

BAB II

KAJIAN PUSTAKA

2.1 Tinjauan Pustaka

Tinjauan pustaka dalam penelitian ini disusun secara sistematis untuk memetakan perkembangan teknologi klasifikasi citra di bidang perikanan, mulai dari metode ekstraksi fitur tradisional hingga pemanfaatan arsitektur *deep learning* yang kompleks. Kajian ini mencakup berbagai metodologi yang telah dikembangkan oleh peneliti sebelumnya. Penelusuran literatur bertujuan untuk mengidentifikasi celah penelitian (*research gap*) serta posisi keaslian dari metode yang diusulkan dalam penelitian ini.

2.1.1 Metode Berbasis Fusi Fitur

Metode berbasis fusi fitur dapat digunakan untuk mengklasifikasikan tingkat kesegaran ikan secara otomatis dan akurat (Balim & Olgun, 2025). Penelitian ini secara khusus mengintegrasikan informasi pelengkap dari dua modalitas data, yaitu citra visual dan data pantulan laser, untuk meningkatkan kinerja klasifikasi. Dalam prosesnya, fitur-fitur visual diekstrak dari gambar ikan menggunakan dua model *deep learning* yang telah dilatih sebelumnya, *EfficientNetB0* dan *ResNet18*, yang berfungsi sebagai ekstraktor fitur yang kuat untuk menangkap karakteristik penting dari citra.

Sementara itu, data pantulan laser yang merupakan sinyal 1-dimensi diolah menggunakan model *CNN*. Model ini dirancang secara khusus untuk mengekstrak fitur-fitur dari spektrum pantulan laser yang terkait dengan kondisi biokimia ikan. Selanjutnya, proses fusi fitur menggabungkan fitur-fitur yang telah diekstrak dari kedua modalitas (citra dan laser) menjadi satu vektor fitur komprehensif. Vektor fitur hasil fusi kemudian dimasukkan ke dalam pengklasifikasi untuk memprediksi dan mengkategorikan tingkat kesegaran ikan ke dalam tiga kelas, yaitu *fresh*, *medium fresh*, dan *Stale*. Penelitian ini menggunakan dataset khusus yang dikumpulkan sendiri oleh peneliti, yang mencakup total 3600 sampel ikan jenis *sea bream*.

Untuk pengembangan di masa depan, peneliti menyarankan untuk melakukan pengujian metode fusi fitur ini pada berbagai spesies ikan lain untuk menguji generalisasi dan kehandalan sistem pada variasi biologis yang lebih luas. Eksplorasi teknik fusi fitur yang lebih canggih, seperti fusi pada level keputusan (*decision-level fusion*) atau fusi pada level yang lebih dalam (*deep-level fusion*), juga perlu diteliti untuk melihat potensi peningkatan kinerja. Selain itu, pengembangan sistem yang dapat mengintegrasikan ekstraksi fitur dan klasifikasi secara *real-time* sangat disarankan untuk aplikasi komersial, misalnya di pabrik pengolahan ikan atau pasar. Terakhir, peneliti menyarankan untuk menyelidiki kemungkinan merancang sebuah jaringan saraf tunggal yang dapat memproses data citra dan data laser secara simultan, alih-alih menggunakan dua model terpisah.

2.1.2 Metode *Ensemble* Heterogen

Metode penggunaan *ensemble* heterogen dari model *deep learning* dapat digunakan untuk meningkatkan kekokohan dan kinerja model klasifikasi citra ikan melalui penggabungan prediksi dari berbagai arsitektur jaringan saraf konvolusional (Wyatt dkk., 2022). Penelitian ini menggunakan dua arsitektur *CNN deep learning* populer, yaitu *ResNet-50* dan *VGG-16*, sebagai fondasi model dasar (*base models*). Model-model *ensemble* kemudian dibangun dengan menggabungkan dua model dasar tersebut, di mana peneliti mengeksplorasi beberapa skema *ensemble* yang berbeda.

Skema yang dieksplorasi meliputi model tunggal *ResNet-50* dan *VGG-16* yang digunakan sebagai *baseline*, serta model *ensemble ResNet-50* dan *VGG-16* yang dibuat dengan melatih ulang beberapa model yang sama pada bagian-bagian data yang berbeda. Skema paling beragam adalah *hybrid ensemble*, yang menggabungkan model-model dari arsitektur yang berbeda. Prediksi akhir dari setiap *ensemble* diperoleh melalui mekanisme voting untuk menentukan kelas yang paling mungkin. Untuk melatih dan menguji model, penelitian ini menggunakan dua dataset citra laut yang berbeda.

Dataset ikan karang, yang terdiri dari 9 spesies ikan karang dengan total 1280 citra, dan dataset ikan perdagangan, yang terdiri dari 3 spesies ikan perdagangan dengan total 1810 citra. Untuk pengembangan di masa depan, penelitian selanjutnya dapat dikembangkan dengan menerapkan dan menguji metode ini pada dataset yang lebih besar dan mencakup lebih banyak spesies ikan, untuk memvalidasi skalabilitas metode *ensemble*. Peneliti juga menyarankan untuk menguji metode *ensemble* ini dengan arsitektur *CNN* yang lebih baru atau berbeda, seperti *MobileNet*, *EfficientNet*, atau *Vision Transformers*, untuk melihat apakah ada peningkatan kinerja lebih lanjut.

Selain itu, eksplorasi metode *ensemble* yang lebih canggih, seperti stacking atau bagging, disarankan karena mungkin dapat memberikan kinerja yang lebih optimal dibandingkan skema *voting*. Terakhir, peneliti dapat mengembangkan pengklasifikasi dengan pengenalan objek, sehingga metode *ensemble* tidak hanya digunakan untuk klasifikasi, tetapi juga mampu mendeteksi dan mengklasifikasikan beberapa ikan dalam satu citra.

2.1.3 Metode *dropout-ELM*

Metode *dropout-ELM* dapat digunakan untuk mengatasi masalah *overfitting* pada pengklasifikasi *extreme learning machine (ELM)*, terutama ketika model dihadapkan pada data berukuran sampel kecil. (Tao & Zhong, 2025). Penelitian ini mengintegrasikan teknik regulasi *dropout* ke dalam arsitektur pengklasifikasi *ELM*. Selama proses pelatihan, *dropout* secara acak menonaktifkan *neuron* pada *hidden layer* dengan probabilitas tertentu (*dropout rate*). Teknik ini secara efektif memaksa model untuk mempelajari representasi yang lebih kuat dan independen, serta secara signifikan mengurangi ketergantungan antar *neuron*, yang pada dasarnya meniru *ensemble* dari banyak sub-jaringan yang "ditipiskan" dan merata-ratakan prediksi mereka selama inferensi.

Untuk mengevaluasi kemampuan algoritma *dropout-ELM*, peneliti menggunakan dua pendekatan dataset: pertama, Dataset Buatan yang dibuat menggunakan perangkat lunak *WEKA* untuk mengontrol karakteristiknya dan memungkinkan evaluasi yang teliti terhadap pengaruh berbagai kondisi pada

kinerja model; dan kedua, Dataset Dunia Nyata yang berupa lima dataset dari *Kaggle* untuk memastikan bahwa temuan penelitian relevan dan dapat diterapkan pada skenario nyata. Semua evaluasi dilakukan di *Python 2.2.0*, dengan spesifikasi perangkat keras *2.40 GHz CPU*, *16 GB RAM*, dan *500 GB hard disk*. Setiap dataset secara acak dibagi menjadi 80% untuk pelatihan dan 20% untuk pengujian, dan hasil yang dilaporkan merupakan rata-rata dari 10 eksperimen independen yang diulang.

Hasil penelitian berhasil menunjukkan efektivitas *dropout-ELM* dalam mengatasi *overfitting* pada dataset buatan dan dataset dunia nyata yang dicirikan dengan ukuran sampel yang kecil. Meskipun penelitian ini berhasil, penelitian lebih lanjut dapat mengeksplorasi penerapannya di domain masalah lain, seperti pengenalan citra, pemrosesan bahasa alami, atau data medis dengan data yang terbatas, untuk memvalidasi generalisasinya. Selain itu, mengoptimalkan dan menyetel parameter *dropout-ELM* menjadi fokus penting, karena rasio dropout dan jumlah neuron pada *hidden layer* memengaruhi kinerja model secara keseluruhan penelitian di masa depan dapat fokus pada pengembangan metode yang lebih canggih untuk menyetel parameter ini secara otomatis.

Terdapat juga potensi untuk menggabungkan *dropout-ELM* dengan teknik lain, seperti teknik *ensemble* yang lebih kompleks, seperti *bagging* atau *boosting* yang disesuaikan untuk konteks data berukuran sampel kecil, guna meningkatkan kinerja lebih lanjut. Terakhir, karena penelitian ini terbatas pada dataset dengan ukuran sampel kecil, evaluasi pada dataset yang lebih besar di masa depan diperlukan untuk menguji skalabilitas dan kekokohan (*robustness*) metode *dropout-ELM* pada data yang lebih kompleks.

2.1.4 Metode *Open-Set Recognition (OSR)*

Metode *open-set recognition* dapat digunakan untuk mengklasifikasikan spesies ikan yang dikenal sambil secara bersamaan mendeteksi instansi ikan dari spesies yang tidak dikenal. Penelitian ini menguji dan membandingkan tiga metode *OSR* terdepan (*state-of-the-art*) untuk mengatasi tantangan dalam klasifikasi citra ikan.(Córdova dkk., 2025). Penelitian ini memfokuskan pada tiga

metode open-set recognition sebagai metode utama untuk klasifikasi ikan. *Multiple Gaussian Prototype Learning (MGPL)* menggunakan conditional variational *auto-encoder* untuk mempelajari representasi fitur laten di mana spesies yang dikenal dan tidak dikenal dapat dipisahkan.

Selama inferensi, *MGPL* merepresentasikan setiap kelas dengan beberapa *prototipe Gaussian*, dan sampel baru diklasifikasikan sebagai spesies yang diketahui jika jaraknya ke prototipe terdekat berada di bawah ambang batas yang telah ditentukan. Selanjutnya, *Probability of Inclusion Support Vector Machine (PISVM)*, sebagai ekstensi dari *Support Vector Machine*, melatih pengklasifikasi biner menggunakan pendekatan *one-vs-all* dan kemudian memasang distribusi Weibull ke skor margin untuk setiap kelas. Sampel diklasifikasikan sebagai spesies yang dikenal jika probabilitas inklusinya berada di atas ambang batas. Terakhir, *Open-Set Nearest Neighbor (OSNN)* adalah metode non-parametrik yang didasarkan pada algoritma *k-nearest neighbor*.

Alih-alih menerapkan ambang batas pada skor kesamaan, *OSNN* menggunakan ambang batas pada rasio jarak antara dua sampel terdekat yang paling mirip; jika rasio ini melebihi ambang batas, sampel diklasifikasikan sebagai tidak dikenal. Khusus untuk metode *OSNN* dan *PISVM*, fitur diekstrak menggunakan model *ResNet-18* yang dilatih pada spesies yang dikenal. Penelitian ini menggunakan dataset *fish detection and Weight Estimation (FDWE)* yang tersedia untuk umum, yang berisi 1086 gambar dengan total 2216 instansi ikan dari sembilan spesies umum di laut utara. Dataset ini mencakup data ground truth seperti spesies, koordinat *bounding box*, berat, dan tingkat oklusi, tetapi memiliki beberapa karakteristik yang menantang.

Tantangan utamanya adalah ketidakseimbangan kelas, di mana lebih dari 50% ikan termasuk dalam kelas mayoritas sementara beberapa spesies lain hanya memiliki kurang dari 60 instansi. Tantangan lain adalah tingginya variasi intra-kelas (variasi besar dalam penampilan ikan di dalam spesies yang sama) dan rendahnya variasi antar-kelas (beberapa spesies memiliki kemiripan yang tinggi dalam bentuk, ukuran, dan warna). Selain itu, adanya oklusi, di mana ikan dapat

tertutup sebagian oleh puing-puing atau ikan lain, membuat tugas klasifikasi menjadi lebih kompleks.

2.1.5 Metode *Enhanced Convolutional Neural Network (ECNN)*

Metode *Enhanced Convolutional Neural Network* dapat digunakan untuk mengatasi kelemahan dan meningkatkan efektivitas model *deep learning* sebelumnya dalam tugas klasifikasi citra. (Yang, 2025). Arsitektur *ECNN* dirancang dengan mengintegrasikan beberapa teknik canggih untuk mengoptimalkan aliran informasi dan ekstraksi fitur. Pertama, *ECNN* menggunakan *skip connections*, mirip dengan *ResNet*, untuk memastikan aliran informasi yang efektif di seluruh lapisan jaringan sehingga membantu mencegah masalah *vanishing gradient* yang sering terjadi pada jaringan yang sangat dalam. Kedua, *ECNN* mengimplementasikan struktur inception yang simetris untuk mengekstrak fitur pada berbagai skala secara paralel, sehingga memungkinkan model untuk menangkap pola dan representasi fitur yang lebih kaya dari citra. Ketiga, alih-alih hanya menggunakan satu jenis *pooling*, *ECNN* menggabungkan lapisan *average pooling* dan *max pooling* yang bertujuan untuk mendapatkan representasi fitur yang lebih komprehensif, di mana *max pooling* fokus pada fitur yang paling dominan dan *average pooling* menangkap fitur latar belakang.

Selain itu, penelitian ini menerapkan teknik augmentasi data dan dropout untuk meningkatkan kemampuan generalisasi model dan mencegah overfitting, karena teknik-teknik ini secara efektif memperbanyak variasi data pelatihan dan mengurangi ketergantungan antar *neuron*. Penelitian ini menggunakan dua dataset citra yang umum digunakan untuk mengevaluasi kinerja model: *MNIST* dataset, yang berisi 70.000 citra tulisan tangan digit dari 0 hingga 9, dan *CIFAR-10* dataset, yang berisi 60.000 citra berwarna dari 10 kelas objek yang berbeda. Kedua dataset ini dipilih karena ukurannya yang relatif kecil dan strukturnya yang terstandarisasi, yang memungkinkan perbandingan kinerja yang adil antara model-model yang berbeda.

Meskipun *ECNN* menunjukkan kinerja yang sangat baik, penelitian selanjutnya disarankan untuk menguji coba arsitektur ini pada dataset yang lebih

besar dan kompleks seperti *ImageNet*, untuk memvalidasi skalabilitas dan kemampuan generalisasi model pada skenario dunia nyata. Selain itu, penelitian di masa depan dapat mengeksplorasi variasi dari arsitektur *ECNN*, seperti mengoptimalkan jumlah lapisan, rasio *dropout*, atau parameter *inception block* untuk menemukan konfigurasi yang lebih optimal. Terakhir, peneliti dapat mengembangkan pengklasifikasi khusus dengan mengadaptasi arsitektur *ECNN* untuk tugas-tugas klasifikasi spesifik lainnya, misalnya, untuk citra medis atau pengenalan objek yang lebih kompleks yang dapat menghasilkan kinerja yang lebih unggul.

2.1.6 Metode Kombinasi Model *Deep Learning*

Metode kombinasi model *deep learning* untuk ekstraksi fitur dengan berbagai algoritma machine learning untuk klasifikasi dapat digunakan untuk mengevaluasi kesegaran ikan secara objektif dan akurat menggunakan gambar tubuh ikan yang dikategorikan sebagai segar atau tidak segar (Yasin dkk., 2023). Penelitian ini bertujuan untuk menyediakan metode yang inil dan hemat biaya untuk evaluasi kesegaran ikan, yang sangat penting dalam industri makanan untuk kontrol kualitas dan keamanan pangan. Metode ini melibatkan dua tahap utama: ekstraksi fitur menggunakan dua model *deep learning* terdepan, *SqueezeNet* dan *InceptionV3*, yang berfungsi menangkap berbagai fitur dari citra ikan. Setelah fitur diekstrak, lima algoritma *machine learning* digunakan untuk klasifikasi kesegaran, yaitu *K-Nearest Neighbors*, *Support Vector Machines*, *Artificial Neural Networks*, *Logistic Regression*, dan *Random Forest*.

Hasil penelitian menunjukkan bahwa kombinasi model *deep learning* dan algoritma *machine learning* ini mampu mencapai tingkat akurasi yang sangat tinggi dalam klasifikasi kesegaran ikan. Secara spesifik, kombinasi *SVM* dan *LR* dengan *SqueezeNet* mencapai akurasi 100%, demikian pula *ANN* dengan *SqueezeNet* juga mencapai akurasi 100%. Kinerja serupa juga terlihat ketika menggunakan *inceptionV3*, di mana kombinasi *SVM* dan *LR* dengan *InceptionV3* mencapai akurasi 100%, dan *ANN* dengan *InceptionV3* juga mencapai akurasi 100%. Kelebihan utama dari penelitian ini adalah akurasi yang sangat tinggi

(hingga 100%) dalam mengklasifikasikan kesegaran ikan. Selain itu, penggunaan *squeezeNet* memberikan efisiensi komputasi karena merupakan jaringan yang ringan, sehingga cocok untuk *deployment* pada perangkat dengan sumber daya terbatas.

Kelebihan lain mencakup fleksibilitas dan generalisasi karena model *deep learning* yang digunakan mampu mengekstrak berbagai fitur dari citra, memungkinkan model *machine learning* untuk generalisasi dengan baik pada data baru. Tetapi, metode ini juga memiliki beberapa kelemahan, yaitu ketergantungan pada kualitas citra, di mana keakuratan model sangat bergantung pada kualitas gambar yang digunakan, dan citra dengan kualitas rendah atau kondisi pencahayaan yang buruk dapat mempengaruhi kinerja model. Selain itu, terdapat kompleksitas implementasi karena implementasi dan pelatihan model *deep learning* membutuhkan sumber daya komputasi yang signifikan serta pengetahuan teknis yang mendalam. Kelemahan terakhir adalah data terbatas, karena dataset yang digunakan dalam penelitian ini mungkin tidak mencakup semua variasi ikan dan kondisi yang dapat ditemui dalam aplikasi nyata, sehingga model mungkin memerlukan penyesuaian lebih lanjut untuk diterapkan secara luas.

2.1.7 Metode *Differential Evolution based Fine-Tuning (DEFT)*

Metode *Differential Evolution based Fine-Tuning* dapat digunakan untuk secara otomatis memilih lapisan mana dari jaringan saraf konvolusional yang harus disesuaikan (*fine-tuned*) demi mencapai kinerja klasifikasi yang optimal (Vrbančić dan Member, 2020). Penelitian ini mengusulkan *DEFT* sebagai solusi untuk tantangan dalam *transfer learning*. Metode ini mengadaptasi algoritma *meta-heuristik Differential Evolution (DE)* yang berfungsi untuk mencari kombinasi lapisan yang paling optimal untuk di *fine-tune*.

Dalam implementasinya, penelitian ini menggunakan pendekatan *transfer learning* di mana bobot dari arsitektur *VGG19* yang sudah dilatih pada dataset *ImageNet* ditransfer ke tugas baru, yaitu identifikasi *osteosarcoma*. Untuk mengevaluasi seberapa baik setiap kombinasi lapisan yang diusulkan oleh algoritma *DE*, fungsi kerugian *Categorical Cross-Entropy (CCE)* digunakan

sebagai fungsi kebugaran (*fitness function*). Kinerja metode *DEFT* kemudian dibandingkan dengan beberapa skenario lain, termasuk Model *CNN* yang dilatih konvensional, Model *pre-trained* dengan semua lapisan di *fine-tune*, dan Model *CNN* dengan lapisan yang di *fine-tune* secara manual (*handpicked*).

Untuk pengembangan lebih lanjut, peneliti menyarankan untuk menerapkan algoritma optimasi lain, seperti *firefly algorithm* atau *particle swarm optimization*, guna mengevaluasi apakah metode tersebut dapat memberikan hasil yang lebih baik atau lebih efisien dalam memilih lapisan. Penelitian selanjutnya juga dapat mengaplikasikan metode *DEFT* pada dataset citra medis yang berbeda, atau bahkan pada tugas klasifikasi citra di luar domain medis, untuk menguji seberapa baik kinerja metode ini tanpa memining domain klasifikasi. Selain itu, peneliti disarankan untuk menjelajahi cara untuk membuat mekanisme pemilihan lapisan menjadi lebih efisien agar dapat memberikan kinerja klasifikasi yang lebih baik dengan kompleksitas waktu yang lebih rendah.

2.1.8 Metode *EfficientNetB0* Feature Extraction dan Ensemble Learning

Metode *EfficientNetB0* feature extraction dan ensemble learning dapat digunakan untuk mengklasifikasikan citra daging sapi dan babi secara kuat. Penulis mengidentifikasi isu serius mengenai pemalsuan daging yang berdampak pada aspek kesehatan serta kepatuhan agama masyarakat. Sistem visi komputer menawarkan solusi pemeriksaan otomatis untuk membedakan tekstur daging babi yang sangat mirip dengan daging sapi. Kebutuhan akan metode otentikasi makanan yang efisien mendorong pengembangan teknologi klasifikasi berbasis kecerdasan buatan ini.

Penelitian ini mengimplementasikan arsitektur *EfficientNetB0* sebagai pengekstraksi fitur utama untuk menangkap karakteristik visual daging secara mendalam. Dengan menggabungkan beberapa algoritma seperti *RLDA*, *Support Vector Machine (SVM)*, dan *Random Forest* ke dalam sebuah sistem pembelajaran ensemble. Mereka menggunakan mekanisme *soft voting* untuk menyatukan prediksi dari berbagai model guna meminimalkan kesalahan klasifikasi. Proses

pengolahan citra ini melibatkan dataset sebanyak 400 gambar dengan variasi pencahayaan yang berbeda di laboratorium.

Eksperimen tersebut menghasilkan tingkat akurasi yang sangat tinggi sebesar 99,0% pada pengujian *hold-out*. Peneliti menerapkan teknik Gradient-weighted *Class Activation Mapping (Grad-CAM)* untuk memberikan interpretasi visual terhadap keputusan model. Peta panas dari *Grad-CAM* menunjukkan bahwa sistem memfokuskan analisis pada area otot dan intensitas saluran warna merah untuk membedakan jenis daging. Hasil ini membuktikan bahwa integrasi *transfer learning* dan *ensemble* dapat memberikan hasil yang sangat presisi dan transparan.

Kesimpulan dari penelitian ini adalah penggunaan model hibrida secara signifikan meningkatkan performa generalisasi dibandingkan model tunggal. Penulis memberikan kontribusi penting dalam bidang teknologi pangan melalui penyediaan sistem klasifikasi yang dapat dipertanggungjawabkan secara visual. Implementasi metode ini dapat membantu otoritas pengawas makanan dalam mendeteksi kontaminasi daging babi secara lebih cepat. Penelitian masa depan dapat memperluas cakupan dataset dengan jenis daging lain untuk meningkatkan fleksibilitas sistem di pasar global.

2.1.9 Metode Deep Convolutional Neural Network (DCNN) dengan Augmentasi

Metode *Deep Convolutional Neural Network (DCNN)* dengan augmentasi data dapat digunakan untuk mengklasifikasikan spesies ikan pada dataset yang terbatas. Penulis menyoroti kekayaan hayati laut di koridor Pulau Verde, Filipina, yang memerlukan sistem pemantauan otomatis bagi perlindungan ekosistem. Klasifikasi spesies ikan secara manual membutuhkan waktu yang lama serta pengetahuan taksonomi yang sangat mendalam dari para ahli. Teknologi pembelajaran mendalam hadir sebagai alat bantu yang krusial untuk mempercepat proses identifikasi biota laut secara akurat.

Dengan memodifikasi arsitektur *VGG16* dengan menerapkan teknik *retraining* dan *fine-tuning* pada lapisan-lapisan saraf konvolusi. Mereka mengatasi kendala kurangnya data gambar melalui penerapan teknik augmentasi data seperti rotasi, pergeseran lebar, dan pembalikan horizontal. Langkah ini menghasilkan

data sintetis yang memperkaya variasi fitur sehingga model dapat belajar secara lebih optimal meskipun menggunakan sampel asli yang sedikit. Penelitian ini memfokuskan pelatihan pada tiga spesies ikan dominan yang memiliki karakteristik visual unik di perairan tersebut.

Hasil pengujian menunjukkan bahwa model tersebut mencapai tingkat akurasi yang luar biasa sebesar 99% melalui optimasi parameter yang tepat. Augmentasi data secara efektif mencegah terjadinya *overfitting* dan meningkatkan kemampuan model dalam mengenali fitur spesifik ikan. Peneliti mendapati bahwa penyesuaian laju pembelajaran (*learning rate*) memberikan pengaruh besar terhadap kecepatan konvergensi model selama proses pelatihan. Keberhasilan ini membuktikan bahwa teknologi *DCNN* sangat hinil dalam menangani tugas klasifikasi citra yang kompleks di lingkungan alami.

Peneliti menegaskan bahwa metodologi ini memberikan fondasi yang kuat bagi pengembangan sistem manajemen perikanan nasional di Filipina. Penggunaan data sintetis terbukti menjadi solusi cerdas bagi penelitian yang terkendala oleh sulitnya mendapatkan sampel gambar berkualitas tinggi. Studi ini merekomendasikan integrasi sistem klasifikasi ini ke dalam perangkat pemantauan bawah air untuk mendukung konservasi laut secara berkelanjutan. Penelitian selanjutnya diharapkan dapat mencakup lebih banyak spesies ikan agar sistem memiliki kemampuan deteksi yang lebih luas.

2.1.10 Ekstraksi Fitur dan Pengenalan Gambar

Metode *Convolutional Neural Networks (CNN)* dapat digunakan untuk melakukan ekstraksi fitur otomatis dan pengenalan gambar yang kompleks. Penulis menjelaskan bahwa struktur *CNN* mengadopsi prinsip kerja sistem visual makhluk hidup yang memproses informasi secara hierarkis. Teknologi ini memungkinkan komputer untuk mengenali pola visual mulai dari garis sederhana hingga bentuk objek yang sangat rumit secara mandiri. Literatur ini menekankan transformasi signifikan dalam bidang visi komputer melalui penggunaan lapisan saraf yang saling terhubung secara fungsional.

Peneliti menguraikan secara sistematis peranan setiap lapisan fungsional seperti lapisan konvolusi, lapisan aktivasi *ReLU*, dan lapisan pengumpulan (*pooling*). Lapisan konvolusi berfungsi untuk mengekstraksi fitur optik dari gambar mentah melalui penggunaan berbagai filter spasial. Sementara itu, lapisan pooling mereduksi dimensi data untuk mempercepat proses komputasi tanpa menghilangkan informasi fitur yang penting. Penulis menggunakan basis data *MNIST* yang berisi puluhan ribu digit tulisan tangan sebagai objek eksperimen untuk menguji kehandalan struktur jaringan tersebut.

Penelitian ini menunjukkan bahwa algoritma *CNN* mampu mengenali karakter digit dengan tingkat presisi yang jauh melampaui metode tradisional. Sistem ini secara otomatis mempelajari representasi fitur yang relevan selama fase pelatihan melalui mekanisme *backpropagation*. Peneliti menekankan bahwa kemampuan ekstraksi fitur otomatis merupakan keunggulan utama yang membedakan *CNN* dari algoritma pembelajaran mesin konvensional. Hasil eksperimen pada dataset *MNIST* memperkuat argumen bahwa arsitektur hierarkis sangat efektif untuk menangani variasi bentuk visual.

Kesimpulannya bahwa *CNN* merupakan arsitektur paling menjanjikan untuk pengembangan teknologi pengenalan gambar di masa depan. Makalah ini memberikan pemahaman teoretis yang mendalam mengenai mekanisme internal jaringan saraf konvolusi bagi para peneliti pemula. Fleksibilitas model *CNN* memungkinkan penerapannya pada berbagai domain seperti pengenalan wajah, deteksi medis, hingga kendaraan otonom. Pengembangan algoritma ini secara berkelanjutan akan terus mendorong batasan kemampuan kecerdasan buatan dalam memahami dunia visual.

2.1.11 Metode *Deep Learning* dan *Machine Learning* Terintegrasi

Metode *deep learning* dan *machine learning* terintegrasi pada aplikasi seluler dapat digunakan untuk mendeteksi tingkat kesegaran mata ikan secara instan. Peneliti menekankan pentingnya konsumsi ikan segar bagi pemenuhan nutrisi dan pencegahan keracunan makanan pada manusia. Identifikasi kesegaran ikan secara tradisional melalui bau atau tekstur kulit seringkali bersifat subjektif dan tidak

konsisten bagi konsumen awam. Oleh karena itu, tim pengembang menciptakan sistem cerdas yang mampu memberikan penilaian objektif berdasarkan karakteristik mata ikan.

Peneliti mengembangkan dataset khusus bernama *Fish Freshness Eye (FFE)* untuk melatih model kecerdasan buatan dengan berbagai tingkat degradasi kualitas. Mereka menerapkan algoritma *SqueezeNet* dan *VGG19* sebagai pengekstraksi fitur untuk mendapatkan representasi numerik dari citra mata ikan. Peneliti membandingkan performa lima algoritma klasifikasi yang berbeda termasuk *artificial neural network (ANN)*, *random forest*, dan *support vector Machine*. Integrasi antara fitur mendalam dan model pembelajaran mesin bertujuan untuk mencapai hasil prediksi yang paling stabil.

Hasil penelitian mencatatkan bahwa kombinasi model *VGG19* dan *ANN* memberikan tingkat keberhasilan klasifikasi tertinggi sebesar 77,3%. Peneliti kemudian menanamkan model terbaik tersebut ke dalam sebuah aplikasi seluler berbasis *Android* untuk memudahkan penggunaan di lapangan. Aplikasi ini memungkinkan pengguna untuk mengambil foto mata ikan dan menerima hasil deteksi kesegaran secara langsung melalui layar *smartphone*. Inovasi ini menjembatani celah antara penelitian laboratorium yang kompleks dengan kebutuhan praktis masyarakat di pasar tradisional.

Penelitian ini memberikan kontribusi nyata dalam upaya penjaminan keamanan pangan melalui pemanfaatan teknologi digital yang terjangkau. Penulis membuktikan bahwa analisis bagian mata ikan secara otomatis dapat menjadi indikator yang *valid* untuk menentukan kualitas protein. Pengembangan aplikasi mobile ini diharapkan dapat meningkatkan literasi konsumen mengenai pentingnya memilih produk perikanan yang berkualitas tinggi.

2.1.12 Metode XgBoost Berbasis Fitur Gist dan GLCM

Metode *XgBoost* berbasis fitur *Gist* dan *GLCM* dapat digunakan untuk meningkatkan akurasi klasifikasi citra ikan di berbagai kondisi lingkungan. Peneliti mengidentifikasi kompleksitas dalam pengenalan spesies ikan akibat adanya variasi bentuk, ukuran, serta gangguan *noise* pada citra bawah air.

Pemilihan fitur yang tepat menjadi faktor penentu keberhasilan sistem klasifikasi dalam membedakan spesies yang memiliki kemiripan visual yang tinggi. Penelitian ini bertujuan untuk menciptakan metode klasifikasi yang tangguh dan efisien bagi industri perikanan serta survei ekologi laut.

Penelitian ini menerapkan tahap pra-pemrosesan yang meliputi penghapusan noise dan pengubahan ukuran gambar secara seragam sebelum melakukan ekstraksi fitur. Mereka menggabungkan fitur *Gist* untuk menangkap struktur spasial global dari seluruh gambar dengan fitur *Gray Level Co-occurrence Matrix (GLCM)* untuk analisis tekstur lokal. Penggabungan kedua jenis fitur ini menciptakan matriks fitur hibrida yang memberikan informasi sangat detail mengenai karakteristik setiap spesies ikan. Penelitian ini menggunakan algoritma *XgBoost* yang dikenal memiliki kecepatan dan performa tinggi dalam menangani data tabular hasil ekstraksi fitur.

Eksperimen yang dilakukan pada dataset *QUT* dan *F4K* menunjukkan hasil yang sangat kompetitif dibandingkan dengan metode lainnya. Algoritma *XgBoost* mencapai akurasi sebesar 90,2% pada dataset *QUT* serta akurasi yang lebih tinggi yaitu 98,08% pada dataset *F4K*. Peneliti menemukan bahwa sinergi antara fitur global dan lokal secara signifikan mampu mengurangi ambiguitas selama proses pengambilan keputusan oleh model. Kehandalan metode ini terbukti konsisten meskipun diuji pada kumpulan data yang memiliki karakteristik lingkungan perairan yang berbeda-beda.

Penelitian ini menyimpulkan bahwa integrasi fitur spasial dan tekstur merupakan pendekatan yang sangat efektif untuk tugas klasifikasi citra ikan yang rumit. Penulis memberikan solusi teknis yang dapat menghemat waktu secara signifikan dalam proses penghitungan dan inventarisasi populasi ikan. Hasil penelitian ini memiliki potensi aplikasi yang luas dalam bidang pertanian, layanan memancing, serta aplikasi survei perikanan otomatis. Pengembangan selanjutnya dapat difokuskan pada pengujian algoritma *deep learning* lainnya untuk dibandingkan dengan efisiensi yang ditawarkan oleh model *XgBoost*.

2.2 Keaslian Penelitian

Tabel 2.1 menyajikan rangkuman kontribusi dan rekomendasi jurnal penelitian terdahulu yang menjadi basis *State of the Art* dalam disertasi ini. Sintesis literatur tersebut dilakukan untuk memastikan bahwa pengembangan model klasifikasi ikan ini berjalan secara fokus dan memiliki urgensi penelitian yang jelas. Penelitian ini mengusulkan sebuah metodologi baru dalam klasifikasi citra ikan dengan mensinergikan algoritma segmentasi *U2-Net* dan arsitektur hibrida *Dual-Backbone* berbasis *EfficientNetB0* dan *InceptionV3*. Keaslian riset ini terletak pada implementasi "*Feature-Level Fusion*" yang secara otomatis mengintegrasikan representasi laten dari dua jalur saraf yang berbeda untuk mengatasi masalah ambiguitas visual pada spesies ikan. Pendekatan ini secara fundamental berbeda dari penelitian (Wyatt dkk., 2022, Akbar dkk., 2025) yang hanya menggunakan metode ensemble pada tingkat keputusan (*late fusion*), karena penelitian ini melakukan penggabungan pengetahuan sejak tahap ekstraksi fitur.

Penelitian ini menerapkan pra-pemrosesan cerdas menggunakan model *U2-Net* melalui library *RemBG* untuk memisahkan objek ikan dari latar belakang secara otomatis. Prosedur ini memberikan keunggulan kritis jika dibandingkan dengan penelitian (Montalbo & Hernandez, 2019, Dhar, 2021) yang masih memproses citra dengan latar belakang utuh atau hanya menggunakan filter *noise* standar. Dengan menghilangkan gangguan visual latar belakang, Subjek (model) dapat melakukan Predikat (ekstraksi fitur) terhadap Objek (biometrik ikan) secara lebih fokus, sehingga meningkatkan presisi pada spesies yang memiliki pola warna serupa dengan lingkungannya.

Strategi *Dual-Backbone* yang menggabungkan *EfficientNetB0* dan *InceptionV3* menciptakan sinergi ekstraksi fitur multi-skala yang tidak dimiliki oleh model tunggal pada penelitian (Yang, 2025, Yasin dkk., 2023, Beyza dkk., 2024). Peneliti memanfaatkan *EfficientNetB0* untuk menangkap fitur semantik yang mendalam melalui penskalaan jaringan yang optimal, sementara *InceptionV3* digunakan untuk menangkap detail morfologi dalam berbagai ukuran filter paralel. Kombinasi ini memungkinkan sistem untuk mengenali detail halus seperti

tekstur sisik sekaligus struktur anatomi ikan secara holistik, yang merupakan lompatan teknologi dibandingkan penggunaan fitur manual *Gist/GLCM* pada (Dhar, 2021).

Mekanisme fusi pada level fitur (*Concatenation Layer*) secara adaptif menyatukan vektor fitur dari kedua *backbone* untuk menghasilkan keputusan klasifikasi yang lebih tangguh. Penelitian ini tidak hanya memilih satu model terbaik, (Vrbanić & Member, 2020) melainkan memaksa kedua arsitektur untuk saling melengkapi dalam satu kesatuan jaringan saraf *end-to-end*. Inovasi ini secara otomatis meminimalkan nilai *loss* dan meningkatkan generalisasi model pada dataset yang luas. Secara keseluruhan, integrasi segmentasi cerdas dan fusi fitur hibrida ini sebagai kontribusi orisinal yang signifikan dan kompetitif di peta penelitian kecerdasan buatan global.

Tabel 2.1 Penelitian sebelumnya yang terkait

No	Referensi	Objek	Tujuan	Metode	Hasil
1.	(Balim dkk. 2025)	Ikan (<i>Gilthead Seabream</i>)	Mengklasifikasi ikan 3 tingkat kesegaran secara otomatis.	Integrasi fitur citra (<i>CNN ResNet50</i>) dan data reflektansi laser menggunakan <i>Feature Fusion</i> .	Akurasi: 98,41%; F1- <i>Score</i> : 98,41%; <i>Precision</i> : 98,45%.
2.	(Charmel ia dkk. 2022)	Ikan nila, khususnya bagian citra insang dan kepala	Untuk mengetahui tingkat akurasi identifikasi kualitas kesegaran ikan	Ekstraksi Fitur dengan <i>HSV</i> menggunakan metode <i>K-Nearest Neighbor (KNN)</i>	nilai akurasi sebesar 94,09% pada tahap pelatihan dan 90% pada tahap pengujian.
3.	(Aloysia Malirmas ele, 2024)	Citra mata ikan yang berjumlah 1.652 dataset	Mengklasifikasi kesegaran ikan berdasarkan citra mata melalui sistem berbasis web	Menggunakan model <i>Deep Learning VGG-16</i> .	Akurasi : 97,14% secara umum dan mencapai : 98,68% ketika diuji dengan data validasi.

Tabel 2.1 Penelitian sebelumnya yang terkait (lanjutan)

No	Referensi	Objek	Tujuan	Metode	Hasil
4.	(Córdova dkk. 2025)	Spesies Ikan Umum	Mengenali spesies baru di luar dataset latih (<i>Open Set</i>).	<i>Open-Set Recognition</i> (OSR) menggunakan <i>Extreme Value Theory</i> (EVT) dan <i>Thresholding Softmax</i> .	<i>Macro F1-Score</i> : 0,94; Mengurangi kesalahan klasifikasi spesies asing.
5.	(Yang dkk. 2025)	Citra Umum (<i>CIFAR-10</i>)	Mencegah <i>overfitting</i> pada dataset kecil.	CNN 6 lapis konvolusi dengan <i>Batch Normalization</i> , <i>Dropout</i> (0.5), dan <i>Adam Optimizer</i> .	Akurasi: 84,95%; <i>Test Loss</i> : 0,4537; <i>Train Loss</i> : 0,2140.
6.	(Yasin dkk. 2023)	Kesegaran Ikan	Membandingkan performa <i>AI</i> konvensional dan <i>Deep Learning</i> .	Ekstraksi fitur menggunakan <i>MobileNetV2</i> yang dikombinasikan dengan klasifikasi <i>SVM</i> .	Akurasi: 98,33%; <i>F1-Score</i> : 0,98; <i>Sensitivity</i> : 0,98.
7.	(Vrbančič & Podgorelec 2020)	Citra & Medis Umum	Mengoptimasi pemilihan lapisan untuk <i>transfer learning</i> .	Metode <i>Differential Evolution based Fine-Tuning (DEFT)</i> untuk mencari parameter lapisan terbaik.	Akurasi: 98,45%; Penurunan <i>Error Rate</i> yang signifikan dibanding <i>Fine-Tuning</i> standar.
8.	(Akbar dkk. 2025)	Daging Sapi & Babi	Otentikasi pangan dengan interpretasi visual.	Fitur <i>EfficientNetB0</i> diolah dengan <i>Ensemble Soft Voting</i> (RLDA, SVM, dan <i>Random Forest</i>).	Akurasi: 99,0%; Interpretasi wilayah fitur menggunakan <i>Grad-CAM</i> .

Tabel 2.1 Penelitian sebelumnya yang terkait (lanjutan)

No	Referensi	Objek	Tujuan	Metode	Hasil
9.	(Montalbo & Hernandez 2019)	Spesies Ikan (Pulau Verde)	Klasifikasi spesies pada kondisi data minim.	<i>Fine-tuning VGG16</i> dengan teknik augmentasi (rotasi, <i>shear</i> , <i>zoom</i>) pada 1.500+ citra.	Akurasi: 99,0%; <i>Loss</i> : 0,08 (sangat rendah menunjukkan model stabil).
10.	(Liu 2018)	Digit MNIST	Analisis teknis ekstraksi fitur hierarkis.	Arsitektur <i>CNN</i> stinir dengan 3 lapis (Konvolusi, <i>Pooling</i> , dan <i>Fully Connected</i>).	Ekstraksi pola visual efektif secara otomatis; Minimalisir <i>Error</i> pada digit tangan.
11.	(Yildiz dkk. 2024)	Mata Ikan	Deteksi kesegaran melalui kamera <i>smartphone</i> .	Ekstraksi fitur VGG19 dan klasifikasi menggunakan <i>Artificial Neural Network</i> (ANN) MLP.	Akurasi: 77,3%; Implementasi aplikasi Android berjalan <i>real-time</i> .
12.	(Dhar & Guha 2021)	Citra Ikan (QUT & F4K)	Klasifikasi spesies dengan fitur tekstur hibrida.	Gabungan fitur global (<i>Gist</i>) dan fitur lokal (<i>GLCM</i>) yang dilatih dengan model <i>XgBoost</i> .	Akurasi: 98,08% (F4K); <i>F1-Score</i> rata-rata di atas 0,91.
13	(Farooq dkk., 2026)	Spesies ikan tropis (Pakistan)	Kerangka kerja multi-kelas untuk klasifikasi spesies ikan menggunakan teknik <i>deep learning</i>	YOLOv4	Akurasi Klasifikasi: 99% (mAP: 99,65%). Hasil ini menunjukkan performa tinggi dalam membedakan antar spesies ikan yang berbeda secara otomatis.

Tabel 2.1 Penelitian sebelumnya yang terkait (lanjutan)

No	Referensi	Objek	Tujuan	Metode	Hasil
14	(Suwarsito dkk., 2022)	10 jenis ikan air tawar Indonesia	Klasifikasi ikan air tawar berdasarkan representasi citra menggunakan metode <i>K-Nearest Neighbor</i>	K-Nearest Neighbor (KNN)	Akurasi Klasifikasi: 70%. Akurasi ini merujuk pada kemampuan sistem dalam mengelompokkan citra ikan ke dalam 10 label spesies yang ditentukan.
15	Prasetyo et al. (2022)	Dataset <i>Fish-gres & Fish4-Knowledge</i>	Jaringan residual multi-level <i>VGGNet</i> untuk klasifikasi spesies ikan	MLR-VGGNet	Akurasi Klasifikasi: 99,69%. Fokus pada peningkatan akurasi pengenalan spesies dengan menangkap fitur detail (tekstur) dan fitur global (bentuk).
16	(Knausgård dkk., 2022)	Ikan di lingkungan alami (<i>temperate fish</i>)	Deteksi dan klasifikasi ikan beriklim sedang: pendekatan berbasis <i>deep learning</i>	<i>YOLO & CNN (SE-Net)</i>	Akurasi Gabungan: Berhasil melakukan Deteksi keberadaan ikan dan dilanjutkan dengan Klasifikasi spesies dengan akurasi 99,27% (pada data uji terkontrol).

Tabel 2.1 Penelitian sebelumnya yang terkait (lanjutan)

No	Referensi	Objek	Tujuan	Metode	Hasil
17	(Senthil Pandi S & Thanam A,2024)	Dataset <i>Fish-gres & Fish4-Knowledge</i>	Mengoptimalkan klasifikasi ikan dengan model hibrida <i>SAE-SVMH</i>	Hibrida <i>SAE-SVM</i>	Akurasi Klasifikasi: Sangat Tinggi. Fokus pada akurasi pengenalan jenis spesies dengan meminimalkan kesalahan (<i>loss</i>) saat ekstraksi fitur otomatis.
18	(Alsmadi & Almarashdeh, 2022)	Studi literatur berbagai algoritma	Survei tentang teknik-teknik klasifikasi ikan	Survei Literatur	Merangkum berbagai hasil penelitian di mana fokus utamanya adalah efektivitas algoritma dalam mengidentifikasi spesies ikan.
19	(Banwari dkk., 2022)	Ikan Labeo rohita (Rohu)	<i>teknik visi komputer untuk estimasi kesegaran dari segmen mata ikan</i>	Ekstraksi fitur warna dan tekstur dari ROI mata yang disegmentasi.	Akurasi Estimasi: 96,67%. Sangat efisien untuk penggunaan dunia nyata di industri
20	Madhubhaskanda dkk., 2023)	Berbagai spesies (Tuna, Sea Bream, dsb.)	Mengkaji penggunaan visi komputer, hidung elektronik (e-nose), dan lidah elektronik (e-tongue).	Tinjauan Literatur (Survei) terhadap pelbagai kaedah prapemrosesan dan algoritma.	Menunjukkan visi komputer sebagai kaedah tidak merusak yang sangat berpotensi

Tabel 2.1 Penelitian sebelumnya yang terkait (lanjutan)

No	Referensi	Objek	Tujuan	Metode	Hasil
21	(Choompol dkk., 2025)	Ikan Silver Barb (Puntius gonionotus)	Mencari model Deep Learning terbaik untuk menjelaskan tahap kesegaran ikan	Perbandingan berbagai model CNN (seperti YOLOv8-v11) menggunakan kaedah TOPSIS.	Tugas Klasifikasi: menstandarkan dataset dan meningkatkan ketepatan melalui prapemrosesan citra
22	(Hoang dkk., 2026)	Ikan dalam dataset <i>Freshness of the Fish Eyes (FFE)</i>	Mengatasi subjektivitas penilaian manusia dengan pendekatan fitur warna dan tekstur yang digabungkan.	Gabungan fitur (<i>LBP, GLCM</i>) dengan pengelasan <i>LightGBM</i> dan <i>ANN</i> .	Akurasi Klasifikasi: 77,56% (<i>LightGBM</i>) dan meningkat sehingga 97,49% (<i>ANN</i>) dengan augmentasi data.
23	(Saleh dkk., 2022)	ikan di habitat bawah air (Liar)	Memberikan pemahaman tentang evolusi algoritma klasifikasi ikan dari tahun 2003 hingga 2021.	injauan Literatur Komprehensif (Survei) mengenai <i>Deep Learning</i> (DL).	Hasil Ulasan: DL mampu mengekstrak ciri tinggi secara otomatis tanpa campurtangan manusia yang banyak.
24	(Chen dkk., 2017)	Ikan di alam liar (Dataset <i>Kaggle</i>)	Mengklasifikasi kan spesies ikan dalam keadaan persekitaran yang tidak terkawal dan pencahayaan berubah	Sistem dua cabang: Cabang Deteksi/Alignment dan Cabang Informasi Konteks.	Tugas Deteksi & Klasifikasi: Berjaya mencapai kedudukan 0,7% teratas dalam pertandingan global <i>Kaggle</i> .

Peneliti terdahulu, seperti (Akgül dkk., 2023, Janakidevi dkk., 2021), telah memelopori penggunaan *transfer learning* melalui arsitektur *VGG-16* untuk mendeteksi kesegaran ikan berdasarkan karakteristik warna mata dan insang.

Sejalan dengan itu, (Kaladevi dkk., 2021, Yasin dkk., 2023) memperkuat literatur dengan membuktikan bahwa jaringan saraf konvolusi dalam mampu memberikan hasil klasifikasi yang jauh lebih akurat dibandingkan metode tradisional dalam mengidentifikasi tingkat degradasi mutu ikan sarden. Di sisi lain, pendekatan berbasis fitur klasik tetap dikembangkan oleh (Dhar & Guha, 2021) melalui kombinasi fitur *Gist* dan *GLCM* yang diproses menggunakan algoritma *XGBoost*. Meskipun metode-metode ini telah mencapai akurasi yang menjanjikan, mayoritas studi tersebut masih memiliki ketergantungan tinggi pada pra-pemrosesan manual atau model jalur tunggal yang rentan terhadap variasi data yang ekstrim dan gangguan latar belakang di lingkungan industri nyata.

Berbeda dengan metode *KNN* atau model tunggal *VGG-16* (Jerandu dkk. 2022, Malirmasele 2024), penelitian ini mampu mengidentifikasi spesies dan klasifikasi kesegaran secara simultan. Integrasi *U2-Net* untuk segmentasi otomatis menjamin akurasi tinggi sebesar 98,10% dengan meminimalisir gangguan latar belakang secara presisi. Penggunaan strategi *Asymmetric Fine-Tuning* dan ambang batas kepercayaan 95% memberikan hasil prediksi yang lebih tangguh dan akuntabel. Secara keseluruhan, disertasi ini menawarkan inovasi teknologi yang lebih komprehensif, otomatis, dan aplikatif untuk standarisasi kualitas perikanan modern.

Dalam konteks inilah, penelitian yang disusun menghadirkan sebuah paradigma baru yang melampaui capaian-capaian teknis pada seluruh literatur sebelumnya melalui konsep *Ensemble Deep Learning* yang jauh lebih kompleks dan kokoh. Keunggulan utama yang menjadi *State of the Art* dalam penelitian ini terletak pada penggunaan strategi *Feature-Level Fusion* yang secara cerdas menggabungkan dua model tulang punggung tangguh, yakni *EfficientNetB0* dan *InceptionV3*. Integrasi pada tingkat fitur ini memungkinkan sistem untuk menangkap pola-pola mikroskopis yang detail sekaligus mempertahankan efisiensi parameter yang tinggi, sebuah kombinasi strategis yang tidak ditemukan pada penelitian berbasis model tunggal atau sistem hibrida sederhana yang diulas dalam survei (García dkk., 2022).

Aspek yang paling krusial sebagai pembeda utama dalam penelitian ini adalah pengenalan mekanisme Optimasi *Dual-Backbone* Adaptif secara Otomatis. Jika penelitian terdahulu masih terbatas pada penyesuaian parameter secara statis atau manual, model yang bangun mampu melakukan adaptasi secara mandiri terhadap karakteristik citra melalui teknik *Asymmetric Fine-Tuning*, teknik ini memastikan bahwa setiap jalur jaringan saraf bekerja secara maksimal dalam mengekstrak representasi visual yang paling relevan untuk klasifikasi spesies sekaligus penilaian tingkat kesegaran secara simultan. Kemampuan adaptasi otomatis ini menjadikan model Ini jauh lebih cerdas dibandingkan model-model dalam referensi sebelumnya yang sering kali terbatas pada satu jenis ikan atau satu jenis parameter kualitas saja.

Secara keseluruhan, kontribusi ilmiah dari penelitian ini memberikan dampak transformatif bagi dunia akademik dan industri perikanan melalui pencapaian akurasi 98,10%. Penelitian ini tidak hanya sekadar mengikuti tren penggunaan kecerdasan buatan, tetapi telah berhasil menciptakan sebuah landasan baru dalam rekayasa fitur digital. Dengan menyatukan segmentasi otomatis, fusi dua jalur *backbone*, dan optimasi adaptif dalam satu sistem terpadu, penelitian ini memposisikan diri sebagai *State of the Art* terbaru yang menjawab tantangan kompleksitas data di lapangan yang belum mampu diatasi secara tuntas oleh penelitian-penelitian sebelumnya dalam satu dekade terakhir.

2.3 Landasan Teori

Landasan teori dalam penelitian ini disusun untuk membangun kerangka konseptual yang komprehensif, mulai dari prinsip dasar *artificial intelligence*, *computer vision*, dan *machine learning* hingga implementasi teknis *convolutional neural network (CNN)* yang menjadi pilar utama ekstraksi fitur otomatis. Penggunaan strategi *transfer learning* dan *fine-tuning* pada arsitektur hibrida *EfficientNetB0* dan *InceptionV3*, serta menjelaskan bagaimana mekanisme *ensemble deep learning* melalui fusi fitur dapat meningkatkan ketangguhan model. Selain itu, pemaparan mengenai segmentasi citra berbasis *U2-Net* dan augmentasi citra disertakan untuk memperkuat pemahaman mengenai proses

standarisasi data masukan, yang pada akhirnya akan dievaluasi secara objektif menggunakan *confusion matrix* guna mengukur akurasi dan performa sistem klasifikasi biometrik ikan yang diusulkan.

2.3.1 *Artificial Intelligence*

Artificial Intelligence adalah bidang ilmu komputer yang luas, fundamental, dan terus berevolusi, yang bertujuan untuk menciptakan sistem yang mampu meniru kemampuan kognitif manusia seperti belajar, bernalar, memahami, dan memecahkan masalah. Definisi ini, yang menjadi pondasi dasar, telah berkembang seiring waktu, bergeser dari fokus teoretis awal menuju pendekatan yang lebih pragmatis dan didorong oleh data. Bidang ini telah mengalami pergeseran paradigma signifikan (Stuart Russell, 1995). Di masa lalu, *AI* didominasi oleh rekayasa pengetahuan manual, di mana para ahli memprogram aturan-aturan eksplisit.

Namun, era modern ditandai dengan dominasi pembelajaran mesin (*machine learning*) dan pembelajaran mendalam (*deep learning*), yang memungkinkan sistem untuk belajar secara mandiri dari data dalam jumlah besar. Pergeseran ini juga mencakup redefinisi tujuan *AI*, alih-alih sekadar memaksimalkan fungsi utilitas yang telah ditentukan, *AI* kini dituntut untuk belajar apa yang harus dimaksimalkan, bahkan di tengah ketidakpastian mengenai tujuan manusia. Dalam praktiknya, evolusi ini telah melahirkan berbagai sub-bidang. Pemahaman Bahasa Alami (*Natural Language Understanding*), adalah salah satu area yang paling maju (Zixing Cai, Lijue Liu, Baifan Chen, 2021). Kemajuan dalam *NLU*, didorong oleh arsitektur revolusioner seperti *Transformer*, telah memungkinkan pengembangan model bahasa besar (*LLM*) (Tarumingkeng, 2024).

Model-model ini, dapat memproses, menafsirkan, dan menghasilkan bahasa manusia dengan cara yang belum pernah terjadi sebelumnya, membentuk fondasi untuk *AI* Generatif yang kini mendominasi inovasi (Brown dkk., 2020). Perkembangan ini juga didukung oleh teknik-teknik optimasi canggih, seperti *SGDR* (*Stochastic Gradient Descent with Warm Restarts*) yang membuat pelatihan model yang kompleks menjadi lebih efisien [30]. Meskipun kemajuan

teoretis dan teknis ini telah mempercepat adopsi *AI* di dunia nyata mengungkapkan bahwa implementasi praktis masih menghadapi tantangan.

Laporan tersebut mencatat bahwa *AI* di tingkat bisnis masih berada pada tahap yang relatif belum matang. Salah satu masalah utamanya adalah kurangnya hasil yang dapat direproduksi (*reproducible results*), yang merupakan prasyarat penting untuk kehandalan dan kepercayaan (Loukides, 2021). Selain itu, seiring dengan pesatnya pertumbuhan *AI*, isu-isu krusial terkait etika, keamanan, dan privasi sering kali dianggap sebagai pertimbangan sekunder, bukan prioritas utama dalam siklus pengembangan. Oleh karena itu, linisan teori *AI* saat ini tidak hanya mencakup arsitektur dan algoritma, tetapi juga tantangan untuk menjembatani kesenjangan antara kemampuan teoretis *AI* dengan implementasi yang bertanggung jawab, inil, dan etis di dunia nyata.

2.3.2 Computer Vision

Visi komputer adalah salah satu cabang penting dari kecerdasan buatan (*AI*) yang bertujuan untuk memungkinkan mesin "melihat" dan memahami informasi dari citra digital dan video. Bidang ini memiliki landasan teoretis yang kuat, terutama pada algoritma numerik (Solomon, 2015). Algoritma ini mencakup metode dari aljabar linear, seperti eliminasi *Gauss*, dekomposisi nilai singular, dan berbagai teknik faktorisasi matriks. Algoritma numerik ini berfungsi sebagai tulang punggung matematis untuk menyelesaikan sistem persamaan linear yang mendasari berbagai tugas visual, seperti kalibrasi kamera, estimasi *pose*, dan rekonstruksi *3D*. Meskipun efektif, pendekatan klasik ini sering kali memerlukan rekayasa fitur manual yang kompleks dan memakan waktu.

Tetapi, visi komputer mengalami transformasi revolusioner dengan munculnya pembelajaran mendalam (Solomon, 2015). Model *deep learning*, khususnya jaringan saraf tiruan konvolusional, secara fundamental mengubah cara mesin memproses data visual. Berbeda dengan metode tradisional, *CNN* memiliki kemampuan untuk secara otomatis mempelajari dan mengekstraksi fitur visual yang relevan dari data, dari yang sederhana seperti tepi dan warna, hingga yang lebih kompleks seperti bentuk dan tekstur. Kemampuan ini secara signifikan

mengurangi kebutuhan akan intervensi manual dan meningkatkan akurasi secara drastis.

Katalisator utama di balik kesuksesan pembelajaran mendalam pada visi komputer adalah konvergensi dengan era analisis big data bagaimana ketersediaan dataset citra yang masif menjadi prasyarat penting (Jan, 2019). Data dalam jumlah besar memungkinkan model *CNN* untuk dilatih secara ekstensif, sehingga mereka dapat mengenali pola yang sangat kompleks dan menghasilkan kemampuan generalisasi yang luar biasa. Visi komputer modern merupakan hasil dari sinergi antara prinsip-prinsip algoritma numerik yang solid, kekuatan pemrosesan fitur otomatis dari *deep learning*, dan ketersediaan data dalam skala besar. Visi komputer tidak hanya berfokus pada pengenalan objek, tetapi juga pada pemahaman yang lebih dalam tentang adegan (*scene understanding*). Misalnya, arsitektur seperti *Vision Transformer (ViT)*, yang muncul lebih baru, menunjukkan bahwa pendekatan yang awalnya dirancang untuk Pemrosesan Bahasa Alami (*NLP*) juga dapat diterapkan pada tugas-tugas visual.

ViT memecah gambar menjadi patch kecil dan memprosesnya secara paralel menggunakan mekanisme *self-attention*, yang menawarkan cara alternatif dan efektif untuk menangani data visual tanpa bergantung pada arsitektur konvolusional tradisional. Visi komputer saat ini adalah bidang yang dinamis, terus berkembang dari fondasi algoritma numerik menuju paradigma *deep learning* yang kuat, dan kini mulai mengintegrasikan arsitektur inovatif dari bidang *AI* lainnya untuk mencapai pemahaman visual yang lebih canggih.

2.3.3 Machine Learning

Machine learning adalah sub-bidang kecerdasan buatan yang mengembangkan algoritma yang dapat belajar dari data dan membuat keputusan, alih-alih diprogram secara eksplisit (Raschka, 2024). Dalam pengolahan citra, *machine learning* berperan penting karena memungkinkan komputer untuk "memahami" konten visual. Model-model ini dilatih untuk mengidentifikasi pola, objek, dan fitur dalam gambar atau video, jauh melampaui kemampuan manipulasi piksel konvensional (Elwirehardja dan Pardamean, 2023). Prosesnya

secara umum melibatkan dua tahap, pelatihan model menggunakan data historis, lalu prediksi pada data baru.

Pengolahan citra modern sangat didominasi oleh *deep learning*, yang merupakan bagian dari *machine learning* yang menggunakan jaringan saraf tiruan dengan banyak lapisan untuk memproses data. Arsitektur berlapis ini memungkinkan model untuk secara otomatis mengekstrak fitur penting dari citra, sebuah proses yang secara tradisional dilakukan secara manual. Jaringan yang dalam ini sangat efektif untuk data yang kompleks seperti citra, suara, dan teks (TV Geetha, 2023). Pendekatan baru dengan arsitektur *Transformer* (Tarumingkeng, 2024), yang awalnya dikembangkan untuk pemrosesan bahasa alami. Model seperti *Vision Transformer* (Zhai dkk., 2021), mengadaptasi arsitektur ini untuk pengolahan citra. *ViT* memecah citra menjadi bagian-bagian kecil dan memprosesnya secara sekuensial, menunjukkan bahwa pendekatan ini bisa menjadi alternatif yang kuat untuk *CNN*.

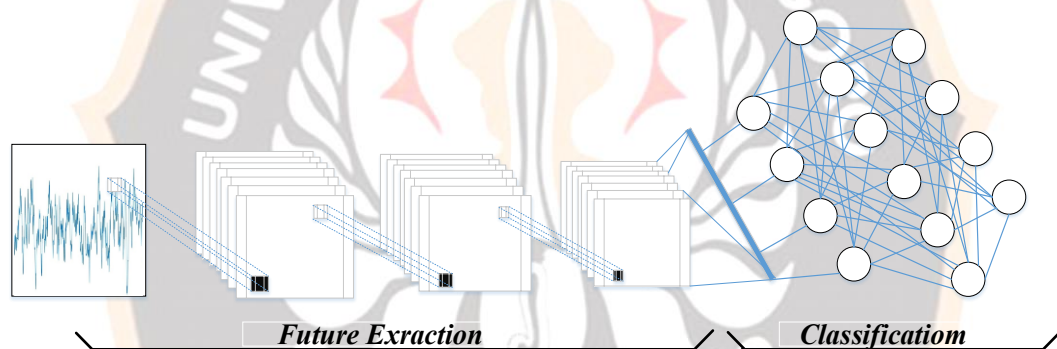
Beberapa tantangan utama yang dihadapi dalam pengolahan citra *ML* adalah *Overfitting*, model terlalu spesifik terhadap data pelatihan dan tidak dapat bergeneralisasi dengan baik pada data baru. *Data Representation*, tantangan untuk merepresentasikan citra dengan cara yang paling informatif agar model dapat memahaminya secara efektif. Dengan menekankan pentingnya kualitas data dan prapemrosesan, seperti augmentasi data, untuk meningkatkan akurasi (Fadhilah, 2024). Interpretasi model dapat memahami dan membuat keputusan tertentu, terutama pada model *deep learning* yang kompleks, masih menjadi tantangan yang dikenal sebagai masalah "kotak hitam" (Raschka, 2024).

2.3.4 *Convolutional Neural Network*

Convolutional Neural Network (CNN) adalah jenis jaringan saraf tiruan yang khusus dirancang untuk memproses data yang memiliki struktur grid seperti citra. *CNN* telah menjadi tulang punggung dari banyak aplikasi *AI* modern, mulai dari pengenalan citra hingga deteksi objek. *CNN* terdiri dari beberapa lapisan, masing-masing memiliki peran khusus dalam mengekstraksi dan mengabstraksikan fitur dari data input. Dikembangkan pada tahun 1990 *convolutional neural network*

adalah algoritma *deep learning* yang pertama kali diperkenalkan melalui publikasi "*Gradient-Based Learning Applied to Document Recognition*" (Lecun dkk., 2023). *CNN* merupakan evolusi dari *Multilayer Perceptron (MLP)* yang secara spesifik diadaptasi untuk pengolahan data dua dimensi. Karena memiliki kedalaman jaringan yang tinggi,

CNN diklasifikasikan sebagai *deep neural network* dan sering diaplikasikan dalam tugas-tugas yang melibatkan data citra. Gambar 2.1 menunjukkan arsitektur umum dari *convolutional neural network* yang merupakan model *deep learning* yang dominan digunakan untuk tugas pemrosesan citra. *CNN* berfungsi mengekstraksi fitur hierarkis dari data masukan citra melalui serangkaian lapisan



Gambar 2.1 Arsitektur CNN

Berikut adalah penjelasan tentang lapisan utama dalam *CNN* :

a. Lapisan Konvolusi (*Convolutional Layer*)

Lapisan konvolusi adalah inti dari *CNN*. Pada lapisan ini, filter (juga disebut kernel) berukuran kecil melintasi data input dan menghasilkan peta fitur.

$$(I * K)(i, j) = \sum_m \sum_n I(i + m, j + n) K(m, n) \quad (2.1)$$

I : Merupakan input image yang akan dikenakan operasi konvolusi. I adalah sebuah matriks yang merepresentasikan intensitas dari citra

K : Kernel atau filter yang akan digunakan untuk melakukan operasi konvolusi pada citra. K merupakan sebuah matriks yang biasanya berukuran lebih kecil dari citra input

$(I * K)(i, j)$: Hasil dari operasi konvolusi antara citra input I dan kernel K pada posisi piksel (i, j) citra output setelah operasi konvolusi dilakukan.

$\sum_m \sum_n$: Notasi penjumlahan ini yang menunjukkan semua hasil perkalian elemen elemen dari kernel K yang dijumlahkan dengan elemen elemen yang sesuai dari citra input I . Indeks m dan n adalah indeks iterasi elemen elemen dari kernel.

$I(i + m, j + n)$: Nilai piksel dari citra input I pada posisi $(i + m, j + n)$. Posisi ini bergeser sesuai dengan posisi m dan n dari kernel K .

$K(m, n)$: Nilai elemen dari kernel K pada posisi (m, n) . Kernel K diterapkan pada citra input I dengan menggeser posisi sesuai dengan indeks m dan n

b. Lapisan Aktivasi (*Activation Layer*)

Setelah lapisan konvolusi, lapisan aktivasi biasanya diterapkan untuk menambahkan non-linearitas ke dalam model. Salah satu fungsi aktivasi yang umum digunakan adalah *ReLU* (*Rectified Linear Unit*).

$$f(x) = \max(0, x) \quad (2.2)$$

$f(x)$: Fungsi aktivasi yang dihasilkan. Output dari fungsi *ReLU*

$\max(0, x)$: Operasi yang dilakukan oleh fungsi *ReLU*. Operasi ini memilih nilai maksimum antara 0 dan x . Artinya jika x lebih besar dari 0, maka $f(x) = x$, jika x kurang dari atau sama dengan 0, maka $f(x) = 0$

Input x : Nilai yang masuk ke neuron sebelum diterapkan fungsi aktivasi *ReLU*

Output $f(x)$: Output dari fungsi aktivasi. Setelah menerapkan fungsi *ReLU*, ini adalah nilai yang dihasilkan dan dikirim ke lapisan berikutnya dalam jaringan saraf

c. Lapisan *Pooling* (*Pooling Layer*)

Lapisan pooling mengurangi dimensi spasial dari peta fitur dan membantu dalam komputasi yang lebih efisien serta menangani *overfitting*. *Max pooling* adalah teknik *pooling* yang sering digunakan.

Persamaan *Max Pooling* :

Jika memiliki matriks input I dan ukuran pooling P :

$$(I_{pool})(i, j) = \max \{I(m, n) | m, n \in P(i, j)\} \quad (2.3)$$

$(I_{pool})(i, j)$: nilai output dari operasi max pooling pada koordinat (i, j) (I, j) (i, j) dalam hasil pooling (pooled output).

$\max \{I(m, n) | m, n \in P(i, j)\}$: Operasi max pooling untuk mencari nilai maksimum $I(m, n)$ dari sekumpulan nilai yang ditentukan oleh m dan n dari jendela polling $P(i, j)$

$I(m, n)$: Nilai input pada koordinat (m, n) dari input asli (biasanya dari citra atau hasil dari lapisan konvolusional sebelumnya)

$m, n \in P(i, j)$: m dan n adalah indeks yang berada dalam jendela polling $P(i, j)$ biasanya merupakan submatriks dari input yang ditentukan oleh ukuran pooling (contoh : 2X2 atau 3X3)

d. Lapisan *Fully Connected* (*Fully Connected Layer*)

Lapisan ini mirip dengan jaringan saraf biasa di mana semua neuron dari lapisan sebelumnya terhubung dengan *neuron* di lapisan ini. Lapisan ini digunakan untuk menggabungkan fitur yang telah diekstraksi dan melakukan klasifikasi.

$$Y = W \cdot x + b \quad (2.4)$$

Y: Output dari lapisan *fully connected*. Biasanya, ini adalah vektor yang berisi nilai-nilai yang dihasilkan setelah input diproses melalui lapisan tersebut.

W: Matriks bobot (*weight matrix*). Bobot adalah parameter yang dipelajari selama pelatihan model. Matriks W mengandung bobot-bobot yang menghubungkan setiap neuron di lapisan input dengan setiap neuron di lapisan output. Jika ada n neuron di lapisan input dan m neuron di lapisan output, maka dimensi W adalah $m \times n$

x: Vektor input untuk lapisan *fully connected*. Vektor ini mengandung nilai-nilai yang diterima oleh lapisan tersebut. Dimensi x harus sesuai dengan jumlah neuron di lapisan input.

b: Vektor bias (*bias vector*). Bias adalah parameter tambahan yang juga dipelajari selama pelatihan model. Bias ditambahkan ke hasil perkalian matriks bobot dan vektor input untuk setiap neuron di lapisan output. Dimensi b biasanya sama dengan jumlah neuron di lapisan output.

e. *Backward Propagation (Backpropagation)*

Untuk melatih CNN, digunakan algoritma *backpropagation* yang mengoptimalkan bobot melalui penurunan gradien (*gradient descent*)

$$W_{new} = W_{old} - \eta \cdot \nabla L(W) \quad (2.5)$$

W_{new} : Matriks bobot baru setelah pembaruan. Ini adalah nilai bobot yang akan digunakan untuk iterasi berikutnya atau *epoch* berikutnya dalam pelatihan.

W_{old} : Matriks bobot lama sebelum pembaruan. Ini adalah nilai bobot yang digunakan pada iterasi atau *epoch* saat ini dalam pelatihan.

η (eta): *Learning rate* (laju pembelajaran). Ini adalah faktor skala yang menentukan seberapa besar langkah pembaruan bobot. *Learning rate* adalah parameter hiper yang dikonfigurasi oleh pengguna dan sangat mempengaruhi konvergensi pelatihan.

$\nabla L(W)$: Gradien dari fungsi *loss* (kerugian) L terhadap bobot W . Gradien ini adalah vektor atau matriks yang mengandung turunan parsial dari fungsi *loss* dengan *respect to* (terhadap) setiap elemen bobot W . Gradien menunjukkan arah dan seberapa besar perubahan yang diperlukan untuk meminimalkan fungsi *loss*.

2.3.5 *Transfer Learning dan Fine-Tuning*

Pembelajaran mendalam terutama untuk tugas-tugas visi komputer, salah satu tantangan paling fundamental adalah kebutuhan akan dataset yang besar dan terlabel secara akurat untuk melatih model dari awal. Masalah ini menjadi semakin krusial di domain-domain spesifik, seperti klasifikasi citra di lingkungan laut yang dinamis (Wyatt dkk., 2022) atau identifikasi citra medis (Yasin dkk., 2023), di mana pengumpulan data masif sulit dilakukan. Untuk mengatasi keterbatasan ini, *transfer learning* hadir sebagai sebuah paradigma metodologi yang sangat efisien. *Transfer learning* adalah teknik yang memanfaatkan pengetahuan yang telah diperoleh oleh model yang dilatih pada satu tugas (*source task*) dan mengaplikasikannya untuk menyelesaikan tugas lain yang terkait (*target task*) (Yasin dkk., 2023).

Dengan metode ini, kebutuhan untuk melatih model dari nol dapat dihindari, sebuah proses yang tidak hanya membutuhkan data dalam jumlah besar tetapi juga sumber daya komputasi yang intensif. Inti dari *transfer learning* adalah proses *fine-tuning*. Proses ini dimulai dengan menggunakan arsitektur model yang sudah terlatih, seperti *EfficientNetB0* atau *InceptionV3*, yang telah belajar untuk mengenali fitur visual dasar dari miliaran gambar. Bobot dari lapisan-lapisan awal model ini, yang bertanggung jawab untuk mendeteksi fitur-fitur umum seperti tepi dan tekstur, sering kali dibekukan (*frozen*). Pembekuan ini bertujuan untuk mempertahankan pengetahuan dasar yang telah dipelajari dari dataset yang luas, karena fitur-fitur tersebut dianggap relevan untuk tugas baru.

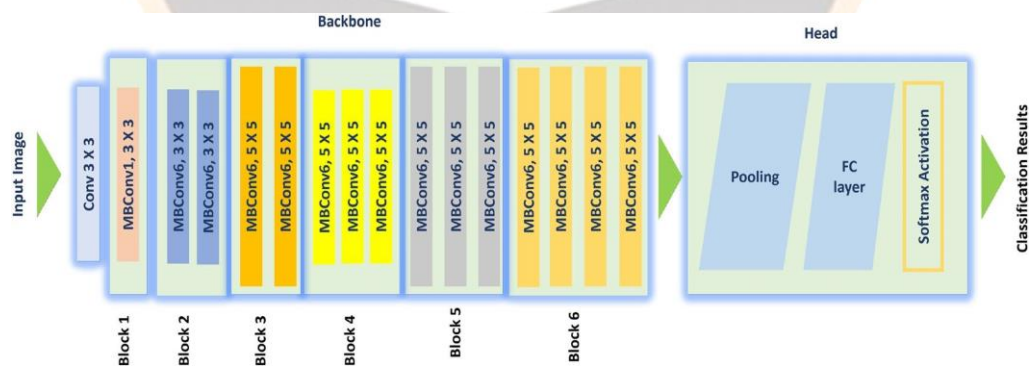
Sementara itu, lapisan-lapisan akhir model, yang bertanggung jawab untuk klasifikasi, dilatih ulang (*unfreeze*) atau diganti dengan lapisan baru yang disesuaikan dengan dataset target yang lebih kecil dan spesifik. Meskipun konsepnya kuat, proses *fine-tuning* tidak luput dari tantangan. Salah satu masalah utamanya adalah menentukan lapisan mana yang paling optimal untuk dilakukan *fine-tune* agar mencapai kinerja terbaik. Pemilihan ini sering kali bersifat manual dan memerlukan eksperimen yang memakan waktu. Guna mengatasi hal ini, penelitian telah mengembangkan metode yang lebih canggih, seperti *Differential Evolution based Fine-Tuning (DEFT)*, yang menggunakan algoritma meta-heuristik untuk secara otomatis memilih kombinasi lapisan yang paling cocok untuk diubah bobotnya. Pendekatan ini membuat proses *fine-tuning* menjadi lebih efisien dan tidak terlalu bergantung pada intuisi peneliti.

Selain itu, dalam proses *fine-tuning*, risiko *overfitting* tetap menjadi perhatian serius, terutama ketika dataset target berukuran kecil. Oleh karena itu, teknik regularisasi seperti *dropout* menjadi sangat krusial (Alnagashi dkk., 2024). *Dropout* secara acak menonaktifkan sebagian *neuron* selama pelatihan, yang memaksa model untuk tidak terlalu bergantung pada neuron tertentu dan meningkatkan kemampuan generalisasi. Lebih lanjut, kinerja model yang di *fine-tune* dapat ditingkatkan dengan menggunakan metode *ensemble*, di mana beberapa model yang dilatih digabungkan untuk menghasilkan prediksi yang lebih robust dan akurat. Pendekatan ini sangat berguna dalam lingkungan yang

menantang, seperti klasifikasi citra di lingkungan laut, di mana kondisi visual dapat sangat bervariasi (Wyatt dkk., 2022).

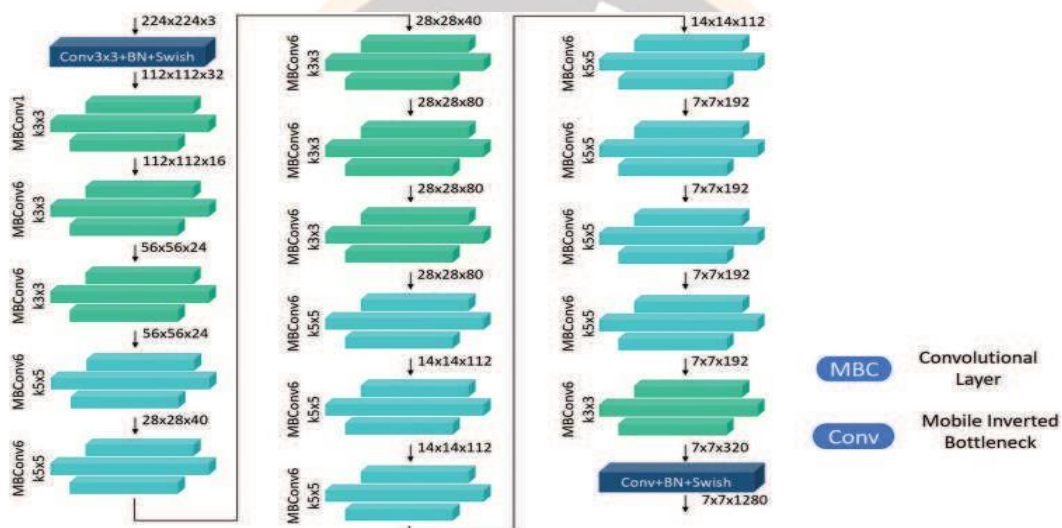
2.3.6 Arsitektur *EfficientNetB0* dan *InceptionV3*

EfficientNetB0 merupakan model dasar yang menggunakan metode penskalaan senyawa untuk mengoptimalkan kinerja jaringan saraf. Secara umum, arsitektur ini menyeimbangkan dimensi kedalaman, lebar, dan resolusi gambar secara bersamaan melalui koefisien tetap. Hal ini membuat *EfficientNetB0* menjadi jaringan yang sangat efisien karena model tersebut menghasilkan akurasi tinggi dengan jumlah parameter yang sangat minimal. Struktur utama *EfficientNetB0* mengandalkan blok *Mobile Inverted Bottleneck Convolution* atau *MBConv* (Ali dkk., 2025). Model ini menerapkan mekanisme perhatian *Squeeze and Excitation* untuk memfokuskan ekstraksi pada fitur yang paling penting (Reports, 2025). Sementara itu penelitian lain membuktikan bahwa *EfficientNetB0* memiliki kemampuan generalisasi yang sangat kuat dalam mengklasifikasikan berbagai pola penyakit tanaman (Access, 2025). Gambar 2.2 merupakan visualisasi perbandingan metode penskalaan tunggal dibandingkan dengan penskalaan senyawa (*compound scaling*) yang menjadi inti kekuatan *EfficientNetB0*.



Gambar 2.2. Arsitektur *EfficientNet-B0* dengan *MBConv* sebagai blok bangunan dasar (Ali dkk., 2025).

Penggunaan bobot awal dari *ImageNet* pada *EfficientNetB0* mempercepat proses konvergensi saat pelatihan berlangsung (Kaushik, 2024). Model ini tetap memberikan performa yang stabil meskipun menangani citra medis yang memiliki variasi tekstur sangat kompleks (Salian & Sawarkar, 2022). Gambar 2.3 merupakan diagram arsitektur global *EfficientNetB0*.

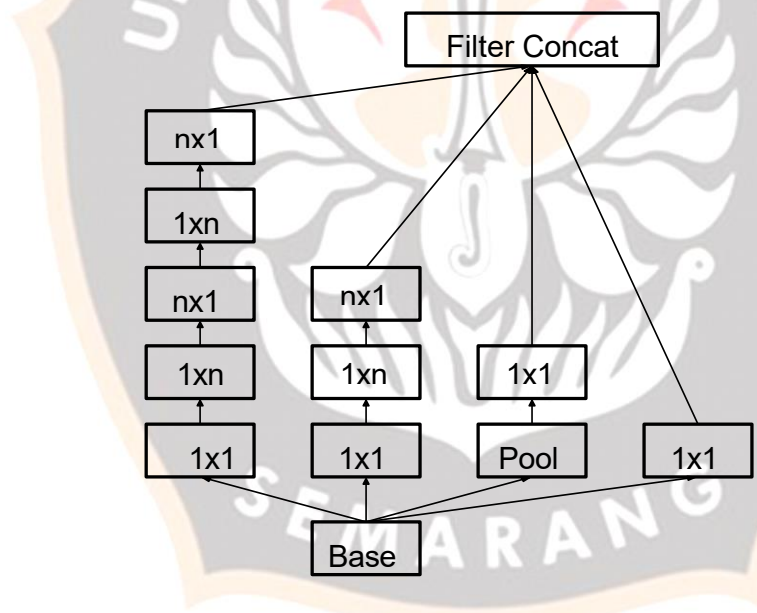


Gambar 2.3 Arsitektur *EfficientNetB0* Global (Hussain dkk., 2023).

Kesimpulan mengenai arsitektur ini menunjukkan bahwa keunggulan utama *EfficientNetB0* terletak pada efisiensi penggunaan parameter dan kecepatan proses inferensi yang sangat tinggi sehingga model tersebut sangat ideal untuk perangkat dengan sumber daya terbatas. Namun, arsitektur ini memiliki kelemahan pada kebutuhan pengaturan *hyperparameter* yang sangat teliti agar metode penskalaannya tidak menyebabkan penggunaan memori yang tidak terduga pada resolusi gambar yang sangat besar. Selain itu, beberapa penelitian menyebutkan bahwa waktu pelatihan terkadang menjadi lebih lama jika perangkat keras tidak memiliki optimasi khusus untuk operasi konvolusi kedalaman yang menjadi inti jaringan ini.

InceptionV3 adalah jaringan saraf dalam yang menggunakan jalur filter paralel untuk menangkap fitur pada berbagai skala gambar. Secara umum, prinsip utama *InceptionV3* adalah memecah konvolusi berukuran besar menjadi rangkaian

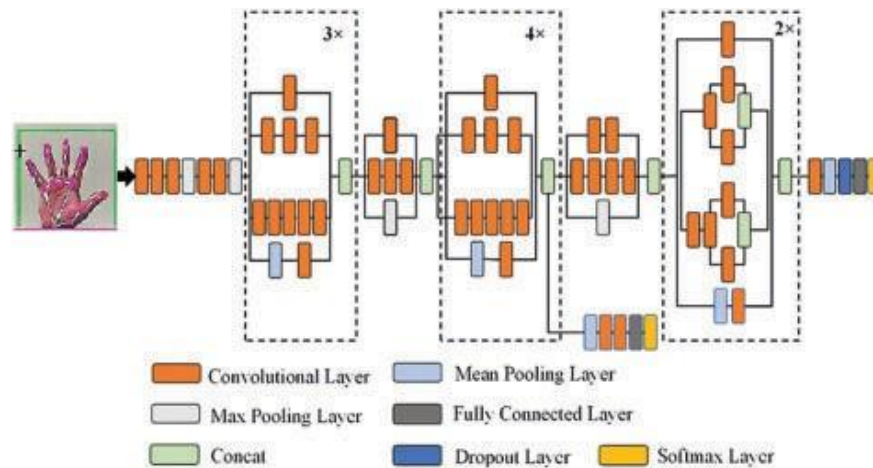
konvolusi yang lebih kecil atau asimetris. Pendekatan ini bertujuan untuk mengurangi beban parameter tanpa menghilangkan detail informasi spasial yang penting di dalam citra. Christian Szegedy memaparkan bahwa penggunaan teknik faktorisasi konvolusi mampu menurunkan biaya komputasi secara signifikan. Dalam penerapan praktis, Adnan Hussain memanfaatkan fleksibilitas *InceptionV3* untuk mengenali gestur tangan yang memiliki detail posisi jari sangat rumit. Arshleen Kaur menekankan bahwa arsitektur ini sangat handal dalam mengenali pola tekstur bawah laut yang tidak teratur karena model tersebut memiliki jalur paralel yang memproses fitur mikro dan makro secara simultan. Gambar 2.4 merupakan ilustrasi faktorisasi konvolusi di mana filter besar dipecah menjadi beberapa filter kecil untuk mengoptimalkan efisiensi parameter.



Gambar 2.4 Modul *Inception* setelah faktorisasi konvolusi $n \times n$. Dalam arsitektur (Szegedy et al., 2014)

Penggunaan teknik label *smoothing* pada *InceptionV3* membantu model mencapai akurasi yang konsisten di tengah gangguan visual (Kaur dkk., 2024). *InceptionV3* tetap menjadi pilihan yang sangat kuat karena model tersebut mampu menangani variasi pencahayaan dan perbedaan skala objek yang ekstrem secara efektif (Kaushik, 2024). Gambar 2.5 merupakan struktur alur kerja (*workflow*) global

InceptionV3 yang menunjukkan transisi data dari lapisan input, melewati blok-blok Inception, hingga mencapai pengklasifikasi akhir.



Gambar 2.5 Arsitektur model *InceptionV3* (Hussain dkk., 2023)

Kesimpulan dari arsitektur ini menunjukkan bahwa kelebihan utama *InceptionV3* berada pada kemampuannya dalam mengekstraksi fitur multi-skala yang sangat kaya dan stabilitasnya yang tinggi saat menghadapi objek dengan ukuran yang sangat beragam. Akan tetapi, *InceptionV3* memiliki kelemahan pada struktur jaringannya yang sangat kompleks dan jumlah parameter yang jauh lebih besar dibandingkan *EfficientNetB0*. Hal tersebut mengakibatkan penggunaan memori yang lebih boros dan kebutuhan daya komputasi yang lebih besar sehingga proses pemrosesan data cenderung lebih lambat pada perangkat keras standar.

2.3.7 Ensemble Deep Learning

Dalam domain pengolahan citra, model *deep learning* tunggal, meskipun memiliki performa yang kuat, sering kali memiliki keterbatasan. Keterbatasan ini mencakup kerentanan terhadap variasi data, ketidakstabilan dalam prediksi, dan potensi *overfitting* pada dataset yang kompleks. Untuk mengatasi isu-isu ini, strategi yang semakin populer dan efektif adalah *deep ensemble learning*, yaitu sebuah metodologi yang menggabungkan beberapa model *deep learning* individual untuk menghasilkan prediksi akhir yang lebih akurat dan *robust* dibandingkan dengan model tunggal mana pun (Ganaie dkk., 2022).

Secara teoretis, kekuatan *ensemble* terletak pada prinsip diversitas. Ketika model-model yang berbeda digabungkan, mereka cenderung memiliki kelemahan yang saling melengkapi. Sebuah model mungkin akurat dalam satu set kondisi, sementara model lain unggul dalam kondisi yang berbeda. Dengan menggabungkan keputusan dari semua model, *ensemble* dapat memitigasi kesalahan individual dan meningkatkan kemampuan generalisasi secara keseluruhan. Konsep ini dapat diimplementasikan melalui beberapa skema, termasuk *bagging*, *boosting*, dan *stacking* (Ganaie dkk., 2022).

Aplikasi *deep ensemble learning* dalam pengolahan citra sangat beragam. Dalam tugas klasifikasi citra yang menantang, seperti yang ditemukan dalam sistem manajemen maritim, *ensemble* digunakan untuk meningkatkan akurasi dan kehandalan sistem (Mesgaribarzi, 2024). Penelitian juga menunjukkan bahwa *deep ensemble learning* dapat meningkatkan kekokohan (*robustness*) klasifikasi, seperti dalam kasus membedakan citra daging sapi dan babi. Dalam penelitian ini, *EfficientNetB0* digunakan sebagai ekstraktor fitur, dan kemudian hasil fiturnya digabungkan melalui metode *ensemble* untuk mencapai performa yang superior (Akbar dkk., 2025). Selain tugas klasifikasi, *deep ensemble learning* juga sangat efektif untuk deteksi objek.

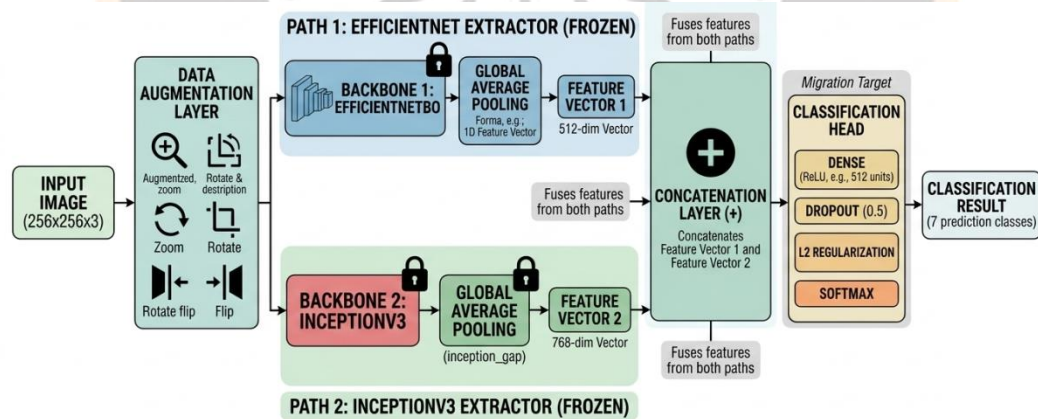
Sebuah studi tentang deteksi kerusakan jalan, misalnya, menunjukkan bahwa menggabungkan beberapa versi dari model deteksi objek canggih, seperti *YOLO-v4*, dalam sebuah *ensemble* dapat secara signifikan meningkatkan skor *F1*, yang merupakan metrik penting untuk mengukur akurasi deteksi (Doshi, 2018). Hal ini menegaskan bahwa *deep ensemble learning* adalah alat yang serbaguna dan tidak terbatas pada satu jenis tugas visual saja. Namun, penerapan *deep ensemble learning* juga menghadapi tantangan, terutama terkait biaya komputasi yang tinggi karena melatih dan menjalankan beberapa model. Untuk mengatasi hal ini, teknik-teknik optimasi seperti *pruning* dapat digunakan (Mesgaribarzi, 2024).

Pruning adalah proses selektif menghilangkan model-model yang kurang efektif atau memiliki korelasi yang tinggi dari *ensemble*, sehingga mempertahankan kinerja optimal sambil mengurangi beban komputasi. Dengan demikian, *deep ensemble learning* menjadi sebuah strategi yang tidak hanya

berfokus pada akurasi, tetapi juga pada efisiensi dan kekinian model dalam lingkungan pengolahan citra yang kompleks. Untuk mengimplementasikan deep ensemble learning secara efektif, beberapa strategi canggih dapat diterapkan.

1. Perancangan *Feature-Level Fusion* dengan optimasi *Dual Backbone*

Sistem memulai seluruh rangkaian proses pengolahan data ketika pengguna memasukkan dataset gambar mentah berukuran 256 X 256 piksel ke dalam alur kerja utama. Lapisan augmentasi data secara cerdas memodifikasi karakteristik visual gambar melalui operasi rotasi, pergeseran lebar, serta penyesuaian kontras untuk memperkaya keberagaman informasi selama masa pelatihan. Arsitektur model kemudian mengarahkan aliran data tersebut ke dalam dua jalur ekstraksi fitur yang beroperasi secara simultan menggunakan backbone *EfficientNetB0* dan *InceptionV3*. Secara bersamaan, jalur kedua menggunakan modul *inception* dari *InceptionV3* untuk mengekstraksi fitur visual pada berbagai skala spasial yang berbeda melalui filter konvolusi yang bervariasi.



Gambar 2.6 Perancangan *Feature-Level Fusion* dengan optimasi *Dual Backbone*

Strategi optimasi sistem ini mengintegrasikan dua model saraf tiruan yang berbeda melalui pendekatan *transfer learning* untuk mencapai tingkat akurasi yang melampaui kemampuan model tunggal. Pengembang memulai tahap optimasi dengan menerapkan teknik *dual preprocessing* yang menyesuaikan rentang nilai piksel gambar secara spesifik untuk masing-masing standar input

EfficientNet dan *Inception*. Proses pelatihan awal mengoptimalkan kinerja sistem dengan menetapkan status *freeze* pada seluruh bobot lapisan *backbone* agar pembaruan parameter hanya terjadi pada bagian kepala klasifikasi. Mekanisme ini memastikan bahwa pengetahuan fitur yang telah dipelajari oleh model prainstal dari dataset *ImageNet* tetap terjaga sambil menyesuaikan lapisan akhir dengan pola data baru secara efisien.

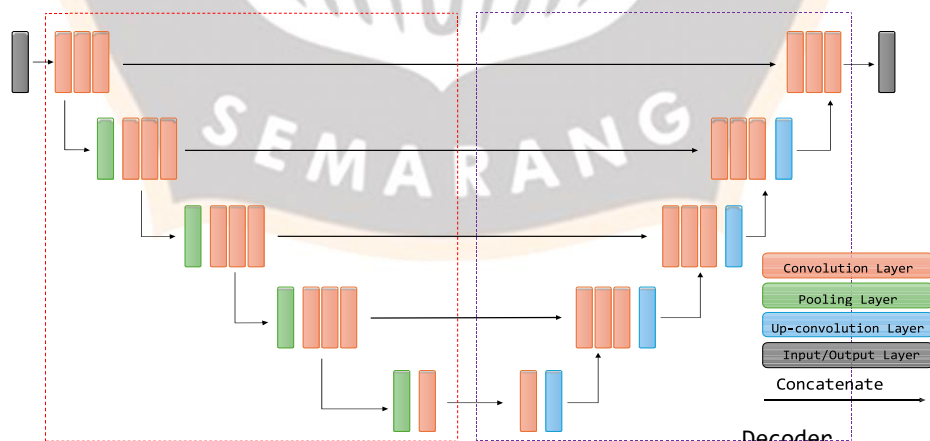
Sistem meningkatkan stabilitas proses pembelajaran melalui penerapan lapisan *batch normalization* dan *dropout* guna memitigasi risiko *overfitting* yang sering muncul pada arsitektur dengan parameter besar. Optimasi pada fase kedua melibatkan teknik *fine-tuning* di mana sistem membuka kunci sebagian lapisan atas dari kedua *backbone* untuk dilatih kembali dengan laju pembelajaran yang sangat kecil. Kombinasi dari kedua arsitektur ini menciptakan keseimbangan yang optimal antara kecepatan pemrosesan data dan ketajaman hasil deteksi. Seluruh rangkaian optimasi ini pada akhirnya menghasilkan model yang memiliki daya generalisasi sangat kuat dalam menghadapi berbagai variasi data di lingkungan pengujian yang nyata.

2. Teknik Regularisasi

Mengingat kompleksitas model *deep learning* dan potensi *overfitting*, terutama pada dataset yang lebih kecil, teknik regularisasi menjadi sangat penting (Yang, 2025). Salah satu teknik yang paling umum adalah *dropout* (Mesgaribarzi, 2024). *Dropout* secara acak menonaktifkan sejumlah neuron selama pelatihan, yang memaksa jaringan untuk tidak terlalu bergantung pada neuron tertentu dan meningkatkan kemampuan generalisasi. Selain *dropout*, teknik lain seperti *pruning* juga dapat digunakan untuk mengoptimalkan ensemble dengan menghilangkan model-model yang tidak efektif, sehingga meningkatkan akurasi dan efisiensi komputasi (Mesgaribarzi, 2024). Secara keseluruhan, *deep ensemble learning* bukanlah sekadar penggabungan model, tetapi merupakan sebuah kerangka kerja terintegrasi yang menggabungkan berbagai strategi canggih dari fusi fitur, perata-rataan model, hingga teknik pelatihan efisien seperti *Snapshot Ensemble* dan regularisasi untuk mencapai kinerja optimal, kekokohan, dan efisiensi dalam pengolahan citra.

2.3.8 Segmentasi Citra

Segmentasi citra adalah proses membagi sebuah gambar digital ke dalam beberapa segmen atau wilayah yang memiliki kemiripan karakteristik visual tertentu (Ghosh dkk., 2026). Teknik ini bertujuan untuk menyederhanakan representasi citra agar objek di dalamnya menjadi lebih mudah untuk dianalisis oleh algoritma komputer (Li dkk., 2022). Dalam pengolahan citra medis maupun industri, segmentasi sering digunakan untuk memisahkan objek utama (*foreground*) dari latar belakang (*background*) agar fitur-fitur penting tidak tercampur dengan informasi yang tidak relevan. Metode segmentasi tradisional biasanya mengandalkan ambang batas warna (*thresholding*) atau deteksi tepi (*edge detection*). Namun, perkembangan *Deep Learning* saat ini memungkinkan proses segmentasi dilakukan secara otomatis dengan tingkat akurasi yang lebih tinggi melalui deteksi objek saliens (*salient object detection*), di mana mesin mampu mengenali objek yang paling menonjol dalam sebuah bingkai citra. Gambar 2.6 merupakan diagram arsitektur *U2-Net* yang menunjukkan konfigurasi *nested U-structure* dan unit *Residual U-block* (RSU) sebagai komponen utama ekstraksi fitur



Gambar 2.7. Ilustrasi arsitektur *U-Net* (Imagery, 2024)

Penelitian ini menerapkan teknik segmentasi citra otomatis pada tahap pra-pemrosesan untuk meningkatkan kualitas data masukan sebelum masuk ke tahap

klasifikasi (Yu dkk., 2022). Penelitian ini menggunakan pustaka perangkat lunak *rembg* yang mengintegrasikan algoritma *deep learning* U2-Net untuk melakukan pemisahan objek secara presisi (Abinaya dkk., 2022). Arsitektur U2-Net tersebut bekerja dengan cara mengekstraksi fitur struktural dari objek melalui mekanisme *nested U-structure* yang sangat dalam. Struktur ini memungkinkan jaringan untuk menangkap informasi kontekstual dari berbagai skala tanpa kehilangan detail resolusi tinggi melalui penggunaan unit *Residual U-block* (RSU). Pemanfaatan *nested U-structure* ini menyerupai prinsip hierarki fitur di mana integrasi berbagai skala fitur sangat krusial untuk menghasilkan peta segmentasi tingkat piksel yang akurat (Imagery, 2024). Sistem memproses citra mentah menggunakan fungsi `remove()` untuk membuang elemen pengganggu di sekitar objek secara otomatis. Melalui algoritma U2-Net, sistem menghasilkan masker transparansi yang secara tajam memisahkan morfologi objek dari tekstur latar belakang yang kompleks. Setelah latar belakang tereliminasi, sistem melakukan konversi ruang warna dari format *RGBA* (dengan lapisan alfa) ke dalam format *RGB* murni menggunakan pustaka Pillow (Yu dkk., 2022). Langkah standardisasi ini sangat penting karena lapisan masukan pada model fusi *EfficientNetB0* dan *InceptionV3* hanya mendukung tiga kanal warna standar tanpa transparansi. Prosedur penghilangan *noise* latar belakang ini memastikan model klasifikasi dapat memfokuskan ekstraksi fitur sepenuhnya pada detail fisik objek. Melalui prosedur pra-pemrosesan tersebut, sistem berhasil mencapai tingkat akurasi yang tinggi karena data masukan telah bersih dari informasi yang tidak relevan.

2.3.9 Augmentasi Citra

Augmentasi citra dalam *Convolutional Neural Networks* (CNN) adalah teknik yang digunakan untuk meningkatkan variasi dan jumlah data pelatihan dengan membuat modifikasi kecil pada citra asli (Montalbo dan Hernandez, 2019). Ini sangat penting untuk meningkatkan performa dan generalisasi model. Dalam pengembangan model *Deep Learning*, augmentasi berfungsi sebagai teknik regularisasi untuk mencegah terjadinya *overfitting*. Dengan memberikan variasi buatan, model dipaksa untuk belajar mengenali objek dari berbagai sudut pandang,

kondisi pencahayaan, dan posisi yang berbeda. Hal ini meningkatkan kemampuan generalisasi model sehingga tetap akurat saat menghadapi data uji yang memiliki karakteristik visual berbeda dari data asli. Dalam penelitian ini augmentasi yang digunakan adalah :

1. *RandomRotation(0.2)* : Model akan memutar citra secara acak hingga 20% dari lingkaran penuh. Hal ini membantu model agar tidak kaku dalam mengenali ikan hanya pada posisi sejajar, melainkan tetap mengenalinya meski posisi ikan miring.
2. *RandomZoom(0.2)* : Teknik ini melakukan perbesaran atau pengecilan citra secara acak. Ini melatih model untuk tetap fokus pada tekstur kulit atau mata ikan terlepas dari jarak kamera saat pengambilan gambar dilakukan.
3. *RandomContrast(0.2)* dan *RandomBrightness(0.2)* : Kedua teknik ini memanipulasi intensitas cahaya dan kontras warna pada citra. Karena kesegaran ikan sering dilihat dari warna kulit dan kejernihan mata, augmentasi ini sangat krusial agar model tidak terkecoh oleh kondisi pencahayaan yang terlalu redup atau terlalu terang saat pengambilan sampel.
4. *RandomTranslation(height_factor=0.1, width_factor=0.1)* : Lapisan ini menggeser posisi objek secara vertikal dan horizontal. Hal ini memastikan model tidak hanya belajar bahwa "ikan harus berada di tengah", sehingga model menjadi lebih tangguh jika posisi ikan sedikit bergeser dari pusat *frame*.

Implementasi lapisan augmentasi data berfungsi sebagai strategi regularisasi untuk memperkuat ketahanan model terhadap variasi input. Peneliti mengintegrasikan berbagai manipulasi geometris dan fotometrik untuk memperkaya keberagaman informasi pada dataset latih. Teknik pembalikan (*flip*), rotasi, dan pergeseran (*translation*) memaksa jaringan untuk mengekstraksi fitur morfologi ikan yang bersifat invarian terhadap posisi. Penyesuaian kontras dan kecerahan secara acak melatih sensitivitas model terhadap perubahan kondisi pencahayaan pada lingkungan pengambilan gambar. Tahapan ini dijalankan secara sistematis setelah proses segmentasi untuk memastikan transformasi hanya diaplikasikan pada objek utama. Pendekatan tersebut meminimalisir risiko

overfitting yang sering muncul pada dataset dengan jumlah sampel terbatas. Penggunaan augmentasi dinamis pada lapisan input meningkatkan kemampuan generalisasi model secara signifikan. Hasil akhirnya adalah model fusi yang mampu menghasilkan prediksi konsisten dan akurat pada data uji yang memiliki karakteristik visual berbeda.

2.3.10 Confusion Matrix

Confusion matrix merupakan salah satu alat evaluasi yang digunakan untuk mengukur kinerja suatu model klasifikasi dengan membandingkan nilai aktual (kelas sebenarnya) dengan nilai prediksi yang dihasilkan. *Confusion matrix* berfungsi untuk memahami perbedaan antar kelas untuk klasifikasi (*supervised learning*) dan melihat sebaran data validitas percobaan secara visual dan bukan ukuran evaluasi model (Markoulidakis, 2024). Pada penelitian ini, matriks konfusi diterapkan untuk klasifikasi 7 kelas yang termasuk ke dalam kategori klasifikasi *multi-class classification*. Gambar 2.8 menyajikan informasi perbandingan antara label aktual hasil pengamatan dengan label prediksi yang dihasilkan oleh model.

		PREDIKSI						
		KF	KN	TF	TN	TD	ToF	ToN
NILAI AKTUAL	Kerisi Fresh	TP 1	error 1,2	error 1,2	error 1,4	error 1,5	error 1,6	error 1,7
	Kerisi NonFresh	error 2,1	TP 2	error 2,2	error 2,4	error 2,5	error 2,6	error 2,7
	Tenggiri Fresh	error 3,1	error 3,2	TP 3	error 3,4	error 3,5	error 3,6	error 3,7
	Tenggiri NonFresh	error 4,1	error 4,2	error 4,3	TP 4	error 4,5	error 4,6	error 4,7
	Tidak_Dikenal	error 5,1	error 5,2	error 5,3	error 5,4	TP5	error 5,6	error 45,7
	Tongkol Fresh	error 6,1	error 6,2	error 6,3	error 6,4	error 6,5	TP 6	error 6,7
	Tongkol NonFresh	error 7,1	error 7,2	error 7,3	error 7,4	error 7,5	error 7,6	TP 7

Keterangan :

KF = Kerisi *Fresh*, KN = Kerisi *NonFresh*, TF = Tenggiri *Fresh*, TN = Tenggiri *NonFresh*, TD = Tidak_Dikenal, ToF = Tongkol *Fresh*, ToN = Tongkol *NonFresh*.

SEKOLAH PASCASARJANA
Gambar 2.8 *Confusion Matrix*

Berdasarkan nilai-nilai di dalam matriks tersebut, dapat dihitung parameter evaluasi sebagai berikut:

1. Akurasi (*Accuracy*) : Merupakan rasio total prediksi benar (diagonal utama) terhadap keseluruhan jumlah data uji.

$$Accuracy = \frac{\sum_{i=1}^6 TP_i}{Total\ Data} \quad (2.1.12)$$

2. Presisi (*Precision*) : Mengukur tingkat keakuratan model dalam mengklasifikasikan instansi ke dalam kelas tertentu dibandingkan dengan total prediksi kelas tersebut.

$$Precision = \frac{TP}{TP+FP} \quad (2.1.13)$$

3. Recall (*Sensitivity*) : Mengukur kemampuan model dalam mengidentifikasi kembali seluruh sampel yang benar-benar termasuk dalam kelas aktual tersebut.

$$Recall = \frac{TP}{TP+FN} \quad (2.1.14)$$

4. *F1-Score* : Memberikan nilai rata-rata harmonik antara presisi dan *recall* untuk memberikan gambaran performa yang lebih seimbang, terutama pada kondisi dataset dengan distribusi kelas yang beragam.

$$F1-Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (2.1.15)$$

Keterangan :

True positive (TP) merupakan jumlah data kelas positif yang terklasifikasi sebagai positif oleh sistem

True Negative (TN) merupakan data kelas negatif yang terklasifikasi sebagai negatif oleh sistem

False Positive (FP) merupakan jumlah data kelas negatif yang terklasifikasi sebagai positif oleh sistem

False Negative (FN) merupakan jumlah data kelas positif yang terklasifikasi sebagai negatif oleh sistem