

ABSTRACT

Facial image-based emotion detection has become a growing area of research with various applications, including psychology, neuroscience, and public services, as it is considered helpful in understanding an individual's emotional state in a short time. Previous studies have employed methods such as Convolutional Neural Networks (CNN), Support Vector Machines (SVM), and hybrid approaches like Convolutional Vision Transformers (CVT) to detect emotions in facial images with relatively high accuracy. However, these methods have certain limitations; most of the research still relies on manual tuning, which is time-consuming, resource-intensive, prone to overfitting, not scalable, and has a low probability of finding the optimal solution. This study aims to implement a Vision Transformer (ViT) model optimized using Grid Search Hyperparameter Tuning to detect emotions from facial images. The research uses the AffectNet dataset, which contains 450,000 facial images covering eight primary emotion classes. The model is evaluated using accuracy, precision, recall, and F1-score metrics to assess its performance. This study is expected to make a significant contribution to the development of facial image-based emotion detection methods with high accuracy through a more efficient and optimal approach.

Keywords: *emotion detection, Vision Transformer, facial image, Hyperparameter Tuning, AffectNet.*