

## BAB III

### ANALISIS SISTEM YANG BERJALAN

#### A. Gambaran Umum Objek Penelitian

Objek penelitian yang akan diteliti adalah model *machine learning* ConvNeXtBase dan EfficientNetV2S, dalam mendeteksi malaria pada citra darah. Kedua model ini dipilih karena memiliki keunggulan dalam memproses citra dengan efisiensi tinggi dan akurasi yang baik. Dataset citra mikroskopis darah yang mengandung parasit malaria dan darah sehat tersedia secara publik pada website Kaggle. Dataset ini telah dilabeli dengan benar untuk menunjang proses pelatihan, validasi, dan pengujian model.

ConvNeXtBase adalah arsitektur convolutional neural network (CNN) modern yang dirancang untuk memperbaiki performa model CNN tradisional dengan memanfaatkan pendekatan arsitektur yang lebih efisien. Model ini memodifikasi elemen-elemen dasar CNN dengan inovasi-inovasi seperti depthwise convolution dan residual connections yang telah terbukti meningkatkan efisiensi komputasi dan akurasi. Sementara itu, EfficientNetV2S merupakan generasi lanjutan dari keluarga EfficientNet yang mengoptimalkan keseimbangan antara ukuran model, akurasi, dan waktu pelatihan menggunakan pendekatan compound scaling. Model ini dirancang untuk mengatasi masalah overfitting pada dataset dengan ukuran terbatas seperti pada penelitian ini.

Penelitian ini akan mengevaluasi performa kedua model berdasarkan metrik-metrik evaluasi seperti akurasi, precision, recall, F1-score, dan waktu komputasi. Analisis ini bertujuan untuk menentukan model yang lebih optimal dalam mendeteksi

keberadaan parasit malaria pada citra darah. Dengan perbandingan ini, diharapkan penelitian ini dapat memberikan kontribusi dalam mengidentifikasi arsitektur model terbaik yang dapat diimplementasikan secara praktis dalam dunia medis, terutama di daerah-daerah dengan sumber daya terbatas.

## **B. Analisis Kesenjangan**

Penelitian sebelumnya telah banyak berkontribusi dalam pengembangan metode diagnosis malaria berbasis citra darah, baik dengan menggunakan algoritma tradisional maupun pendekatan berbasis deep learning. Beberapa penelitian menggunakan algoritma seperti K-Nearest Neighbor (K-NN) dan Support Vector Machine (SVM), yang menunjukkan hasil akurasi yang cukup baik, namun memiliki keterbatasan dalam fleksibilitas dan efisiensi untuk dataset yang lebih besar dan kompleks. Pendekatan berbasis deep learning, terutama Convolutional Neural Network (CNN), menunjukkan performa yang lebih unggul dalam mendeteksi malaria.

Namun, sebagian besar penelitian berbasis CNN masih berfokus pada arsitektur model konvensional seperti VGG dan ResNet, serta belum mengevaluasi performa arsitektur modern seperti ConvNeXtBase dan EfficientNetV2S, yang dirancang untuk efisiensi dan akurasi yang lebih baik.

Beberapa kesenjangan pada CNN yang peneliti temukan adalah sebagai berikut:

## 1. Kinerja Model yang Beragam Berdasarkan Teknik Fine-Tuning

Penelitian yang dilakukan oleh Shekar (2022) menunjukkan bahwa teknik fine-tuning pada model VGG memberikan hasil akurasi tertinggi (99,89%), dibandingkan dengan pendekatan basic CNN dan frozen layer VGG. Hal ini mengindikasikan bahwa optimalisasi lebih lanjut pada arsitektur CNN, seperti penggunaan hyperparameter tuning yang lebih mendalam atau pemilihan arsitektur yang lebih modern, masih perlu dieksplorasi untuk mendapatkan hasil yang lebih stabil dan optimal dalam deteksi malaria.

## 2. Ketidakkonsistenan Penggunaan Fitur Saliency

Yohannes menerapkan teknik berbasis saliency seperti Region Contrast Saliency, Frequency-Tuned Saliency, Histogram Contrast Saliency, dan Spectral Residual Saliency, namun hasil menunjukkan bahwa model tanpa fitur saliency justru mencapai akurasi lebih tinggi (95,06%). Kesenjangan ini menunjukkan bahwa metode saliency yang diterapkan dalam CNN tidak selalu meningkatkan performa model, sehingga penelitian lanjutan perlu mengeksplorasi lebih dalam bagaimana teknik saliency dapat dikombinasikan dengan pendekatan CNN secara lebih efektif.

### 3. Efektivitas Model Ringan CNN untuk Implementasi di Sumber Daya Terbatas

Model ringan seperti ResNet-34, ResNet-101, AlexNet, dan VGG-19 telah diuji dalam berbagai penelitian dengan tingkat akurasi yang cukup tinggi (97,58% pada ResNet-101 dan 97% pada ResNet). Meskipun demikian, penelitian ini masih terbatas pada pengaturan learning rate tanpa melakukan optimasi lebih lanjut pada arsitektur model. Selain itu, belum ada eksplorasi mendalam terkait efisiensi komputasi model ringan dalam kondisi nyata, seperti implementasi pada perangkat seluler atau sistem dengan daya pemrosesan terbatas.

### 4. Tantangan pada Realtime Object Detection

Penelitian yang menggunakan metode deteksi objek secara real-time dengan YOLOv4 dan YOLOv5 menunjukkan variasi akurasi yang signifikan, yaitu 90,51% pada YOLOv4 dan hanya 60,7% pada YOLOv5. Hal ini menunjukkan bahwa meskipun CNN dapat digunakan dalam sistem deteksi waktu nyata, performa model masih sangat bergantung pada arsitektur dan kualitas dataset yang digunakan. Oleh karena itu, pengembangan lebih lanjut diperlukan untuk meningkatkan stabilitas dan keandalan model CNN dalam deteksi malaria secara real-time.

Untuk memberikan gambaran yang lebih jelas mengenai posisi penelitian ini dalam konteks penelitian terdahulu, berikut disajikan Table 3.1 analisis kesenjangan yang merangkum kelemahan penelitian sebelumnya dan bagaimana penelitian ini bertujuan untuk mengatasinya.

**Table 3. 1**  
**Kesenjangan Dengan Penelitian Terdahulu**

Penelitian Terdahulu	Metode yang Digunakan	Hasil dan Kelemahan	Kesenjangan
Aisyah dan Anraeni dengan judul penelitian, Analisis penerapan metode K- Nearest Neighbor (KNN) pada dataset citra penyakit malaria	K-NN	Akurasi yang di dapatkan adalah 89%, kurang fleksibel untuk dataset yang lebih kompleks.	Tidak mampu menangani dataset yang lebih besar dan kompleks.
Nur Ain Banyal, SURIANTI dan Rachman Dayat dengan judul penelitian, Klasifikasi Citra Plasmodium Penyebab Penyakit Malaria Dalam Sel Darah Merah Manusia Dengan Menggunakan Metode Multi Class Support Vector Machine (Svm)	SVM	Akurasi 73,33%; dataset kecil; kesalahan deteksi akibat pewarnaan preparat yang kurang optimal.	Akurasi rendah dan kesalahan deteksi akibat fitur manual pada dataset.
Yohannes, Siska Devella, dan Kelvin Arianto. Dengan judul penelitian, Deteksi Penyakit Malaria Menggunakan Convolutional Neural Network Berbasis Saliency	CNN berbasis saliency	Akurasi tertinggi 95,06% tanpa saliency; fitur saliency mengurangi stabilitas model.	Model berbasis saliency tidak stabil untuk deteksi malaria.
E Laxmi Lydia et al. Dengan judul penelitian, Image Classification using Deep Neural Networks for Malaria Disease Detection	ResNet-34	Akurasi 96,91%; fokus hanya pada pengaturan <i>learning rate</i> tanpa optimasi arsitektur lebih lanjut.	Tidak mengevaluasi atau membandingkan arsitektur lain yang lebih baru.

Theopilus Sasongko, dengan judul Identifikasi Plasmodium Mendeteksi Penyakit dengan Menggunakan Algorithm Convolutional Neural Network	Bayu dengan penelitian, Citra Guna Parasit Malaria	ResNet-101, AlexNet, VGG-19	ResNet-101 mencapai akurasi 97,58%; fokus hanya pada <i>learning rate</i> .	Tidak membahas model modern yang dirancang untuk efisiensi lebih tinggi.
Dhevisha Sukumarran et al, dengan judul penelitian Automated Identification of Malaria-Infected Cells and Classification of Human Malaria Parasites Using a Two-Stage Deep Learning Technique		CNN YOLOv4 dan YOLOv5.	YOLOv4 mencapai akurasi 90,51%; YOLOv5 hanya 60,7%.	Tidak membahas model untuk klasifikasi citra malaria pada skala piksel yang lebih kecil.

Sumber: Olahan Penulis

### C. Metodologi Penelitian

#### 1. Teknik Pengumpulan Data

Pada penelitian ini peneliti memerlukan data. Pengumpulan data ini dilakukan dengan cara, sebagai berikut:

##### a. Observasi

Penulis melakukan observasi terhadap dataset citra darah yang tersebar di platform berbagi dataset seperti kaggle, dan huggingface. Peneliti mendapatkan dataset berjumlah 27.588 citra darah yang terbaik menjadi dua folder yaitu: *parasitized* (terinfeksi) dan *Uninfected* (tidak terinfeksi). Dataset diperoleh dari National Library of Medicine (NIH) melalui platform Kaggle.

b. Studi Literatur/Pustaka

Peneliti juga akan menerapkan teknik sekunder ini dalam proses pengumpulan data dengan cara melihat jurnal-jurnal terdahulu yang berkaitan dengan penelitian identifikasi malaria menggunakan CNN.

2. Teknik Analisis Data

CRISP-DM (Cross-Industry Standard Process for Data Mining) adalah metodologi standar untuk proyek *data mining* yang dirancang agar fleksibel dan dapat diterapkan pada berbagai konteks penelitian atau bisnis. Metode ini terdiri dari 6 fase utama yang bersifat adaptif, artinya setiap fase dapat disesuaikan dengan kebutuhan proyek. Berikut penjelasan rinci tiap fase:

a. Fase Pemahaman Bisnis/Penelitian (*Business/Research Understanding Phase*)

- (1) Mulailah dengan menyampaikan secara jelas tujuan dan kebutuhan proyek, baik dari perspektif bisnis maupun unit penelitian secara keseluruhan.
- (2) Selanjutnya, ubah tujuan dan batasan tersebut menjadi perumusan yang spesifik dalam definisi masalah data mining.
- (3) Terakhir, susun strategi awal untuk mewujudkan tujuan yang telah ditentukan.

b. Pemahaman Data (*Data Understanding*)

- (1) Langkah Pertama adalah mengumpulkan data.
- (2) Selanjutnya, lakukan analisis eksplorasi data untuk memahami karakteristik data dan menemukan wawasan awal.
- (3) Evaluasi kualitas data yang tersedia.

- (4) Terakhir, pilih data yang relevan dan memiliki potensi mengandung pola yang dapat digunakan untuk langkah selanjutnya.

**c. Fase Persiapan Data (Data Preparation Phase):**

- (1) Fase ini melibatkan semua langkah yang diperlukan untuk menyiapkan set data akhir, termasuk data awal, data mentah, dan data yang belum bersih .
- (2) Pilih kasus dan variabel yang relevan untuk analisis lebih lanjut.
- (3) Lakukan transformasi pada variabel tertentu sesuai kebutuhan.
- (4) Bersihkan data mentah agar siap digunakan dalam proses pemodelan.

**d. Fase Model (Modeling Phase):**

- (1) Teknik pemodelan yang sesuai dipilih dan diterapkan pada data.
- (2) Pengaturan model dikalibrasi untuk mendapatkan hasil yang optimal.
- (3) Sering kali, berbagai teknik diterapkan untuk memecahkan masalah data mining yang sama.
- (4) Jika diperlukan, fase ini dapat kembali ke tahap persiapan data untuk menyesuaikan data dengan kebutuhan spesifik dari teknik yang digunakan.

**e. Fase Evaluasi (Evaluation Phase):**

- (1) Model yang dihasilkan dari fase pemodelan harus dievaluasi untuk menilai kualitas dan efektivitasnya sebelum diimplementasikan.
- (2) Evaluasi dilakukan untuk memastikan model telah memenuhi tujuan yang ditetapkan pada tahap awal.
- (3) Peninjauan dilakukan untuk mengidentifikasi apakah ada aspek penting dari masalah bisnis atau penelitian yang belum dipertimbangkan dengan baik.

(4) Keputusan akhir dibuat mengenai penggunaan hasil dari proses data mining.

**f. Fase Penerapan Data (Deployment Phase):**

(1) Proses pembuatan model bukan akhir dari proyek; model yang telah dibuat harus diterapkan secara nyata.

(2) Contoh penerapan sederhana adalah menghasilkan laporan berdasarkan hasil model.

(3) Penerapan yang lebih kompleks dapat mencakup implementasi proses data mining di departemen lain.

(4) Dalam konteks bisnis, hasil model sering kali digunakan oleh pelanggan untuk pengambilan keputusan atau penerapan lanjutan.

**3. Teknik Pengukuran Data**

**1. *Confusion Matrix***

Untuk menilai validitas dan keakuratan data yang digunakan dalam penelitian ini, dilakukan pengukuran kinerja model klasifikasi dengan menggunakan *Confusion Matrix*. *Confusion Matrix* adalah matriks 2x2 yang terdiri dari empat elemen:

(1) True Positive (TP): Jumlah prediksi benar untuk kelas positif.

(2) True Negative (TN): Jumlah prediksi benar untuk kelas negatif.

(3) False Positive (FP): Jumlah prediksi salah di mana data negatif diprediksi sebagai positif.

(4) False Negative (FN): Jumlah prediksi salah di mana data positif diprediksi sebagai negatif.

Contoh *confusion matrix* dapat dilihat pada Tabel 3.2

**Table 3. 2**  
**Contoh Confusion Matrix**

	<i>Predicted Parasitized</i>	<i>Predicted Uninfected</i>
<i>Actual Parasitized</i>	<i>True Positive</i>	<i>False Negative</i>
<i>Actual Uninfected</i>	<i>False Positive</i>	<i>True Negative</i>

Sumber: PRATIWI et al (2021)

Berdasarkan Confusion Matrix, metrik-metrik berikut digunakan untuk mengevaluasi kinerja model:

(1) Akurasi

Akurasi mengukur seberapa banyak prediksi yang benar dari total data, berikut adalah rumusnya:

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN}$$

Akurasi menunjukkan proporsi data yang berhasil diklasifikasikan dengan benar oleh model.

(2) Presisi

Presisi mengukur seberapa andal model dalam memprediksi kelas positif:

$$\text{Presisi} = \frac{TP}{TP+FP}$$

Nilai presisi yang tinggi menunjukkan bahwa model jarang menghasilkan positif palsu.

(3) Recall

Recall mengukur seberapa baik model dapat mendeteksi kelas positif:

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

Recall yang tinggi menunjukkan bahwa model mampu mengidentifikasi sebagian besar data positif.

(4) *F1-Score*

F1-Score adalah rata-rata harmonis dari presisi dan recall, yang memberikan keseimbangan antara keduanya:

$$\text{Recall} = 2 \cdot \frac{\text{Presisi} \cdot \text{Recall}}{\text{Presisi} + \text{Recall}}$$

F1-Score sangat berguna ketika terdapat ketidakseimbangan kelas pada dataset.