ABSTRAK

Penelitian pengenalan ekspresi wajah umumnya masih terfokus pada emosi dasar, padahal interaksi manusia di dunia nyata lebih sering menampilkan ekspresi wajah majemuk (*Compound Facial Expressions*), yaitu gabungan dari dua atau lebih emosi dasar. Pengenalan ekspresi majemuk atau *Compound Expression Recognition* (CER) merupakan tantangan signifikan karena fitur visualnya yang subtil dan ambigu. Solusi yang ada saat ini, seperti *ensemble learning*, sering kali membutuhkan sumber daya komputasi yang besar sehingga tidak efisien untuk implementasi praktis. Di sisi lain, arsitektur modern yang dirancang untuk efisiensi seperti EfficientNetV2 masih jarang dieksplorasi untuk tugas CER yang kompleks ini. Oleh karena itu, penelitian ini bertujuan mengisi kesenjangan tersebut dengan mengevaluasi secara sistematis kinerja arsitektur EfficientNetV2-S sebagai solusi yang akurat sekaligus efisien secara komputasi untuk mengatasi permasalahan klasifikasi ekspresi wajah majemuk.

Metode penelitian yang digunakan adalah eksperimen kuantitatif dengan mengimplementasikan model Convolutional Neural Network (CNN). Penelitian ini memanfaatkan subset ekspresi majemuk dari dataset publik *Real-world Affective Faces Database* (RAF-DB) yang berisi 3.954 citra dalam 11 kelas emosi. Dataset dibagi secara terstratifikasi menjadi 80% data latih, 10% validasi, dan 10% uji. Proses pra-pemrosesan data meliputi pengubahan ukuran seluruh citra menjadi 224x224 piksel dan penerapan teknik augmentasi data pada set pelatihan untuk meningkatkan variasi. Model utama dibangun menggunakan arsitektur EfficientNetV2-S dengan pendekatan *transfer learning* dari bobot ImageNet. Kinerja model ini kemudian dievaluasi melalui 24 skenario pengujian *hyperparameter* yang berbeda dan dibandingkan secara langsung dengan arsitektur generasi sebelumnya, yaitu EfficientNet-B0, untuk memvalidasi keunggulannya.

Hasil penelitian membuktikan bahwa arsitektur EfficientNetV2-S merupakan solusi yang kompetitif. Konfigurasi terbaiknya mencapai akurasi pengujian sebesar 60.85%, terbukti lebih unggul dibandingkan EfficientNet-B0 (58.8%) dan dicapai dengan waktu komputasi yang lebih singkat. Temuan paling signifikan adalah bahwa performa optimal diperoleh tanpa melakukan *fine-tuning*, yang mengindikasikan bahwa bobot pra-terlatih dari *ImageNet* sudah sangat efektif sebagai pengekstraksi fitur tetap. Meskipun demikian, teridentifikasi adanya tantangan *overfitting* yang ditandai oleh kesenjangan antara akurasi pelatihan yang tinggi dan akurasi validasi yang stagnan. Sebagai kontribusi utama, model ini mencapai nilai rata-rata diagonal *confusion matrix* (*average*) sebesar 45.95%, sebuah hasil yang mampu bersaing dengan penelitian rujukan pada dataset RAF-DB, sekaligus membuktikan potensinya sebagai alternatif modern yang efisien.

.

Kata Kunci: Ekspresi Wajah Majemuk, Convolutional Neural Network, EfficientNetV2, *Transfer Learning*

ABSTRACT

Facial expression recognition research generally focuses on basic emotions, whereas real-world human interactions more frequently display compound facial expressions, a combination of two or more basic emotions. Compound Expression Recognition (CER) presents a significant challenge due to its subtle and ambiguous visual features. Current solutions, such as ensemble learning, often require substantial computational resources, rendering them inefficient for practical implementation. However, modern architectures designed for efficiency, like EfficientNetV2, remain largely unexplored for this complex CER task. Therefore, this research aims to fill this gap by systematically evaluating the performance of the EfficientNetV2-S architecture as an accurate and computationally efficient solution for the classification of compound facial expressions.

The research method employed is a quantitative experiment implementing a Convolutional Neural Network (CNN) model. This study utilizes the compound expression subset from the public Real-world Affective Faces Database (RAF-DB), which contains 3,954 images across 11 emotion classes. The dataset was stratified into 80% training, 10% validation, and 10% test sets. The data preprocessing process included resizing all images to 224x224 pixels and applying data augmentation techniques to the training set to increase variation. The primary model was built using the EfficientNetV2-S architecture with a transfer learning approach from ImageNet weights. The model's performance was then evaluated through 24 different hyperparameter testing scenarios and directly compared with the previous-generation architecture, EfficientNet-B0, to validate its superiority.

The results demonstrate that the EfficientNetV2-S architecture is a competitive solution. Its best configuration achieved a test accuracy of 60.85%, proving superior to EfficientNet-B0 (58.8%) while also requiring less computation time. The most significant finding was that optimal performance was achieved without fine-tuning, indicating that the pre-trained ImageNet weights are highly effective as a fixed feature extractor. Nevertheless, an overfitting challenge was identified, marked by a gap between high training accuracy and stagnant validation accuracy. As a primary contribution, this model achieved an average diagonal confusion matrix score of 45.95%, a result that is comparable to benchmark studies, thereby proving its potential as an efficient, modern alternative.

Keywords: *Compound Facial Expressions, Convolutional Neural Network, EfficientNetV2, Transfer Learning*