496
7	BatchNormalization_2	(None, 112, 112, 64)	256
8	MaxPooling2D_2	(None, 56, 56, 64)	0
9	Dropout_2	(None, 56, 56, 64)	0
10	Conv2D_3	(None, 56, 56, 128)	73856
11	BatchNormalization_3	(None, 56, 56, 128)	512
12	MaxPooling2D_3	(None, 28, 28, 128)	0
13	Dropout_3	(None, 28, 28, 128)	0
14	Conv2D_4	(None, 28, 28, 256)	295168
15	BatchNormalization_4	(None, 28, 28, 256)	1024
16	MaxPooling2D_4	(None, 14, 14, 256)	0
17	Dropout_4	(None, 14, 14, 256)	0
18	Conv2D_5	(None, 14, 14, 512)	1180160
19	BatchNormalization_5	(None, 14, 14, 512)	2048
20	MaxPooling2D_5	(None, 7, 7, 512)	0
21	Dropout_5	(None, 7, 7, 512)	0
22	GlobalAveragePooling2D	(None, 512)	0
	Total Parameter		1572544

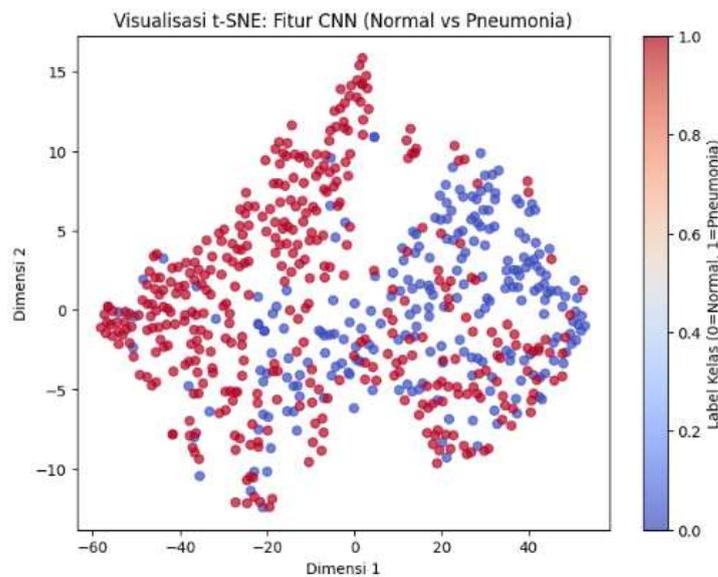
Sumber : (Penulis, 2025)

Setelah lima blok konvolusi, digunakan *GlobalAveragePooling2D*, yang mengubah matriks fitur menjadi vektor fitur berdimensi 512. Layer ini mengambil rata-rata dari setiap fitur map dan cocok digunakan sebagai input bagi model

klasifikasi lain, seperti *Support Vector Machine* (SVM) dalam penelitian ini.

Dari tabel tersebut, diketahui bahwa total parameter dalam arsitektur ini berjumlah 1.572.544, dengan 1.570.560 parameter dapat dilatih (*trainable*) dan 1.984 parameter tidak dapat dilatih (*non-trainable*), yang sebagian besar berasal dari layer normalisasi.

Arsitektur ini dirancang untuk menyeimbangkan antara kedalaman model dan kemampuan generalisasi, serta mampu menangani kompleksitas pola citra X-ray medis yang sering kali memiliki karakteristik visual yang halus dan tidak eksplisit. Model CNN ini telah menunjukkan kemampuan yang baik dalam mengekstraksi fitur yang merepresentasikan kondisi paru-paru secara signifikan.



Gambar 4.3 Visualisasi Fitur Ekstraksi CNN Menggunakan t-SNE

Sumber : (Penulis, 2025)

Gambar 4.3 merupakan hasil visualisasi dua dimensi dari fitur-fitur yang telah diekstraksi oleh model CNN menggunakan teknik *t-distributed Stochastic Neighbor Embedding* (t-SNE). Visualisasi ini bertujuan untuk menggambarkan sejauh mana fitur-fitur yang diekstraksi dari citra X-ray dada mampu

membedakan antara dua kelas, yaitu Normal (biru) dan Pneumonia (merah).

Setiap titik pada grafik mewakili satu citra dari data uji yang telah diproyeksikan ke dalam ruang berdimensi dua. Warna biru merepresentasikan kelas Normal, sedangkan warna merah merepresentasikan kelas Pneumonia. Hasil ini menunjukkan bahwa model CNN mampu menghasilkan representasi fitur yang dapat memisahkan antara dua kelas tersebut dengan cukup baik. Meskipun terdapat beberapa tumpang tindih (*overlap*) antar kelas, sebagian besar titik dari masing-masing kelas membentuk kelompok (*cluster*) yang terpisah, menandakan bahwa fitur yang diekstraksi oleh CNN mengandung informasi diskriminatif yang relevan.

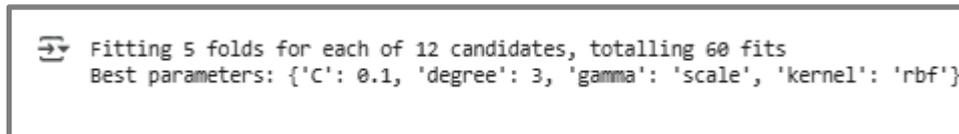
4.4 Pelatihan dan Optimasi Model SVM

Setelah proses ekstraksi fitur dilakukan menggunakan arsitektur CNN, langkah selanjutnya adalah melakukan pelatihan dan optimasi model klasifikasi menggunakan *Support Vector Machine* (SVM). Tujuan dari tahap ini adalah membangun model klasifikasi yang mampu membedakan antara citra X-ray dada yang menunjukkan kondisi normal dan pneumonia berdasarkan fitur yang telah diekstrak.

Fitur dari dataset pelatihan (*train*) dan validasi (*val*) digabungkan dan kemudian dinormalisasi menggunakan metode *StandardScaler*. Proses normalisasi bertujuan untuk menyetarakan skala dari setiap fitur agar algoritma SVM dapat bekerja secara optimal. Normalisasi dilakukan dengan menghitung mean dan standar deviasi dari data pelatihan, kemudian transformasi ini diterapkan ke seluruh data pelatihan, validasi, dan pengujian.

Untuk mendapatkan performa model terbaik, dilakukan proses *Grid Search* terhadap beberapa kombinasi parameter SVM. Parameter yang diuji meliputi:

- 1) C : parameter regularisasi yang diuji dengan nilai [0.01, 0.1].
- 2) Kernel : jenis kernel yang digunakan, yaitu 'rbf', 'poly', dan 'sigmoid'.
- 3) gamma : nilai gamma yang digunakan adalah 'scale'.
- 4) degree : hanya berlaku untuk kernel 'poly', diuji dengan nilai [3, 5].



```

⇒ Fitting 5 folds for each of 12 candidates, totalling 60 fits
Best parameters: {'C': 0.1, 'degree': 3, 'gamma': 'scale', 'kernel': 'rbf'}

```

Gambar 4.4 Hasil Parameter Pada Proses Grid Search

Sumber : (Penulis, 2025)

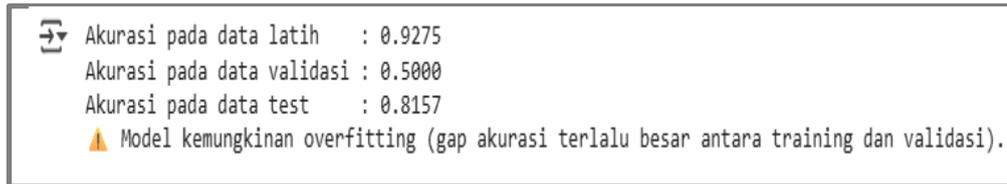
Proses Grid Search menghasilkan total 60 kombinasi pelatihan, yang diperoleh dari 12 kombinasi parameter (C, kernel, gamma, dan degree) yang diuji masing-masing melalui 5-fold cross-validation. Hal ini berarti setiap kombinasi parameter dievaluasi sebanyak lima kali dengan data pelatihan yang berbeda-beda untuk memperoleh hasil yang lebih reliabel dan mencegah *overfitting* terhadap subset tertentu.

Dari seluruh kombinasi tersebut, Grid Search menemukan bahwa performa terbaik diperoleh pada parameter C : 0.1 , kernel : 'rbf', gamma : 'scale', dan degree : 3 walaupun nilai degree tidak digunakan pada kernel rbf. Kombinasi parameter ini memberikan performa tertinggi berdasarkan rata-rata akurasi selama validasi silang, dan kemudian digunakan sebagai model final untuk tahap pengujian dan evaluasi akhir. Pemilihan kernel rbf menunjukkan bahwa hubungan antara fitur-fitur hasil ekstraksi CNN dengan label kelas bersifat non-linear, sehingga membutuhkan pemisahan dengan *hyperplane* non-linier untuk

mendapatkan akurasi yang optimal.

4.5 Evaluasi Performa Model

Setelah proses pelatihan dan optimasi model selesai dilakukan, langkah selanjutnya adalah mengevaluasi performa model dengan menggunakan data uji (*test set*).



```

⇒ Akurasi pada data latih      : 0.9275
   Akurasi pada data validasi   : 0.5000
   Akurasi pada data test      : 0.8157
   ⚠ Model kemungkinan overfitting (gap akurasi terlalu besar antara training dan validasi).

```

Gambar 4.5 Hasil Evaluasi Nilai Akurasi

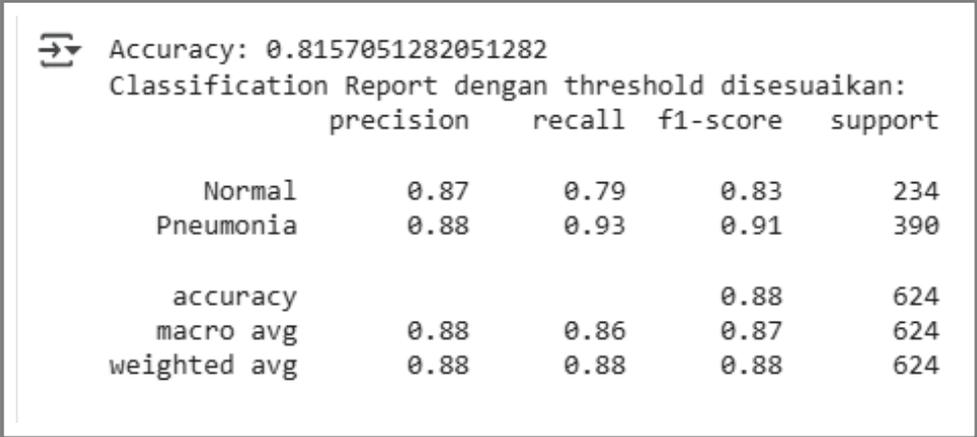
Sumber : (Penulis, 2025)

Berdasarkan hasil tersebut, terlihat adanya perbedaan yang cukup signifikan antara akurasi data latih dan data validasi. Secara sekilas, hal ini dapat menimbulkan indikasi bahwa model tersebut menderita *overfitting*, yaitu suatu kondisi di mana model terus-menerus beradaptasi dengan data sehingga tidak mampu melakukan generalisasi yang tepat terhadap data baru. Namun, dalam konteks penelitian ini, penurunan akurasi yang drastis pada data validasi tidak sepenuhnya mencerminkan *overfitting*, melainkan disebabkan oleh jumlah data validasi yang sangat kecil, yaitu hanya terdiri dari 16 gambar (8 gambar kelas “Normal” dan 8 gambar kelas “Pneumonia”).

Jumlah data validasi yang sangat terbatas menyebabkan evaluasi pada data tersebut tidak cukup representatif untuk mengukur performa model secara objektif. Dalam statistik, ukuran sampel yang terlalu kecil dapat menghasilkan estimasi performa yang bias dan tidak stabil. Oleh karena itu, penurunan akurasi pada data validasi ini bukan disebabkan oleh buruknya kemampuan generalisasi model, melainkan lebih kepada keterbatasan kuantitas data validasi yang tersedia.

Sebaliknya, model menunjukkan performa yang cukup baik pada data uji, dengan akurasi mencapai 81,57%. Data uji merupakan data yang benar-benar tidak dilibatkan dalam proses pelatihan maupun validasi, sehingga hasil akurasi pada data uji menjadi indikator utama untuk menilai kemampuan generalisasi model secara obyektif. Akurasi yang stabil dan tinggi pada data uji menunjukkan bahwa model masih mampu melakukan prediksi dengan baik terhadap data baru, sehingga dapat disimpulkan bahwa model tidak mengalami *overfitting* secara signifikan.

Untuk meningkatkan performa dalam mendeteksi pneumonia, terutama dalam mengurangi *false negative* (kasus pneumonia yang salah terdeteksi sebagai normal), dilakukan penyesuaian *threshold* pada skor prediksi SVM. *Threshold* disesuaikan ke nilai 0.6 agar model lebih sensitif terhadap kemungkinan adanya pneumonia. Setelah penyesuaian *threshold*, diperoleh hasil evaluasi sebagai berikut:



```

Accuracy: 0.8157051282051282
Classification Report dengan threshold disesuaikan:
      precision    recall  f1-score   support

 Normal         0.87      0.79      0.83       234
 Pneumonia      0.88      0.93      0.91       390

 accuracy              0.88       624
 macro avg           0.88      0.86      0.87       624
 weighted avg        0.88      0.88      0.88       624

```

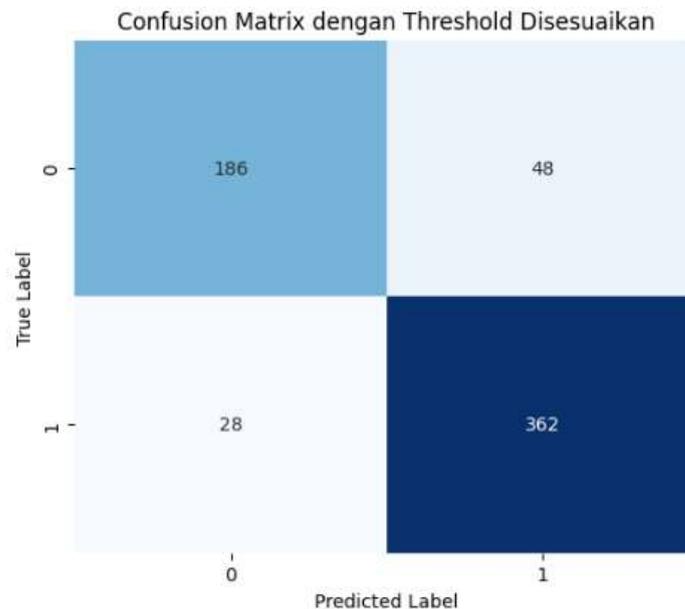
Gambar 4.6 Hasil Evaluasi Dengan Threshold

Sumber : (Penulis, 2025)

Model menunjukkan *recall* yang sangat tinggi untuk kelas Pneumonia, yaitu sebesar 93%, yang berarti model mampu mendeteksi hampir semua kasus

pneumonia secara benar. Hal ini penting dalam konteks medis, karena kesalahan dalam mendeteksi pasien yang sebenarnya terinfeksi pneumonia (*false negative*) dapat berakibat fatal.

Namun, peningkatan sensitivitas ini diiringi dengan sedikit penurunan *precision* pada kelas Normal menjadi 87%, karena model cenderung sedikit "*over-detect*" kasus pneumonia, menyebabkan beberapa gambar normal diklasifikasikan sebagai pneumonia. Strategi ini dinilai tepat karena lebih baik menghasilkan *false positive* dibandingkan *false negative* dalam aplikasi medis. Hasil tersebut juga didukung oleh *confusion matrix* berikut:



Gambar 4.7 Visualisasi Heatmap Hasil Evaluasi

Sumber : (Penulis, 2025)

Berdasarkan *confusion matrix*, terdapat 48 citra normal yang salah diklasifikasikan sebagai pneumonia, dan hanya 28 citra pneumonia yang salah diklasifikasikan sebagai normal. Hal ini menunjukkan bahwa model sangat efektif dalam mendeteksi pneumonia, meskipun terdapat beberapa kesalahan dalam mendeteksi citra normal.

Secara keseluruhan, model hybrid CNN-SVM yang dikembangkan dalam penelitian ini memiliki performa yang baik dalam klasifikasi citra X-ray dada. Dengan akurasi tinggi, *recall* yang sangat baik untuk kelas pneumonia, dan tingkat *overfitting* yang rendah, model ini dapat diandalkan dalam membantu proses deteksi dini penyakit pneumonia berbasis citra medis.

4.6 Penyesuaian Threshold dan Analisis

Setelah Pada tahap ini, dilakukan penyesuaian *threshold* terhadap hasil prediksi dari model *Support Vector Machine* (SVM). Secara default, SVM menggunakan *threshold* 0 untuk memisahkan dua kelas, namun dalam konteks medis, kesalahan dalam mendeteksi penyakit (*false negative*) dapat menimbulkan risiko besar bagi pasien. Oleh karena itu, dilakukan penyesuaian *threshold* untuk meningkatkan sensitivitas model terhadap deteksi pneumonia.

Penyesuaian *threshold* dilakukan dengan menaikkan nilai ambang keputusan (*decision threshold*) dari 0 ke 0.6, sehingga model lebih toleran dalam mengklasifikasikan suatu gambar sebagai pneumonia. Hasil dari penyesuaian *threshold* ini menunjukkan bahwa model menjadi lebih sensitif terhadap gejala pneumonia, ditandai dengan meningkatnya nilai *recall* pada kelas pneumonia dari sebelumnya menjadi 93%.

Namun, perubahan ini menyebabkan peningkatan jumlah *false positive*, yaitu kasus gambar X-ray yang sebenarnya normal namun diklasifikasikan sebagai pneumonia. Meskipun demikian, *trade-off* ini dianggap wajar dan dapat diterima, mengingat dalam praktik medis, lebih baik memberikan perhatian atau pemeriksaan lebih lanjut pada kasus yang dicurigai sebagai pneumonia daripada melewatkan kasus yang benar-benar positif. Dari *confusion matrix* setelah

penyesuaian threshold, diperoleh hasil:

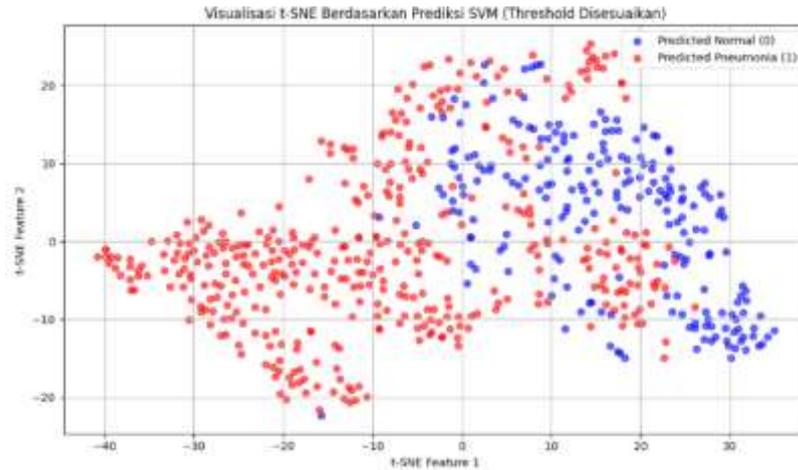
- 1) True Positive (TP): 186
- 2) True Negative (TN): 362
- 3) False Positive (FP): 48
- 4) False Positive (FP): 28

Analisis ini menunjukkan bahwa hanya 28 kasus pneumonia yang tidak terdeteksi, sedangkan sebagian besar kasus pneumonia berhasil dikenali. Sementara itu, kesalahan deteksi pada kasus normal (48 kasus) masih dalam batas yang dapat diterima untuk konteks aplikasi klinis awal.

Dengan demikian, penyesuaian threshold ini berhasil meningkatkan kemampuan model dalam mendeteksi pneumonia secara lebih akurat dan responsif. Langkah ini penting untuk mengoptimalkan kinerja model dalam konteks *screening* penyakit, terutama pada sistem pendukung keputusan medis berbasis citra X-ray dada.

4.7 Visualisasi Prediksi Model

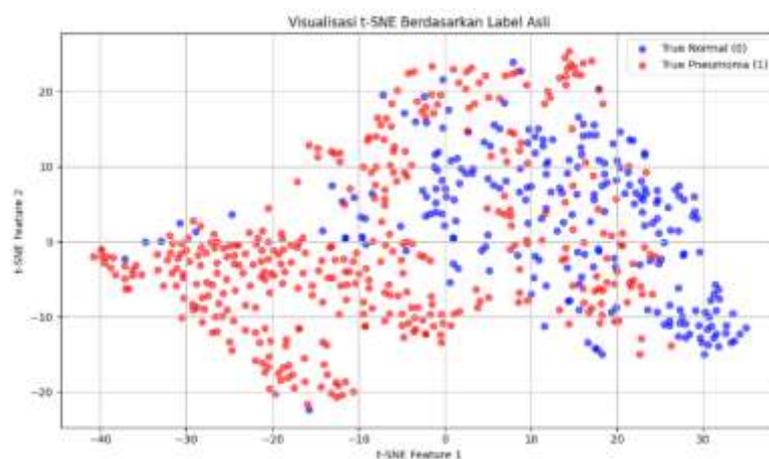
Untuk memberikan pemahaman visual terhadap hasil klasifikasi model SVM setelah proses ekstraksi fitur menggunakan CNN, dilakukan visualisasi dua dimensi menggunakan metode *t-distributed Stochastic Neighbor Embedding* (t-SNE). Visualisasi ini bertujuan untuk merepresentasikan pemisahan antara kelas "Normal" dan "Pneumonia" secara visual setelah fitur diekstraksi dan diproyeksikan ke dalam ruang dua dimensi.



Gambar 4.8 Visualisasi t-SNE Berdasarkan Prediksi SVM (Threshold Disesuaikan)

Sumber : (Penulis, 2025)

Gambar 4.8 menunjukkan visualisasi t-SNE berdasarkan hasil prediksi model SVM dengan threshold yang telah disesuaikan. Terlihat bahwa mayoritas titik telah terklasifikasi ke dalam kelompok yang konsisten, meskipun masih terdapat sejumlah titik yang berada pada area tumpang tindih antar kelas, yang mengindikasikan adanya prediksi yang kurang tepat yang sulit dibedakan secara visual.



Gambar 4.9 Visualisasi t-SNE Berdasarkan Label Asli

Sumber : (Penulis, 2025)

Gambar 4.9 menampilkan visualisasi t-SNE berdasarkan label asli dari data uji. Titik-titik biru menunjukkan citra dengan label asli "Normal", dan titik-titik merah menunjukkan citra dengan label "Pneumonia". Terlihat bahwa penyebaran data dari dua kelas juga cukup jelas terpisah, meskipun tetap terdapat wilayah yang berdekatan yang menunjukkan kemiripan pola fitur di antara sebagian data dari kedua kelas.

Dari perbandingan kedua visualisasi ini, dapat disimpulkan bahwa model SVM telah mampu memetakan fitur-fitur hasil ekstraksi dari CNN ke dalam ruang vektor yang cukup representatif untuk membedakan dua kelas. Keberhasilan pemisahan ini juga selaras dengan hasil evaluasi numerik sebelumnya, seperti akurasi, *precision*, *recall*, dan *F1-score*.

4.8 Implementasi Model ke Aplikasi

Setelah selesainya proses evaluasi dan pelatihan model, langkah selanjutnya adalah memvisualisasikan model yang dikembangkan ke dalam aplikasi berbasis web untuk membantu pengguna dalam melakukan prediksi pneumonia dari X-ray. Implementasi ini bertujuan agar model dapat digunakan oleh pengguna non-teknis secara praktis dan interaktif.

Aplikasi ini dibangun menggunakan pustaka Gradio, yang memungkinkan pembuatan antarmuka pengguna grafis (GUI) secara cepat dan mudah diakses melalui web. Pengguna cukup mengunggah citra X-ray dada, lalu sistem akan secara otomatis melakukan proses prediksi berdasarkan model hybrid CNN-SVM yang telah dilatih sebelumnya.



Gambar 4.10 Interface Model

Sumber : (Penulis, 2025)

Adapun tahapan dalam proses kerja aplikasi ini adalah sebagai berikut:

- 1) Pengguna mengunggah citra X-ray dada ke dalam aplikasi.
- 2) Gambar yang diunggah dikonversi ke format RGB dan diubah ukurannya menjadi 224×224 piksel.
- 3) Model CNN yang telah dilatih akan mengekstraksi fitur dari gambar tersebut.
- 4) Fitur yang dihasilkan dinormalisasi menggunakan StandardScaler yang telah dilatih pada data pelatihan.
- 5) Hasil normalisasi fitur kemudian diklasifikasikan oleh model SVM untuk memprediksi apakah gambar termasuk dalam kelas “Normal” atau “Pneumonia”.
- 6) Hasil prediksi ditampilkan secara langsung pada antarmuka dalam bentuk label yang mudah dipahami.

Dengan adanya implementasi ini, model yang telah dibangun tidak hanya

terbatas pada penggunaan dalam lingkungan pengembangan (eksperimen), tetapi juga dapat diakses secara praktis oleh pengguna akhir sebagai alat bantu awal dalam mendeteksi pneumonia secara otomatis dari citra X-ray dada.

4.9 Pembahasan dan Analisis

Berdasarkan hasil evaluasi performa model yang telah dilakukan, model hybrid CNN-SVM menunjukkan kemampuan yang cukup baik dalam mendeteksi pneumonia berdasarkan citra X-ray dada. Model ini mampu mencapai akurasi sebesar 92,75% pada data latih dan 81,57% pada data uji, yang menunjukkan bahwa model memiliki kemampuan generalisasi yang memadai terhadap data baru.

Meskipun akurasi pada data validasi tercatat hanya 50%, hal ini tidak serta-merta menandakan bahwa model mengalami *overfitting* secara signifikan. Rendahnya akurasi pada data validasi lebih disebabkan oleh jumlah data validasi yang sangat terbatas, yaitu hanya 16 gambar. Jumlah ini tidak cukup untuk memberikan estimasi performa yang stabil dan representatif, sehingga hasil validasi dianggap kurang dapat dijadikan tolak ukur utama. Sebagai gantinya, performa pada data uji dijadikan acuan utama karena mencerminkan kemampuan model terhadap data yang benar-benar belum pernah dilihat sebelumnya.

Proses ekstraksi fitur menggunakan CNN telah berhasil mereduksi dimensi data citra menjadi fitur-fitur numerik yang informatif dan relevan, yang kemudian diklasifikasikan menggunakan algoritma *Support Vector Machine* (SVM). Penggunaan metode GridSearchCV dan Stratified K-Fold Cross Validation dalam pelatihan SVM juga telah membantu dalam menemukan parameter model terbaik yang mampu mengoptimalkan kinerja klasifikasi.

Hasil evaluasi lebih lanjut menggunakan metrik seperti *precision*, *recall*, *f1-score*, dan *confusion matrix* menunjukkan bahwa model memiliki kemampuan deteksi yang tinggi terhadap kelas pneumonia (*recall* mencapai 93%). Hal ini sangat penting dalam konteks deteksi penyakit, karena kesalahan dalam mendeteksi kasus pneumonia (*false negative*) dapat berakibat serius. Oleh karena itu, penyesuaian *threshold* prediksi dilakukan secara manual untuk mengurangi risiko *false negative*, yang terbukti mampu meningkatkan sensitivitas model terhadap kasus pneumonia.

Visualisasi menggunakan t-SNE terhadap hasil prediksi dan label asli juga menunjukkan pemisahan yang cukup jelas antara kelas Normal dan Pneumonia, baik berdasarkan prediksi model maupun berdasarkan label asli. Hal ini memberikan indikasi tambahan bahwa fitur yang diekstraksi oleh CNN cukup efektif dalam memisahkan kedua kelas tersebut di ruang fitur berdimensi rendah.

Namun, selama proses penelitian ini, terdapat beberapa tantangan yang harus dihadapi. Salah satu tantangan utama adalah jumlah data validasi yang sangat terbatas, yakni hanya terdiri dari 16 gambar, yang menyebabkan hasil evaluasi pada data validasi menjadi tidak stabil dan kurang representatif. Hal ini menyulitkan dalam menilai performa model secara obyektif di tahap validasi, sehingga analisis lebih difokuskan pada hasil uji.

Selain itu, integrasi antara model CNN sebagai *feature extractor* dan SVM sebagai *classifier* juga memerlukan penanganan teknis yang kompleks, mulai dari proses normalisasi fitur, penyimpanan antar tahap, hingga pengujian performa akhir. Di sisi lain, keterbatasan perangkat komputasi seperti keterbatasan memori dan waktu sesi pada platform Google Colab menjadi kendala teknis yang cukup

menyita waktu, terutama saat proses pelatihan model dan ekstraksi fitur dalam jumlah besar. Terakhir, implementasi model ke dalam aplikasi menggunakan Gradio juga menuntut konversi pipeline dan penyesuaian input gambar secara *real time*, sehingga memerlukan perhatian ekstra dalam menjaga keakuratan hasil prediksi dan kenyamanan pengguna. Meskipun berbagai tantangan tersebut dihadapi, penelitian ini tetap dapat diselesaikan dengan baik dan menghasilkan model yang berfungsi dengan optimal.

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan maka dapat diambil beberapa kesimpulan sebagai berikut:

- 1) Penelitian ini berhasil mengembangkan sistem deteksi pneumonia berbasis citra X-ray dada menggunakan pendekatan hybrid CNN-SVM, di mana CNN digunakan untuk ekstraksi fitur dan SVM untuk klasifikasi dua kelas, yaitu “Normal” dan “Pneumonia”.
- 2) Model hybrid CNN-SVM menunjukkan performa yang baik dengan akurasi sebesar 92.75% pada data latih dan 81.57% pada data uji. Meskipun akurasi validasi hanya 50%, hal ini disebabkan oleh jumlah data validasi yang sangat terbatas, bukan karena overfitting. Evaluasi lanjutan menunjukkan model memiliki sensitivitas tinggi terhadap kasus pneumonia, dengan recall mencapai 93% dan f1-score sebesar 91% untuk kelas Pneumonia.
- 3) Tantangan dalam penelitian ini meliputi keterbatasan jumlah data validasi, kompleksitas dalam proses ekstraksi fitur dan optimasi parameter, serta integrasi model ke dalam aplikasi berbasis Gradio. Namun, hasil implementasi menunjukkan bahwa model dapat diterapkan secara praktis dalam aplikasi sederhana untuk mendukung diagnosis awal pneumonia.

5.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan mengenai "Pengolahan Citra Digital untuk Ekstraksi Fitur pada Hybrid Model CNN-SVM dalam Deteksi Penyakit Pneumonia Berbasis X-ray Dada", maka dapat diambil saran yaitu untuk pengembangan lebih lanjut, klasifikasi dapat ditingkatkan menjadi multi-kelas, misalnya membedakan antara pneumonia bakteri, jamur dan virus. Selain itu, pengujian pada data X-ray dari rumah sakit atau institusi medis nyata, serta keterlibatan dokter spesialis radiologi, penting untuk menilai kesiapan sistem dalam konteks klinis yang sebenarnya.

Penelitian lanjutan juga dapat mengkaji perbandingan performa antara pendekatan hybrid CNN-SVM dengan model *deep learning end-to-end* seperti CNN + Softmax, atau model *pretrained (transfer learning)* seperti VGG16, ResNet, dan EfficientNet untuk mengetahui efektivitas relatif dari setiap pendekatan. Serta Penambahan jumlah data validasi dan pengujian sangat disarankan agar evaluasi performa model menjadi lebih representatif.

DAFTAR PUSTAKA

- What Is Medical Imaging? | All Allied Health Schools. (n.d.). Retrieved November 15, 2024, from <https://www.allalliedhealthschools.com/blog/what-is-medical-imaging/>
- Rahman, A. K., Purdy, A. G., & Ender, P. T. (2020). COVID-19 pneumonia in patients on chronic hydroxychloroquine therapy: Three cases of COVID-19 pneumonia. *Case Reports in Infectious Diseases*, 2020, Article ID 8822753. <https://doi.org/10.1155/2020/8822753>
- Fida, W. N. R. (2020). *Studi ekologi gambaran faktor pejamu dan faktor lingkungan fisik hunian dengan kejadian pneumonia pada balita di Provinsi Jawa Barat tahun 2014–2017* (Skripsi, Universitas Islam Negeri Syarif Hidayatullah Jakarta). Fakultas Ilmu Kesehatan, Program Studi Kesehatan Masyarakat.
- World Health Organization. (2023). *Pneumonia to kill nearly 11 million children by 2030*. Save the Children. <https://www.savethechildren.net/news/pneumonia-kill-nearly-11-million-children-2030>
- Kementerian Kesehatan Republik Indonesia. (2022). *Pneumonia menjadi ancaman kesakitan dan kematian di dunia*. Direktorat Jenderal Pencegahan dan Pengendalian Penyakit. <https://p2p.kemkes.go.id/pneumonia-menjadi-ancaman-kesakitan-dan-kematian-di-dunia/>
- Mercer Marsh Benefits. (2023). *Health on Demand 2023: Mengungkap Manfaat Kesejahteraan yang Paling Bernilai bagi Karyawan di Indonesia*. Marsh. <https://www.marsh.com/id/en/services/employee-health-benefits/insights/health-on-demand.html>
- Alsharif, R., Al-Issa, Y., Alqudah, A. M., Abu Qasmieh, I., Mustafa, W. A., & Alquran, H. (2021). PneumoniaNet: Automated detection and classification of pediatric pneumonia using chest X-ray images and CNN approach. *Electronics*, 10(23), 2949. <https://doi.org/10.3390/electronics10232949>
- Jawaz, I., & Rahmadewi, R. (2024). Sistem deteksi pneumonia paru-paru dengan pengolahan citra digital dan machine learning. *Electron: Jurnal Ilmiah Teknik Elektro*, 5(1), 138-146. <https://doi.org/10.33019/electron.v5i1.114>
- Rahmawati, A., Yulianti, I., Nurajizah, S., Hidayatulloh, T., & Sari, A. O. (2025). Analisis Performa Model ResNet-50 Pada Diagnosis Pneumonia Balita Berdasarkan Citra Radiografi Thorax. *Computer Science (CO - SCIENCE)*, 5(1), 42-48. <https://doi.org/10.31294/coscience.v5i1.7618>
- Sharma, M., & Parveen, R. (2021). The application of image processing in liver cancer detection. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 12(10). 10.14569/IJACSA.2021.0121050
- Sari, L., & Tampubolon, M. S. (2022). *Deteksi Pneumonia Menggunakan Convolutional Neural Network*. Repository Institut Teknologi Indonesia. <https://repository.itl.ac.id/handle/123456789/1124>
- Rahman, S., Sarker, S., Miraj, M. A. A., Nihal, R. A., Haque, A. K. M. N., &

- Noman, A. A. (2021). Deep learning-driven automated detection of COVID-19 from radiography images: A comparative analysis. *Cognitive Computation*, 13(5), 1–30. <https://doi.org/10.1007/s12559-020-09779-5>
- Khairullah, I. K., Yusa, A., Hartanto, A. D., Hartatik, H., & Kusnawi, K. (2020). *Deteksi citra digital menggunakan algoritma CNN dengan model normalisasi RGB*. *INTECHNO Journal*, 2(2), 1–10. <https://doi.org/10.24076/intechnojournal.2020v2i2.1545>
- Hall, J. E., & Guyton, A. C. (2020). *Guyton and Hall textbook of medical physiology* (14th ed.). Elsevier.
- Zhang, K., Yao, E., Aung, T., & Chuang, P.-T. (2024). *The alveolus: Our current knowledge of how the gas exchange unit of the lung is constructed and repaired*. UCSF Cardiovascular Research Institute. <https://cvri.ucsf.edu/publications/alveolus-our-current-knowledge-how-gas-exchange-unit-lung-constructed-and-repaired>
- Ou, X., Meng, J., Ma, C., Wan, H., Chen, Y., & Luo, F. (2025). *Numerical simulation of voluntary respiration in a model of the whole human airway tree*. *Biomechanics and Modeling in Mechanobiology*. <https://doi.org/10.1007/s10237-025-01932-z>
- Häder, J., Meiners, S., & Mühlfeld, C. (2023). *Imaging of infectious lung diseases*. Springer. https://link.springer.com/chapter/10.1007/174_2024_554
- Chen, B., Liu, W., Chen, Y., She, Q., Li, M., Zhao, H., Zhao, W., Peng, Z., & Wu, J. (2021). Effect of Poor Nutritional Status and Comorbidities on the Occurrence and Outcome of Pneumonia in Elderly Adults. *Frontiers in Medicine*, 8, 719530. <https://doi.org/10.3389/fmed.2021.719530>
- Fauziyah, A. H. (2020). Deteksi Pneumonia pada Anak-Anak dari Citra X-Ray Berbasis Convolutional Neural Network. *Tugas Akhir, Institut Teknologi Sepuluh Nopember Surabaya*. <http://repository.its.ac.id/id/eprint/80480>
- Mishra, A. (2024). A comprehensive review of Artificial Intelligence and Machine Learning: Concepts, trends, and applications. *International Journal of Scientific Research in Science and Technology*, 11(5), 126–142. <https://doi.org/10.32628/IJSRST2411587>
- Harahap, M., Laia, E. M., Sitanggang, L. S., Sinaga, M., Sihombing, D. F., & Husein, A. M. (2022). Deteksi Penyakit Covid-19 Pada Citra X-Ray Dengan Pendekatan Convolutional Neural Network (CNN). *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 6(1), 70–77. <https://doi.org/10.29207/resti.v6i1.3373>
- Mardianto, M. F. F., Yoani, A., Soewignjo, S., Putra, I. K. P. K. A., & Dewi, D. A. (2024). Classification of Pneumonia from Chest X-ray images using Support Vector Machine and Convolutional Neural Network. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 15(6). <https://doi.org/10.14569/IJACSA.2024.01506104>
- Zhang, D., Ren, F., Li, Y., Na, L., & Ma, Y. (2021). Pneumonia Detection from

- Chest X-ray Images Based on Convolutional Neural Network. *Electronics*, 10(13), 1512. <https://doi.org/10.3390/electronics10131512>
- Artiyati, N., & Ahdika, A. (2024). *Implementation of the Support Vector Machine (SVM) Method for Classification of Underdeveloped Areas in Indonesia*. *Jurnal Media Informatika Budidarma*, 8(1), 36–43. <https://doi.org/10.30865/mib.v8i2.7598>
- Putri, A. N., Aryanti, A., & Soim, S. (2024). Implementasi algoritma SVM non-linear pada klasifikasi analisis sentimen perkembangan AI di sektor pendidikan. *Building of Informatics, Technology and Science (BITS)*, 6(2), 703–711. <https://paperity.org/p/356546961/implementasi-algoritma-svm-non-linear-pada-klasifikasi-analisis-sentimen-perkembangan-ai>
- Jollyta, D., Gusrianty, G., Prihandoko, P., & Sukrianto, D. (2023). N-gram and kernel performance using Support Vector Machine algorithm for fake news detection system. *ILKOM Jurnal Ilmiah*, 15(3), 398–404. <https://doi.org/10.33096/ilkom.v15i3.1770.398-404>
- Alfarizi, M. R. S., Al-Farish, M. Z., Taufiqurrahman, M., Ardiansah, G., & Elgar, M. (2023). Penggunaan Python sebagai bahasa pemrograman untuk machine learning dan deep learning. *Karimah Tauhid*, 2(1). <https://paperity.org/p/342223681/penggunaan-python-sebagai-bahasa-pemrograman-untuk-machine-learning-dan-deep-learning>
- DigitalOcean. (2025). *Best Python libraries for machine learning in 2025*. DigitalOcean. <https://www.digitalocean.com/community/conceptual-articles/python-libraries-for-machine-learning>
- Hermanto, K., Salim, D., Wu, B., Salim, O. R., & Gunadi, R. B. (2023). Penggunaan Python untuk menganalisis pola penyebaran COVID-19 di masa pandemi. *Journal of Student Development Information System (JoSDIS)*, 3(2), 62–75. <https://jurnal.ulb.ac.id/index.php/JoSDIS/article/view/4548>
- J. Mohajon, “Confusion Matrix for Your Multi-Class Machine Learning Model,” <https://towardsdatascience.com/>, 2020. <https://towardsdatascience.com/confusion-matrix-for-your-multi-class-machine-learning-model-ff9aa3bf7826>
- Hidayat, T. (2021). *Penerapan Algoritma Convolutional Neural Network Dalam Identifikasi Citra Daging Merah Konsumsi Jenis Lokal* (Tesis, Universitas Nusa Mandiri). <https://repository.nusamandiri.ac.id/repo/17071/Penerapan-Algoritma-Convolutional-Neural-Network-Dalam-Identifikasi-Citra-Daging-Merah-Konsumsi-Jenis-Lokal>
- Yopento, J., Ernawati, E., & Coastera, F. F. (2021). Identifikasi pneumonia pada citra X-ray menggunakan metode convolutional neural network (CNN) berdasarkan ekstraksi fitur Sobel. *Rekursif: Jurnal Informatika*, 9(2), 45-52. <https://doi.org/10.33369/rekursif.v10i1.17247>
- Khoirunisa, F. E., & Charibaldi, N. (2024). Effect of using GLCM and LBP+HOG feature extraction on SVM method in classification of human skin disease

- type. *Inform: Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi*, 9(2), 145–150. <https://doi.org/10.25139/inform.v9i2.8275>
- Al Reshan, M. S., Gill, K. S., Anand, V., Gupta, S., Alshahrani, H., Sulaiman, A., & Shaikh, A. (2023). Detection of pneumonia from chest X-ray images utilizing MobileNet model. *Healthcare*, 11(11), 1561. <https://doi.org/10.3390/healthcare11111561>
- Mabrouk, A., Díaz Redondo, R. P., Dahou, A., Abd Elaziz, M., & Kayed, M. (2022). Pneumonia Detection on Chest X-ray Images Using Ensemble of Deep Convolutional Neural Networks. *Applied Sciences*, 12(13), 6448. <https://doi.org/10.3390/app12136448>
- Maysanjaya, I. M. D. (2020). *Klasifikasi pneumonia pada citra X-ray paru-paru dengan Convolutional Neural Network*. *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, 9(2), 190–195. <https://doi.org/10.22146/jnteti.v9i2.66>
- Putri, V. K. M. (2022, 11 Oktober). Paru-paru: Pengertian dan bagiannya. *Kompas.com*. <https://www.kompas.com/skola/read/2022/10/11/070000569/paru-paru--pengertian-dan-bagiannya>
- Ahmad, I. (2020). Confusion matrix. Dalam *40 Algorithms Every Programmer Should Know* (hlm. 85–88). Packt Publishing. <https://www.oreilly.com/library/view/40-algorithms-every/9781789801217/c3e888a0-b279-46e9-b16e-2ded81e49e02.xhtml>
- Trivusi. (2022, 4 Juli). Apa itu Kernel Trick? Pengertian dan Jenis-jenis Fungsi Kernel SVM. *Trivusi*. <https://www.trivusi.web.id/2022/04/fungsi-kernel-svm.html>

LAMPIRAN

A. Lampiran Code Program

1) Penetapan Seed

```
import random
import numpy as np
import tensorflow as tf
import os

seed = 42
random.seed(seed)
np.random.seed(seed)
tf.random.set_seed(seed)
os.environ['PYTHONHASHSEED'] = str(seed)
```

2) Preprocessing

```
from tensorflow.keras.preprocessing.image import ImageDataGenerator
from google.colab import drive

drive.mount('/content/drive')

# 1. Direktori dataset
data_dir = '/content/drive/MyDrive/dataset/archive (3)/chest_xray'
train_dir = data_dir + '/train'
val_dir = data_dir + '/val'
test_dir = data_dir + '/test'

batch_size = 16

# 3. Generator dengan augmentasi (untuk training CNN saja)
train_datagen_aug = ImageDataGenerator(
    rescale=1./255,
    rotation_range=15,
    width_shift_range=0.1,
    height_shift_range=0.1,
    zoom_range=0.1,
    horizontal_flip=True
)

train_data_aug = train_datagen_aug.flow_from_directory(
    train_dir,
    target_size=(224, 224),
    batch_size=batch_size,
    class_mode='binary',
    shuffle=True,
    seed=seed
)
```

```

# 4. Generator tanpa augmentasi (untuk ekstraksi fitur CNN ke .npy)
pure_datagen = ImageDataGenerator(rescale=1./255)

train_data_no_aug = pure_datagen.flow_from_directory(
    train_dir,
    target_size=(224, 224),
    batch_size=batch_size,
    class_mode='binary',
    shuffle=False
)

val_data = pure_datagen.flow_from_directory(
    val_dir,
    target_size=(224, 224),
    batch_size=batch_size,
    class_mode='binary',
    shuffle=False
)

test_data = pure_datagen.flow_from_directory(
    test_dir,
    target_size=(224, 224),
    batch_size=batch_size,
    class_mode='binary',
    shuffle=False
)

```

3) Cek Resize, normalisasi, dan augmentasi

```

▶ # Ambil satu batch dari data training
images, labels = next(train_data_aug)

# Lihat bentuk dari gambar (batch_size, height, width, channels)
print("Ukuran gambar:", images.shape)

```

```

▶ print("Nilai piksel minimum:", images.min())
print("Nilai piksel maksimum:", images.max())

```

```

▶ import matplotlib.pyplot as plt

# Tampilkan 9 gambar dari batch
plt.figure(figsize=(10, 10))
for i in range(9):
    plt.subplot(3, 3, i+1)
    plt.imshow(images[i])
    plt.axis('off')
plt.suptitle("Contoh Gambar Hasil Augmentasi")
plt.show()

```

4) Membuat Model Ekstraksi Fitur

```

from tensorflow.keras.models import Model
from tensorflow.keras.layers import Conv2D, MaxPooling2D, Dropout, GlobalAveragePooling2D, Input, BatchNormalization
from tensorflow.keras import regularizers
from tensorflow.keras.initializers import HeNormal

# 1. Input shape untuk gambar 224x224 RGB
input_shape = (224, 224, 3)
inputs = Input(shape=input_shape)

# 2. Convolutional Block 1
x = Conv2D(32, (3, 3), activation='relu', padding='same', kernel_initializer=HeNormal(), kernel_regularizer=regularizers.l2(0.001))(inputs)
x = BatchNormalization()(x)
x = MaxPooling2D(pool_size=(2, 2))(x)
x = Dropout(0.4)(x)

# 3. Convolutional Block 2
x = Conv2D(64, (3, 3), activation='relu', padding='same', kernel_initializer=HeNormal(), kernel_regularizer=regularizers.l2(0.001))(x)
x = BatchNormalization()(x)
x = MaxPooling2D(pool_size=(2, 2))(x)
x = Dropout(0.4)(x)

# 4. Convolutional Block 3
x = Conv2D(128, (3, 3), activation='relu', padding='same', kernel_initializer=HeNormal(), kernel_regularizer=regularizers.l2(0.001))(x)
x = BatchNormalization()(x)
x = MaxPooling2D(pool_size=(2, 2))(x)
x = Dropout(0.4)(x)

# 5. Convolutional Block 4
x = Conv2D(256, (3, 3), activation='relu', padding='same', kernel_initializer=HeNormal(), kernel_regularizer=regularizers.l2(0.001))(x)
x = BatchNormalization()(x)
x = MaxPooling2D(pool_size=(2, 2))(x)
x = Dropout(0.4)(x)

# 6. Convolutional Block 5
x = Conv2D(512, (3, 3), activation='relu', padding='same', kernel_initializer=HeNormal(), kernel_regularizer=regularizers.l2(0.001))(x)
x = BatchNormalization()(x)
x = MaxPooling2D(pool_size=(2, 2))(x)
x = Dropout(0.4)(x)

# 7. Global Average Pooling untuk ekstraksi fitur akhir
x = GlobalAveragePooling2D()(x)

# 8. Buat model CNN sebagai feature extractor
cnn_feature_extractor = Model(inputs=inputs, outputs=x)

# 9. Tampilkan ringkasan model
cnn_feature_extractor.summary()

```

5) Visualisasi t-SNE Fitur CNN

```

from sklearn.manifold import TSNE
import matplotlib.pyplot as plt
import numpy as np

# Ekstrak fitur dari test_data
features = cnn_feature_extractor.predict(test_data)

# Ambil label asli (0 = Normal, 1 = Pneumonia)
labels = test_data.classes

# Ubah ke dimensi 2 menggunakan t-SNE
tsne = TSNE(n_components=2, random_state=seed)
features_2d = tsne.fit_transform(features)

# Visualisasikan
plt.figure(figsize=(8,6))
plt.scatter(features_2d[:,0], features_2d[:,1], c=labels, cmap='coolwarm', alpha=0.7)
plt.title("Visualisasi t-SNE: Fitur CNN (Normal vs Pneumonia)")
plt.xlabel("Dimensi 1")
plt.ylabel("Dimensi 2")
plt.colorbar(label="Label Kelas (0=Normal, 1=Pneumonia)")
plt.show()

```

6) Ekstraksi Fitur dan Simpan ke .Npy

```

import numpy as np
from tqdm import tqdm

# Fungsi untuk ekstraksi fitur dari data generator menggunakan CNN
def extract_features(data_generator, model):
    features = []
    labels = []

    for i in tqdm(range(len(data_generator))):
        imgs, lbls = data_generator[i]
        feats = model.predict(imgs, verbose=0)
        features.append(feats)
        labels.append(lbls)

    return np.vstack(features), np.concatenate(labels)

# Ekstraksi fitur dari train, val, dan test
train_features, train_labels = extract_features(train_data_no_aug, cnn_feature_extractor)
val_features, val_labels = extract_features(val_data, cnn_feature_extractor)
test_features, test_labels = extract_features(test_data, cnn_feature_extractor)

# Simpan ke file .npy
np.save('/content/drive/MyDrive/train_features.npy', train_features)
np.save('/content/drive/MyDrive/train_labels.npy', train_labels)
np.save('/content/drive/MyDrive/val_features.npy', val_features)
np.save('/content/drive/MyDrive/val_labels.npy', val_labels)
np.save('/content/drive/MyDrive/test_features.npy', test_features)
np.save('/content/drive/MyDrive/test_labels.npy', test_labels)

```

7) Training SVM dengan Grid Search

```

from sklearn.svm import SVC
from sklearn.model_selection import StratifiedKFold, GridSearchCV
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import accuracy_score
import numpy as np

# Load data training dan validasi
train_features = np.load('/content/drive/MyDrive/train_features.npy')
train_labels = np.load('/content/drive/MyDrive/train_labels.npy')
val_features = np.load('/content/drive/MyDrive/val_features.npy')
val_labels = np.load('/content/drive/MyDrive/val_labels.npy')

# Gabungkan data train dan val
X = np.vstack([train_features, val_features])
y = np.concatenate([train_labels, val_labels])

# Normalisasi
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

# Setup parameter GridSearch
param_grid = {
    'C': [0.01, 0.1],
    'kernel': ['poly', 'sigmoid', 'rbf'],
    'gamma': ['scale'],
    'degree': [3, 5] # hanya berlaku untuk kernel poly
}

```

```

# Gunakan Stratified K-Fold agar proporsi kelas tetap seimbang
skf = StratifiedKFold(n_splits=5, shuffle=True, random_state=seed)

# Setup GridSearchCV dengan Stratified K-Fold
svm = SVC()
grid = GridSearchCV(estimator=svm, param_grid=param_grid, cv=skf, n_jobs=-1, verbose=2)
grid.fit(X_scaled, y)

# Model terbaik
best_svm = grid.best_estimator_
print("Best parameters:", grid.best_params_)

```

8) Proses Scaler

```

from sklearn.preprocessing import StandardScaler
import numpy as np

# Load fitur
train_features = np.load('/content/drive/MyDrive/train_features.npy')
val_features = np.load('/content/drive/MyDrive/val_features.npy')
test_features = np.load('/content/drive/MyDrive/test_features.npy')

# Scaling
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(train_features)
X_val_scaled = scaler.transform(val_features)
X_test_scaled = scaler.transform(test_features)

```

9) Cek Overfitting

```

from sklearn.metrics import accuracy_score, confusion_matrix, classification_report
import numpy as np

# Load label
y_train = np.load('/content/drive/MyDrive/train_labels.npy')
val_labels = np.load('/content/drive/MyDrive/val_labels.npy')
test_labels = np.load('/content/drive/MyDrive/test_labels.npy')

# Prediksi
train_preds = best_svm.predict(X_train_scaled)
val_preds = best_svm.predict(X_val_scaled)
test_preds = best_svm.predict(X_test_scaled)

# Hitung akurasi
train_acc = accuracy_score(y_train, train_preds)
val_acc = accuracy_score(val_labels, val_preds)
test_acc = accuracy_score(test_labels, test_preds)

# Threshold untuk analisis overfitting
threshold = 0.10 # misal 10% gap toleransi

# Tampilkan hasil akurasi
print(f"Akurasi pada data latih : {train_acc:.4f}")
print(f"Akurasi pada data validasi : {val_acc:.4f}")
print(f"Akurasi pada data test : {test_acc:.4f}")

```

```

# Analisis sederhana
if train_acc - val_acc > threshold:
    print(" ⚠ Model kemungkinan overfitting (gap akurasi terlalu besar antara training dan validasi).")
elif val_acc < test_acc and abs(test_acc - train_acc) < threshold:
    print(" 📊 Model generalisasi cukup baik.")
elif abs(train_acc - val_acc) < threshold and abs(train_acc - test_acc) < threshold:
    print(" ✅ Model tampak seimbang dan tidak overfitting.")
else:
    print(" 🗨 Model overfitting ringan atau data imbalance).")

# Dapatkan skor keputusan SVM
test_scores = best_svm.decision_function(X_test_scaled)

# Set threshold baru untuk kurangi false negative pneumonia (adjust sesuai kebutuhan)
new_threshold = 0.6

# Prediksi ulang berdasarkan threshold manual
test_preds_threshold = (test_scores >= new_threshold).astype(int)

```

10) Evaluasi Model

```

from sklearn.metrics import classification_report, confusion_matrix, accuracy_score
import numpy as np

# Load test set
test_features = np.load('/content/drive/MyDrive/test_features.npy')
test_labels = np.load('/content/drive/MyDrive/test_labels.npy')

# Evaluasi model terbaik
y_pred = best_svm.predict(X_test_scaled)

# Evaluasi performa
print("Accuracy:", accuracy_score(test_labels, y_pred))

print("Classification Report dengan threshold disesuaikan:")
print(classification_report(test_labels, test_preds_threshold, target_names=["Normal", "Pneumonia"]))

print("Confusion Matrix dengan threshold disesuaikan:")
print(confusion_matrix(test_labels, test_preds_threshold))

```

11) Visualisasi Heatmap Confussion Matrix

```

import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.metrics import confusion_matrix

# Hitung confusion matrix berdasarkan prediksi dengan threshold disesuaikan
cm = confusion_matrix(test_labels, test_preds_threshold)

plt.figure(figsize=(6, 5))
sns.heatmap(cm, annot=True, fmt='d', cmap='Blues', cbar=False)
plt.title('Confusion Matrix dengan Threshold Disesuaikan')
plt.xlabel('Predicted Label')
plt.ylabel('True Label')
plt.show()

```

12) Visualisasi t-SNE Setelah Klasifikasi SVM

```

import matplotlib.pyplot as plt
from sklearn.manifold import TSNE
import numpy as np

# Jalankan t-SNE pada fitur test yang sudah diskalakan
tsne = TSNE(n_components=2, perplexity=30, n_iter=1000, random_state=seed)
X_test_tsne = tsne.fit_transform(X_test_scaled)

# Visualisasi berdasarkan prediksi dengan threshold disesuaikan
plt.figure(figsize=(10, 6))
colors = ['blue', 'red']
label_names = ['Normal (0)', 'Pneumonia (1)']

for i, label in enumerate(np.unique(test_preds_threshold)):
    indices = test_preds_threshold == label
    plt.scatter(X_test_tsne[indices, 0], X_test_tsne[indices, 1],
               c=colors[i], label=f'Predicted {label_names[int(label)]}', alpha=0.6)

plt.title("Visualisasi t-SNE Berdasarkan Prediksi SVM (Threshold Disesuaikan)")
plt.xlabel("t-SNE Feature 1")
plt.ylabel("t-SNE Feature 2")
plt.legend()
plt.grid(True)
plt.tight_layout()
plt.show()

```

```

# Opsional: Visualisasi berdasarkan label asli (untuk perbandingan)
plt.figure(figsize=(10, 6))
for i, label in enumerate(np.unique(test_labels)):
    indices = test_labels == label
    plt.scatter(X_test_tsne[indices, 0], X_test_tsne[indices, 1],
               c=colors[i], label=f'True {label_names[int(label)]}', alpha=0.6)

plt.title("Visualisasi t-SNE Berdasarkan Label Asli")
plt.xlabel("t-SNE Feature 1")
plt.ylabel("t-SNE Feature 2")
plt.legend()
plt.grid(True)
plt.tight_layout()
plt.show()

```

13) Menyimpan Model yang Sudah Dilatih ke Google Drive

```

import joblib

# Simpan CNN Feature Extractor
cnn_feature_extractor.save('/content/drive/MyDrive/cnn_feature_extractor.h5')

# Simpan SVM model terbaik dari GridSearch
joblib.dump(best_svm, '/content/drive/MyDrive/svm_model.pkl')

# Simpan Scaler
joblib.dump(scaler, '/content/drive/MyDrive/scaler.pkl')

```

14) Instal Gradio

```
!pip install gradio
import gradio as gr
```

15) Load Model yang Sudah Disimpan

```
from tensorflow.keras.models import load_model
import joblib
import numpy as np
from PIL import Image
import cv2

# Load CNN feature extractor
cnn_model = load_model('/content/drive/MyDrive/cnn_feature_extractor.h5')

# Load SVM model
svm_model = joblib.load('/content/drive/MyDrive/svm_model.pkl')

# Load scaler
scaler = joblib.load('/content/drive/MyDrive/scaler.pkl')
```

16) Membuat Fungsi Prediksi Untuk Interface

```
def predict_pneumonia(img):
    # Konversi dan resize gambar
    img = img.convert("RGB")
    img = img.resize((224, 224))
    img_array = np.array(img) / 255.0 # Normalisasi
    img_array = np.expand_dims(img_array, axis=0) # Tambahkan batch dimensi

    # Ekstraksi fitur dengan CNN
    features = cnn_model.predict(img_array)

    # Normalisasi fitur
    features_scaled = scaler.transform(features)

    # Prediksi dengan SVM
    prediction = svm_model.predict(features_scaled)[0]

    # Label
    label = "Pneumonia" if prediction == 1 else "Normal"
    return label
```

17) Membuat Interface Model

```
interface = gr.Interface(
    fn=predict_pneumonia,
    inputs=gr.Image(type="pil"),
    outputs=gr.Label(num_top_classes=2),
    title="Deteksi Pneumonia dari X-ray Dada",
    description="Upload gambar X-ray dada untuk mendeteksi apakah terkena pneumonia atau tidak"
)

interface.launch()
```

18) SK Pembimbing


MAJELIS PENDIDIKAN TINGGI PENELITIAN & PENGEMBANGAN PIMPINAN PUSAT MUHAMMADIYAH
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UMSU Terakreditasi A Berdasarkan Keputusan Badan Akreditasi Nasional Perguruan Tinggi No. 99/SK/BAN-PT/Akred/PT/02/2019
Pusat Administrasi: Jalan Mukhtar Basri No. 3 Medan 20238 Telp. (061) 6622400 - 66224567 Fax. (061) 6625474 - 6631003
Website: www.umhu.ac.id Email: info@umhu.ac.id

PENETAPAN DOSEN PEMBIMBING
PROPOSAL/SKRIPSI MAHASISWA
NOMOR : 244/IL3-AU/UMSU-09/F/2025

Assalamu'alaikum Warahmatullahi Wabarakatuh

Dekan Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Muhammadiyah Sumatera Utara, berdasarkan Persetujuan permohonan judul penelitian Proposal / Skripsi dari Ketua / Sekretaris

Program Studi : Teknologi Informasi
Pada tanggal : 04 Februari 2025

Dengan ini menetapkan Dosen Pembimbing Proposal / Skripsi Mahasiswa.

Nama : Muhammad Aulia Irza
NPM : 2109020171
Semester : VII (Tujuh)
Program studi : Teknologi Informasi
Judul Proposal / Skripsi : Pengolahan Citra Digital untuk Ekstraksi Fitur pada Hybrid Model CNN-SVM dalam Deteksi Penyakit Pneumonia Berbasis X-Ray Dada.

Dosen Pembimbing : Dr. Irvan, M.Si.

Dengan demikian di izinkan menulis Proposal / Skripsi dengan ketentuan

1. Penulisan berpedoman pada buku panduan penulisan Proposal / Skripsi Fakultas Ilmu Komputer dan Teknologi Informasi UMSU
2. Pelaksanaan Sidang Skripsi harus berjarak 3 bulan setelah dikeluarkannya Surat Penetapan Dosen Pembimbing Skripsi.
3. **Proyek Proposal / Skripsi dinyatakan " BATAL "** bila tidak selesai sebelum Masa Kadaluarsa tanggal - **04 Februari 2026**
4. Revisi judul.....

Wassalamu'alaikum Warahmatullahi Wabarakatuh

Ditetapkan di : Medan
 Pada Tanggal : 05 Sya'ban 1446 H
 04 Februari 2025M


 a.n. Dekan
 Wakil Dekan I
Hujia Maulana, S.T.,M.Kom.
 NIDN : 0121119102







C.c. File

PENGOLAHAN CITRA DIGITAL UNTUK EKSTRAKSI FITUR PADA HYBRID MODEL CNN-SVM DALAM DETEKSI PENYAKIT PNEUMONIA BERBASIS X-RAY DADA

ORIGINALITY REPORT

15% SIMILARITY INDEX	10% INTERNET SOURCES	8% PUBLICATIONS	4% STUDENT PAPERS
--------------------------------	--------------------------------	---------------------------	-----------------------------

PRIMARY SOURCES

1	docplayer.info Internet Source	1%
2	ejournal.unesa.ac.id Internet Source	1%
3	Fuat Turk, Erkan Akkur, Osman Eroglu. "BI-RADS categories and breast lesions classification of mammographic images using artificial intelligence diagnostic model", Neural Network World, 2023 Publication	1%
4	Raihan Raihan, Cecep Nurul Alam, Wildan Budiawan Zulfikar. "Deteksi Pneumonia pada Citra Akhir X - Ray Dada Menggunakan Convolutional Neural Networks Berdasarkan Fitur Prewitt Operator", INTERNAL (Information System Journal), 2025 Publication	1%
5	repository.uin-suska.ac.id Internet Source	1%
6	Ilham Jawaz, Reni Rahmadewi. "Sistem Deteksi Pneumonia Paru-Paru dengan Pengolahan Citra Digital dan Machine Learning", ELECTRON Jurnal Ilmiah Teknik Elektro, 2024 Publication	<1%

7	etheses.uin-malang.ac.id Internet Source	<1 %
8	Submitted to Perpustakaan Student Paper	<1 %
9	www.coursehero.com Internet Source	<1 %
10	Submitted to Sriwijaya University Student Paper	<1 %
11	Nur Pratama, Aswan Supriyadi Sunge, Eko Budiarto. "Penerapan Model MobileNetV2 Untuk Prediksi Tingkat Roasting Biji Kopi Berbasis Gambar Pada Bot Telegram", RIGGS: Journal of Artificial Intelligence and Digital Business, 2025 Publication	<1 %
12	Submitted to Universitas Sultan Ageng Tirtayasa Student Paper	<1 %
13	Widyawati Widyawati, Rafli Sidik, Ely Nuryani, Persis Haryo Winasis. "RANCANG BANGUN SISTEM DETEKSI KATARAK MENGGUNAKAN METODE CONVOLUTIONAL NEURAL NETWORK", Journal of Innovation And Future Technology (IFTECH), 2025 Publication	<1 %
14	Ananda Rizki Dani, Irma Handayani. "Classification of Yogyakarta Batik Motifs Using GLCM and CNN Methods", Jurnal Teknologi Terpadu, 2024 Publication	<1 %
15	repository.ub.ac.id Internet Source	<1 %

- | | | |
|----|---|------|
| 16 | 123dok.com
Internet Source | <1 % |
| 17 | semnasristek.sakaintek.com
Internet Source | <1 % |
| 18 | Juanto Simangunsong, Nurmala Dewi Simanjuntak, Aprima Anugerah Matondang. "Penerapan Transfer Learning untuk Klasifikasi Citra Bunga Berbasis Convolutional Neural Network", Jurnal Minfo Polgan, 2025
Publication | <1 % |
| 19 | Mohammad Bayu Anggara. "Comparison of Naïve Bayes and SVM Methods in Sentiment Analysis of User Reviews on the RSUD AL IHSAN Mobile Application", Competitive, 2025
Publication | <1 % |
| 20 | Submitted to Universitas Jember
Student Paper | <1 % |
| 21 | lib.unnes.ac.id
Internet Source | <1 % |
| 22 | dspace.uii.ac.id
Internet Source | <1 % |
| 23 | Rija Muhamad Yazid, Fajri Rakhmat Umbara, Puspita Nurul Sabrina. "Deteksi Ujaran Kebencian dengan Metode Klasifikasi Naïve Bayes dan Metode N-Gram pada Dataset Multi-Label Twitter Berbahasa Indonesia", Informatics and Digital Expert (INDEX), 2023
Publication | <1 % |
| 24 | repository.umsu.ac.id
Internet Source | <1 % |
| 25 | Yonal Supit. "Penerapan Algoritma Support Vector Machine (SVM) untuk Deteksi Penyakit | <1 % |

Mata Katarak", **INDONESIAN JOURNAL ON DATA SCIENCE, 2025**

Publication

-
- | | | |
|-----------|---|------|
| 26 | digilib.uinsby.ac.id
Internet Source | <1 % |
|-----------|---|------|
-
- | | | |
|-----------|---|------|
| 27 | eprints.umsb.ac.id
Internet Source | <1 % |
|-----------|---|------|
-
- | | | |
|-----------|---|------|
| 28 | Ni Putu Irene Pasca Nurahdika, Joko Sutopo. "Analisa Performa Convolution Neural Network Dalam Pengenalan Tulisan Tangan Aksara Lontara", <i>INTECOMS: Journal of Information Technology and Computer Science</i> , 2025
Publication | <1 % |
|-----------|---|------|
-
- | | | |
|-----------|---|------|
| 29 | Tazkira Turahman, Erfan Hasmin, Komang Aryasa. "Analisis Perbandingan Metode Convolutional Neural Network (CNN) dan MobileNet dalam Klasifikasi Penyakit Daun Padi", <i>Jurnal JTik (Jurnal Teknologi Informasi dan Komunikasi)</i> , 2024
Publication | <1 % |
|-----------|---|------|
-
- | | | |
|-----------|--|------|
| 30 | Submitted to Universitas Islam Riau
Student Paper | <1 % |
|-----------|--|------|
-
- | | | |
|-----------|---|------|
| 31 | idmetafora.com
Internet Source | <1 % |
|-----------|---|------|
-
- | | | |
|-----------|---|------|
| 32 | ejournal.st3telkom.ac.id
Internet Source | <1 % |
|-----------|---|------|
-
- | | | |
|-----------|--|------|
| 33 | Handry Eldo, Ayuliana Ayuliana, Dikky Suryadi, Giatika Chrisnawati, Loso Judijanto. "Penggunaan Algoritma Support Vector Machine (SVM) Untuk Deteksi Penipuan pada Transaksi Online", <i>Jurnal Minfo Polgan</i> , 2024
Publication | <1 % |
|-----------|--|------|
-

34	journal.trunojoyo.ac.id Internet Source	<1 %
35	Submitted to State Islamic University of Alauddin Makassar Student Paper	<1 %
36	Submitted to Universidad de Alcalá Student Paper	<1 %
37	telkomnika.uad.ac.id Internet Source	<1 %
38	www.anekapendidikan.com Internet Source	<1 %
39	eprints.dinus.ac.id Internet Source	<1 %
40	muhamadizetmutaqien.blogspot.com Internet Source	<1 %
41	www.slideshare.net Internet Source	<1 %
42	Submitted to Universitas Putera Batam Student Paper	<1 %
43	adoc.tips Internet Source	<1 %
44	ejournal.unibba.ac.id Internet Source	<1 %
45	repository.ittelkom-pwt.ac.id Internet Source	<1 %
46	repository.teknokrat.ac.id Internet Source	<1 %
47	Submitted to unikadelasalle Student Paper	<1 %

48 Jesselyn Mu, Muhammad Adrezo, Ahmed Nizhan Haikal. "Identifikasi Wajah Asli dan Buatan Deepfake Menggunakan Metode Convolutional Neural Network", Teknika, 2024
Publication <1 %

49 digilib.unila.ac.id
Internet Source <1 %

50 hes.unida.gontor.ac.id
Internet Source <1 %

51 link.springer.com
Internet Source <1 %

52 repositori.usu.ac.id
Internet Source <1 %

53 Submitted to Tarumanagara University
Student Paper <1 %

54 kc.umn.ac.id
Internet Source <1 %

55 openlibrarypublications.telkomuniversity.ac.id
Internet Source <1 %

56 publikasi.dinus.ac.id
Internet Source <1 %

57 repository.stiedewantara.ac.id
Internet Source <1 %

58 repository.uinjkt.ac.id
Internet Source <1 %

59 www.scribd.com
Internet Source <1 %

60 pt.scribd.com
Internet Source <1 %

repository.upi.edu

61

Internet Source

<1 %

62

repository.usu.ac.id

Internet Source

<1 %

63

www.alinea.id

Internet Source

<1 %

64

www.researchgate.net

Internet Source

<1 %

65

Alfataniah Nur Fajrina, Zein Hanni Pradana, Sevia Indah Purnama, Shinta Romadhona. "Penerapan Arsitektur EfficientNet-B0 Pada Klasifikasi Leukimia Tipe Acute Lymphoblastik Leukimia", Jurnal Riset Rekayasa Elektro, 2024

Publication

<1 %

66

Ferrel Sumera, Abdul Haris Junus Ontowirjo, Pinrolinvic D.K. Manembu. "Dryer System Performance Prediction Using Machine Learning Approach", Jurnal Teknik Elektro dan Komputer, 2025

Publication

<1 %

67

Gregorius Albertus Setu Gado, Putri Noraisya Primandari. "Android-Based Classification System for Cocoa Pod Diseases Using NasNet-Mobile CNN", Jurnal Teknologi Terpadu, 2025

Publication

<1 %

68

Junaiti Sahar, Henny Permatasari, Anung Ahadi Pradana, Ummi Malikal Balqis. "Elderly empowerment with evidence-based depression problems using an android application", Community Empowerment, 2021

Publication

<1 %

69

[Ton Duc Thang University](https://www.tonduc thang university.edu.vn)

Publication

<1 %

70	issuu.com Internet Source	<1 %
71	journal.isas.or.id Internet Source	<1 %
72	jurnal.syntaxliterate.co.id Internet Source	<1 %
73	medium.com Internet Source	<1 %
74	nasional.sindonews.com Internet Source	<1 %
75	podme.id Internet Source	<1 %
76	raniyulianty.com Internet Source	<1 %
77	repositori.unsil.ac.id Internet Source	<1 %
78	repository.its.ac.id Internet Source	<1 %
79	repository.pnj.ac.id Internet Source	<1 %
80	repository.unair.ac.id Internet Source	<1 %
81	riolharvinosa.blogspot.com Internet Source	<1 %
82	Widiana Kholisatun Nisa', Nur Azizah, Soffiana Agustin. "Klasifikasi Jenis Daging Sapi Dan Kuda Menggunakan Metode Support Vector Machine (SVM)", INTRO : Journal Informatika dan Teknik Elektro, 2024 Publication	<1 %

83 Yudisman Ferdinan Bili, Tundo, Nandang Sutisna, Atsilah Daini Putri, Dita Tri Yuliantoro, Laily Nurmayanti. "Prediksi Motif Batik dengan Menggunakan Metode Gabor Filter Convolution Neural Network", Jurnal JTIC (Jurnal Teknologi Informasi dan Komunikasi), 2025
Publication

84 www.repository.trisakti.ac.id
Internet Source

85 Afriani Afriani, Herry Sujaini, Niken Candraningrum. "Analisis Perbandingan Metode Pengklasifikasi Gambar Jenis Tulisan Kaligrafi Arab", Jurnal Edukasi dan Penelitian Informatika (JEPIN), 2024
Publication

86 Ertina Septia Adini, Efida Azizah, Siti Hariati Hastuti, Muhammad Ghazali. "Analisis Faktor-Faktor yang Mempengaruhi Jumlah Kasus Tuberculosis di Kabupaten Lombok Timur menggunakan Model Spatial Autoregressive Poisson", Jambura Journal of Probability and Statistics, 2025
Publication

87 Rahmat Arief Setyadi, Sayuti Rahman, Dionikxon Manurung, Mardiatul Hasanah, Asmah Indrawati. "IMPLEMENTASI TRANSFER LEARNING UNTUK KLASIFIKASI PENYAKIT PADA DAUN CABAI MENGGUNAKAN CNN", Djtechno: Jurnal Teknologi Informasi, 2024
Publication
