

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan Berdasarkan hasil penelitian dan pembahasan yang telah dilakukan mengenai klasifikasi tingkat keparahan Retinopati diabetik menggunakan metode *Stacking Ensemble*, dapat ditarik kesimpulan sebagai berikut:

1. Penelitian ini berhasil merancang dan menerapkan tiga arsitektur *Deep Learning* sebagai model *base learners*, yaitu *EfficientNetV2-S* dengan efisiensi blok *Fused-MBConv*, *ResNet50* yang menggunakan mekanisme *skip connections* untuk mengatasi *vanishing gradient*, dan *DenseNet121* yang menerapkan *dense connectivity* untuk memaksimalkan penggunaan ulang fitur. Ketiga model dilatih menggunakan *dataset* gabungan (APTOS 2019, IDRiD, dan Messidor-2) yang telah melalui tahapan pra-pemrosesan meliputi *ROI Extraction*, *Gamma correction*, *CLAHE*, *Resizing*, dan *Balancing Data*.
2. Penelitian ini juga berhasil merancang arsitektur *Stacking Ensemble* yang diimplementasikan dengan mengintegrasikan probabilitas prediksi dari ketiga *base learners*, yaitu *EfficientNetV2-S*, *ResNet50*, dan *DenseNet121*, untuk membentuk vektor *meta-features* berdimensi 15. Vektor ini kemudian diproses oleh algoritma *Gradient Boosting Classifier* sebagai *meta-learner*, yang bekerja dengan membangun serangkaian *decision trees* secara sekuensial untuk meminimalkan fungsi *loss* menggunakan metode *gradient descent*. Mekanisme ini memungkinkan *meta-learner* untuk mempelajari pola kesalahan yang kompleks dari model *baseline* dan secara efektif mengoreksi bias prediksi, sehingga menghasilkan keputusan klasifikasi akhir yang lebih akurat.
3. Hasil pengujian dan evaluasi performa dari masing-masing model *base learners* dan *Stacking Ensemble* pada tiga skenario pembagian Data (80:10:10, 70:15:15, dan 60:20:20) dengan tiga metode sampling (*Random sampling*, *Undersampling*, dan *No Augmentation*). *EfficientNetV2-S* dengan *Random sampling* pada skenario 70:15:15 mencapai performa terbaik dengan akurasi 83,11%, *Precision* 0,8304, *recall* 0,8311, *F1-Score* 0,8289, dan QWK 0,9298. *ResNet50* mencapai performa terbaik pada skenario 80:10:10 dengan *Random sampling* dengan akurasi 79,50%

dan QWK 0,9203. *DenseNet121* mencapai performa terbaik pada skenario 70:15:15 dengan *Random sampling* dengan akurasi 79,78% dan QWK 0,9182. *Stacking Ensemble* menunjukkan performa terbaik pada skenario 70:15:15 dengan *Random sampling* mencapai akurasi 82,78% dan QWK 0,9293. Analisis *Confusion matrix* menunjukkan bahwa model-model terbaik mampu mengklasifikasikan pada kelas *Grade 0* dan *Grade 4*. Metode *Random sampling* secara konsisten menghasilkan performa terbaik di seluruh arsitektur dengan rata-rata akurasi 79,98% dan QWK 0,9114, diikuti *No Augmentation* dengan rata-rata akurasi 76,76% dan QWK 0,8705, sedangkan *Undersampling* menghasilkan performa terendah dengan rata-rata akurasi 63,31% dan QWK 0,8478.

4. Hasil penelitian menemukan fakta menarik bahwa meskipun *EfficientNetV2-S* mampu mencatatkan performa paling unggul karena menduduki peringkat 1 dari skenario 70:15:15, tetapi metode *Stacking Ensemble* juga terbukti memiliki ketahanan yang lebih unggul karena mempertahankan peringkat 2 dan 3 dari keseluruhan arsitektur.

5.2 Saran Pengembangan

Berdasarkan hasil penelitian, keterbatasan yang ditemukan, dan batasan penelitian yang telah ditetapkan, beberapa saran untuk pengembangan penelitian selanjutnya adalah:

- 1) Penelitian ini dibatasi pada pengembangan model dan evaluasi metrik (*Accuracy*, QWK, *Precision*, *recall*, *F1-Score*, *Confusion matrix*) tanpa implementasi sistem. Penelitian selanjutnya dapat mengembangkan aplikasi berbasis *web* atau *mobile* untuk *deployment* sistem dalam *tele-ophthalmology programs* atau *screening campaigns* di daerah tak terjangkau, dengan fitur tambahan seperti *quality assessment* citra fundus, *report generation*, dan integrasi dengan *Electronic Health Record* (EHR).
- 2) Penelitian ini terbatas pada penggunaan *Gradient Boosting* sebagai *meta-learner*. Penelitian selanjutnya dapat mengeksplorasi *meta-learners* alternatif seperti *XGBoost*, *LightGBM*, *CatBoost*, atau bahkan *Neural Network-based meta-learner* (seperti *Multi-Layer Perceptron*) yang mungkin dapat mempelajari pola integrasi yang lebih kompleks dari *meta-features*, serta membandingkan performa relatif

terhadap *Gradient Boosting*.

- 3) Penelitian ini menggunakan Data gabungan dari tiga *dataset* benchmark (APTOS 2019, IDRiD, Messidor-2). Penelitian selanjutnya dapat melakukan *external validation* pada *dataset* independen dari sumber berbeda (misalnya, EyePACS, DDR, DRIVE) atau *dataset* lokal Indonesia untuk mengkonfirmasi generalisabilitas model pada populasi dengan karakteristik demografis, etnis, dan prevalensi penyakit yang berbeda.
- 4) Menambahkan arsitektur CNN modern lainnya (seperti *Vision Transformer/ViT*, *ConvNeXt*, *EfficientNetV2-M/L*, atau *Swin Transformer*) sebagai *base learners* tambahan untuk meningkatkan *diversity* dalam *Ensemble*, berpotensi meningkatkan performa *Stacking* dengan mengeksplorasi mekanisme pembelajaran yang lebih beragam.
- 5) Mengeksplorasi metode *ordinal regression* yang lebih eksplisit seperti CORAL (*Consistent Rank Logits*), *Ordinal Entropy*, atau *Soft Labels* yang telah terbukti untuk masalah klasifikasi *ordinal*, berpotensi meningkatkan QWK lebih lanjut dengan mempelajari struktur *ordinal* secara lebih efektif.
- 6) Melakukan uji klinis prospektif dengan melibatkan *ophthalmologist* untuk membandingkan performa sistem CAD dengan *Diagnosis* manual, mengukur *inter-observer agreement*, dan mengevaluasi dampak sistem terhadap *workflow* klinis, *diagnostic Accuracy*, dan *patient outcomes* dalam setting klinis yang sesungguhnya.
- 7) Berkolaborasi dengan Kementerian Kesehatan, rumah sakit rujukan, atau program *tele-ophthalmology* nasional untuk mengintegrasikan sistem CAD yang dikembangkan ke dalam *screening programs* Retinopati diabetik di Indonesia, berkontribusi langsung terhadap peningkatan akses layanan kesehatan mata dan pencapaian SDGs poin 3 (*Good Health and Well-Being*).