UNIVERSITAS
GADJAH MADA
Analisis Stacking Ensemble Berbasis Multiple CNN untuk Klasifikasi Ekspresi Wajah pada Dataset Raf-DB
Ahmad Fida' Ulhaq, Prof. Ir. Oyas Wahyunggoro, M.T., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

INTISARI

Pengenalan ekspresi wajah merupakan salah satu permasalahan dalam bidang *Computer Vision* yang menghadapi tantangan dalam menghasilkan klasifikasi yang konsisten pada citra dengan kondisi *in-the-wild*. Variasi karakteristik citra dan ketidakseimbangan distribusi kelas pada *dataset* RAF-DB dapat menyebabkan perbedaan kemampuan model dalam mengenali setiap kategori ekspresi. Perbedaan karakteristik arsitektur *Convolutional Neural Network* (CNN) juga memungkinkan setiap model memiliki keunggulan dan keterbatasan dalam mengklasifikasikan ekspresi wajah. Penelitian ini menganalisis performa model CNN tunggal serta penerapan *stacking ensemble* untuk mengetahui pengaruh penggabungan prediksi beberapa model terhadap performa klasifikasi ekspresi wajah.

Penelitian dilakukan menggunakan *dataset* RAF-DB yang terdiri atas tujuh kelas ekspresi wajah. Tiga arsitektur CNN, yaitu InceptionV3, ResNet-50, dan ConvNeXt-Tiny, digunakan sebagai *base-learner* melalui proses *fine-tuning* dengan bobot awal pralatih ImageNet. Augmentasi citra dan *Weighted Random Sampling* diterapkan pada data pelatihan untuk mendukung pembelajaran pada distribusi kelas yang tidak seimbang. Prediksi probabilitas dari ketiga *base-learner* digunakan sebagai *meta-feature* melalui mekanisme *Out-of-Fold Prediction* dengan *Stratified K-Fold Cross-Validation* sebanyak lima *fold*. Lima algoritma, yaitu Logistic Regression, Support Vector Machine (SVM), Random Forest, Multi-Layer Perceptron (MLP), dan XGBoost, dievaluasi sebagai kandidat meta-learner. Optimasi *hyperparameter* dilakukan menggunakan *Grid Search*, sedangkan evaluasi performa mencakup *accuracy*, *macro precision*, *macro recall*, dan *macro F1-score*.

Hasil penelitian menunjukkan bahwa ConvNeXt-Tiny menghasilkan performa terbaik di antara model tunggal dengan *accuracy* sebesar 88,90% dan *macro F1-score* sebesar 82,53%. Pada pendekatan *stacking ensemble*, Random Forest memperoleh *macro F1-score* tertinggi sebesar 82,82%, dengan *accuracy* sebesar 89,02%, *macro precision* sebesar 86,08%, dan *macro recall* sebesar 80,52%. Sementara itu, Logistic Regression menghasilkan *accuracy* tertinggi sebesar 89,28%, sedangkan MLP memperoleh *macro F1-score* sebesar 82,45% Berdasarkan *macro F1-score* sebagai acuan utama untuk menilai keseimbangan performa pada distribusi kelas yang tidak seimbang, Random Forest ditetapkan sebagai *meta-learner* terbaik. Dibandingkan ConvNeXt-Tiny, Random Forest meningkatkan *macro F1-score* sebesar 0,29%. Hasil ini menunjukkan bahwa *stacking ensemble* memberikan peningkatan performa agregat yang terbatas dan bahwa hasilnya bergantung pada algoritma *meta-learne*r yang digunakan.

Kata kunci: *Facial Expression Recognition, Convolutional Neural Network, Stacking Ensemble*, RAF-DB, *Meta-Learner*

UNIVERSITAS
GADJAH MADA
Analisis Stacking Ensemble Berbasis Multiple CNN untuk Klasifikasi Ekspresi Wajah pada Dataset Raf-DB
Ahmad Fida' Ulhaq, Prof. Ir. Oyas Wahyunggoro, M.T., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

ABSTRACT

Facial Expression Recognition (FER) remains a challenging task in computer vision, particularly when classifying facial images captured under in-the-wild conditions. Variations in image characteristics and class imbalance in the RAF-DB dataset can affect a model's ability to recognize different facial expression categories. Furthermore, differences in Convolutional Neural Network (CNN) architectures may result in varying strengths and limitations in facial expression classification. This study analyzes the performance of individual CNN models and the application of a stacking ensemble to investigate the effect of combining predictions from multiple models on facial expression classification performance.

The experiments were conducted using the RAF-DB dataset, which consists of seven facial expression classes. Three CNN architectures, namely InceptionV3, ResNet-50, and ConvNeXt-Tiny, were employed as base learners through fine-tuning with ImageNet-pretrained weights. Image augmentation and Weighted Random Sampling were applied to the training data to address the imbalanced class distribution. The probability predictions generated by the three base learners were used as meta-features through an Out-of-Fold Prediction mechanism based on five-fold Stratified K-Fold Cross-Validation. Five algorithms—Logistic Regression, Support Vector Machine (SVM), Random Forest, Multi-Layer Perceptron (MLP), and XGBoost—were evaluated as candidate meta-learners. Hyperparameter optimization was performed using Grid Search, and model performance was evaluated based on accuracy, macro precision, macro recall, and macro F1-score.

The results indicate that ConvNeXt-Tiny achieved the best performance among the individual models, with an accuracy of 88.90% and a macro F1-score of 82.53%. For the stacking ensemble, Random Forest achieved the highest macro F1-score of 82.82%, with an accuracy of 89.02%, macro precision of 86.08%, and macro recall of 80.52%. Meanwhile, Logistic Regression achieved the highest accuracy among the stacking models at 89.28%, while MLP obtained a macro F1-score of 82.45%. Based on the macro F1-score, which was used as the primary metric for assessing balanced performance across imbalanced classes, Random Forest was identified as the best-performing meta-learner. Compared with ConvNeXt-Tiny, Random Forest improved the macro F1-score by 0.29%. These findings indicate that the stacking ensemble provides a modest improvement in aggregate performance, with the resulting performance depending on the meta-learning algorithm employed.

Keywords : Facial Expression Recognition, Stacking Ensemble, Convolutional Neural Network, RAF-DB, Meta-Learner