

INTISARI

Ekspresi wajah merupakan salah satu cara paling dasar bagi manusia dalam mengungkapkan emosi yang dirasakannya. Dalam melakukan klasifikasi ekspresi wajah, *machine learning* dapat digunakan untuk mengenali fitur pada wajah dan mengotomatisasi proses klasifikasi. *Transfer learning* pada *Convolutional Neural Network* (CNN) dapat digunakan dalam proses klasifikasi tersebut. Setiap arsitektur *transfer learning* CNN memiliki variasi dalam strukturnya dengan kelebihan masing-masing. Dalam memanfaatkan kelebihan tersebut, dapat digunakan *ensemble learning* berupa *blending* dan *voting* yang memanfaatkan variasi arsitektur CNN. Metode tersebut akan dibandingkan dengan FER2013 sebagai dataset perbandingan untuk menentukan metode yang lebih efektif dalam meningkatkan performa *transfer learning* CNN. Metode tersebut juga akan dibandingkan dengan performa model tunggal *transfer learning* CNN untuk mengetahui peningkatan akurasi yang diperoleh. Lalu, perbedaan jumlah model dasar yang digunakan dalam *ensemble learning* juga akan dibandingkan dengan singkat untuk mengetahui apakah penggunaan lima model dasar lebih baik dibandingkan tiga model dasar.

Arsitektur *transfer learning* CNN yang digunakan meliputi VGG16, ResNet50, InceptionV3, EfficientNetV2B2, dan DenseNet121. Arsitektur tersebut digunakan sebagai model dasar dalam *ensemble learning* dan model tunggal dalam proses evaluasi. Metrik evaluasi yang digunakan adalah akurasi, *precision*, *recall*, *f1-score*, serta *confusion matrix*, dengan *loss* digunakan khusus untuk arsitektur CNN. Proses pelatihan model tunggal dilakukan untuk mengetahui performa model tunggal dan sebagai model dasar dalam penerapan metode *blending* dan *voting*. Metode *voting* dilakukan berupa *soft voting* dan *hard voting*. Sementara itu, *blending* dilakukan dengan *meta learner* berupa *Logistic Regression*, *Support Vector Machine*, *K-Nearest Neighbor*, dan *Random Forest*, dengan hiperparameter terbaik yang dicari menggunakan *stratified k-fold cross validation*. Kombinasi tiga model juga diuji dengan kombinasi lima model dalam setiap metode tersebut.

Analisis kinerja memberikan hasil berupa performa lima model dasar yang secara konsisten mengungguli beberapa kombinasi tiga model dasar yang digunakan. Lalu, metode *blending* dengan *meta learner Logistic Regression* memiliki akurasi sebesar 72,78% yang menyamai akurasi dari *soft voting*. Metode *blending* memiliki keunggulan dari segi *precision* sebesar 73,12% dan *f1-score* sebesar 72,72% yang lebih tinggi dibandingkan *precision* sebesar 73,05% dan *f1-score* sebesar 72,56% yang dimiliki oleh metode *soft voting*. Oleh karena itu, dapat disimpulkan bahwa metode *blending* dengan *meta learner Logistic Regression* merupakan metode terbaik, dengan metode *soft voting* menjadi alternatif yang memiliki penerapan yang lebih simpel dengan performa yang hampir setara. Selain itu, didapat pula peningkatan akurasi pada metode *blending* dengan *meta learner Logistic Regression* dan metode *soft voting* sebesar 3,12% dari model tunggal tertinggi VGG16 sebesar 69,66%. Dari penelitian yang sudah dilakukan, pengembangan model klasifikasi ekspresi wajah ke depannya dapat menggunakan metode *voting* dan *blending* tersebut.

Kata kunci : Klasifikasi Ekspresi Wajah, *Transfer Learning*, CNN, *Blending*, *Voting*

ABSTRACT

Facial expressions are one of the most fundamental ways for humans to express their emotions. In facial expression recognition, machine learning can be utilized to identify facial features and automate the classification process. Transfer learning on Convolutional Neural Network (CNN) can be utilized for this classification process. Each CNN transfer learning architecture has variations in its models, each with its own advantages. To leverage these advantages, ensemble learning methods such as blending and voting can be used to utilize the diversity of CNN architectures. These methods will be compared using FER2013 dataset as benchmark to determine which method is more effective in improving the performance of CNN transfer learning. These methods will also be compared with the performance of single CNN transfer learning models to observe the accuracy improvement obtained. Furthermore, the difference in the number of base models used in ensemble learning will be briefly compared to determine whether using five base models is better than three base models.

The CNN transfer learning architectures used in this research are VGG16, ResNet50, InceptionV3, EfficientNetV2B2, and DenseNet121. Those architectures are used as base models in ensemble learning and single model for evaluation. The evaluation metrics used including accuracy, precision, recall, f1-score, and confusion matrix, where loss evaluation is specific to CNN architectures. Each single model is trained to evaluate its performance as each single model and as base models for blending and voting methods. The voting methods are performed using soft voting and hard voting. Meanwhile, blending are performed with meta learner models including Logistic Regression, Support Vector Machine, K-Nearest Neighbor, and Random Forest, with the best hyperparameters obtained using stratified k-fold cross validation. The combination of three models is also tested against the combination of five models in each method.

Performance analysis using the evaluation metrics resulted in the performance of the five base models consistently outperforming several combinations of three base models. Then, the blending method with the Logistic Regression as meta learner achieved an accuracy of 72.78%, which matches the accuracy of the soft voting method. The blending method has advantages in terms of precision at 73.12% and f1-score at 72.72%, which are higher than the precision of 73.05% and f1-score of 72.56% achieved by the soft voting method. Therefore, it can be concluded that the blending method with Logistic Regression as the meta learner is the best method, with the soft voting method serving as an alternative method that is simpler to implement with performance nearly equivalent to the blending method. Furthermore, an accuracy improvement of 3.12% was obtained for the blending method with the Logistic Regression meta learner and the soft voting method, compared to the highest single model, VGG16, which achieved 69.66%. From the research conducted, future development of facial expression classification models can utilize the voting and blending methods.

Keywords : Facial Expression Recognition, Transfer Learning, CNN, Blending, Voting