

PERANCANGAN ALGORITMA ATRIBUSI UCAPAN SINTETIS DENGAN MENGGUNAKAN METODE TRANSFER LEARNING

Nama Mahasiswa : Satria Putra Akbar
NIM : 04201082
Dosen Pembimbing Utama : Mifta Nur Farid, S.T., M.T..
Dosen Pembimbing Pendamping : Kharis Sugiarto, SST.,M.T.

ABSTRAK

Teknologi *synthetic speech* saat ini mampu meniru ucapan manusia secara sempurna. Meskipun teknologi ini menciptakan peluang untuk pengembangan aplikasi suara yang canggih, namun juga membawa risiko serangan serius seperti manipulasi suara, peniruan suara, dan penggandaan suara. Beberapa upaya telah dilakukan terkait atribusi ucapan sintetis dengan berbagai macam pemodelan. Namun, beberapa metode tersebut memerlukan spesifikasi tinggi dan waktu komputasi yang lama dalam pelatihan. Oleh karena itu, diperlukan penggunaan metode lain dengan spesifikasi yang ada dengan hasil tingkat akurasi tinggi serta mengurangi waktu komputasi. Penelitian ini bertujuan untuk merancang algoritma atribusi ucapan sintetis yang efisien dengan menggunakan metode *Transfer Learning* dan arsitektur CNN, yakni *InceptionV3*, *MobileNetV2*, dan *NASNetMobile*. Model-model ini diuji dengan tanpa *fine-tuning* dan dengan *fine-tuning* serta melakukan analisis terhadap pengurangan jumlah layer, ini diharapkan mampu menyederhanakan model dan meningkatkan efisiensi pelatihan model tanpa mengorbankan akurasi. Hasil penelitian menunjukkan bahwa model *MobileNetV2* memberikan akurasi terbaik mencapai 93% baik dengan maupun tanpa *fine-tuning*, dengan waktu pelatihan hanya 52 menit. Pengurangan layer pada *MobileNetV2* menunjukkan akurasi 90% dengan 50 layer, tetapi menurun menjadi 23% setelah *fine-tuning*. Model *InceptionV3* mencapai akurasi 91% setelah *fine-tuning*, meningkat dari 88% tanpa *fine-tuning*, dengan waktu pelatihan 143 menit. Pengurangan layer pada *InceptionV3* menunjukkan akurasi 97% dengan 150 layer, sementara akurasi 89% tercapai dengan 249 layer, yang juga mengurangi waktu pelatihan menjadi 86 menit. *NASNetMobile* menunjukkan akurasi 88% tanpa *fine-tuning*, tetapi menurun menjadi 82% setelah *fine-tuning*, dengan waktu pelatihan 66 menit. Pengurangan layer pada *NASNetMobile* menunjukkan akurasi 92% dengan 600 layer, yang juga mengurangi waktu pelatihan. Evaluasi menggunakan *confusion matrix* menunjukkan bahwa *fine-tuning* secara signifikan meningkatkan presisi, *recall*, dan *f1-score*, khususnya untuk kelas-kelas yang sulit dikenali. Berdasarkan hasil tersebut, model *MobileNetV2* terbukti lebih stabil dalam kedua pendekatan memberikan hasil yang efisien dan akurat, serta memiliki potensi besar untuk diterapkan pada sistem atribusi ucapan sintetis dengan keterbatasan sumber daya.

Kata Kunci : Atribusi, *Confusion matrix*, *Synthetic Speech*, *Transfer Learning*