ABSTRAK

Alexithymia merupakan gangguan psikologis yang ditandai oleh kesulitan individu dalam mengenali, memahami, dan mengungkapkan emosinya sendiri. Kondisi ini berdampak langsung pada kemampuan penderita berkomunikasi dengan orang-orang di sekitarnya, sehingga interaksi sosial sehari-hari sering kali terhambat. Untuk mengatasi keterbatasan tersebut, penderita membutuhkan alat bantu yang dapat mengidentifikasi emosi secara langsung (*real-time*) ketika berinteraksi dengan pendamping atau orang terdekat. Teknologi *Facial Emotion Recognition* (FER) berbasis kecerdasan buatan menjadi salah satu solusi yang berpotensi menjawab kebutuhan ini. Meskipun demikian, sebagian besar penelitian FER yang telah dilakukan selama ini lebih berfokus pada peningkatan akurasi model, sementara aspek waktu inferensi (*inference time*) kerap luput dari perhatian. Padahal, dalam konteks interaksi sehari-hari, sistem yang merespons secara lambat justru berpotensi mengganggu kelancaran komunikasi sosial penderita *Alexithymia*. Berdasarkan kesenjangan tersebut, penelitian ini menghadirkan analisis komparatif secara sistematis untuk menentukan arsitektur model yang paling optimal, yakni arsitektur yang mampu menyeimbangkan akurasi tinggi dengan latensi rendah, sebagai landasan pengembangan alat bantu yang efektif dan andal bagi penderita *Alexithymia*.

Penelitian ini membandingkan kinerja dua arsitektur, yaitu ResNet-50 dan InceptionV3, yang keduanya dikembangkan menggunakan pendekatan *Transfer Learning* pada *dataset* RAF-DB *in-the-wild* yang berjumlah 15.339 gambar. Tahap awal penelitian dimulai dengan pengumpulan data, dilanjutkan dengan proses praproses berupa deteksi dan pemotongan otomatis area wajah menggunakan *Haar-Cascade Classifier*. Selanjutnya, data dibagi dengan proporsi 80:10:10 untuk mencegah terjadinya kebocoran data (*data leakage*) antara data latih, validasi, dan uji. Guna mengatasi ketidakseimbangan jumlah data pada tiap kelas emosi, diterapkan teknik *hybrid resampling* sehingga setiap kelas memiliki jumlah gambar yang setara, yaitu 1.968 gambar per kelas. Kedua arsitektur kemudian diuji melalui 12 skenario eksperimen dengan variasi parameter, meliputi *learning rate*, *Dropout*, dan jumlah lapisan yang dibuka (*Unfreeze Layer*).

Hasil pengujian menunjukkan bahwa arsitektur ResNet-50 secara konsisten unggul dibandingkan InceptionV3, baik dari segi akurasi maupun kecepatan pemrosesan. Akurasi tertinggi yang dicapai ResNet-50 adalah sebesar 84,62% pada Skenario 2, yaitu pada kombinasi *Dropout* 0,5/0,4, *learning rate* tertentu, serta 20% *Unfreeze Layer*, dengan waktu inferensi yang tergolong efisien, yakni 223,72 ms per gambar. Sebagai perbandingan, InceptionV3 hanya mencapai akurasi 83,31% dengan waktu inferensi yang lebih lama, yaitu 288,44 ms. Analisis *Pareto Frontier* menunjukkan bahwa ResNet-50 memberikan keseimbangan terbaik antara akurasi dan latensi.

Kata Kunci: *Alexithymia*, *Facial Emotion Recognition*, ResNet-50, InceptionV3, *Transfer Learning*

ABSTRACT

Alexithymia is a psychological disorder characterized by an individual's difficulty in recognizing, understanding, and expressing their own emotions. This condition directly impacts the affected individual's ability to communicate with those around them, often hindering daily social interactions. To overcome these limitations, individuals with this condition need tools that can identify emotions in real time when interacting with caregivers or loved ones. Artificial intelligence-based Facial Emotion Recognition (FER) technology is one potential solution to address this need. However, most FER research conducted to date has focused primarily on improving model accuracy, while the aspect of inference time has often been overlooked. In the context of daily interactions, a system that responds slowly has the potential to disrupt the smooth flow of social communication for individuals with Alexithymia. To address this gap, this study presents a systematic comparative analysis to determine the most optimal model architecture one capable of Balancing high accuracy with low latency as the foundation for developing effective and reliable assistive tools for individuals with Alexithymia.

This study compares the performance of two architectures ResNet-50 and InceptionV3 both of which were developed using a Transfer Learning approach on the RAF-DB in-the-wild dataset consisting of 15,339 images. The initial phase of the study began with data collection, followed by a Pre-processing step involving the automatic detection and cropping of facial areas using the Haar-Cascade Classifier. Next, the data was split in an 80:10:10 ratio to prevent data leakage between the training, validation*, and test sets. To address the imbalance in the number of images across each emotion class, a hybrid resampling technique was applied so that each class had an equal number of images1,968 images per class. Both architectures were then tested across 12 experimental scenarios with varying parameters, including learning rate, Dropout, and the number of Unfrozen layers.*

The test results show that the ResNet-50 architecture consistently outperforms InceptionV3 in terms of both accuracy and processing speed. The highest accuracy achieved by ResNet-50 was 84.62% in Scenario 2 specifically, with a Dropout combination of 0.5/0.4, a specific learning rate, and 20% Unfrozen layers with a relatively efficient inference time of 223,72 ms per image. By comparison, InceptionV3 only achieved an accuracy of 83.31% with a longer inference time of 288.44 ms. The Pareto Frontier analysis shows that ResNet-50 provides the best balance between accuracy and latency.

Keywords: *Alexithymia, Facial Emotion Recognition, ResNet-50, InceptionV3, Transfer Learning*