

TESIS
MODEL *STACKING* CNN
UNTUK KLASIFIKASI CITRA BATUAN



Disusun oleh:

RUDIYANTO

22.55.1249

Konsentrasi: Business Intelligence

FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA

2025

TESIS
MODEL *STACKING* CNN
UNTUK KLASIFIKASI CITRA BATUAN

CNN STACKING MODEL FOR ROCK IMAGE CLASSIFICATION

Diajukan untuk memenuhi salah satu syarat mencapai derajat Pascasarjana

Program Studi S2 Teknik Informatika



Disusun oleh:

RUDIYANTO

22.55.1249

Konsentrasi: Business Intelligence

FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA

2025

HALAMAN PERSETUJUAN

**MODEL STACKING CNN
UNTUK KLASIFIKASI CITRA BATUAN**

CNN STACKING MODEL FOR ROCK IMAGE CLASSIFICATION

yang disusun dan diajukan oleh

Rudfianto

22.55.1249

telah disetujui oleh Dosen Pembimbing Tesis
pada tanggal 02 Desember 2025

Dosen Pembimbing,



Prof. Arief Setyanto, S.Si., M.T., Ph.D.
NIK. 190302036

HALAMAN PENGESAHAN

**MODEL STACKING CNN
UNTUK KLASIFIKASI CITRA BATUAN**

CNN STACKING MODEL FOR ROCK IMAGE CLASSIFICATION

yang disusun dan diajukan oleh

Rudlyanto

22.55.1249

Telah dipertahankan di depan Dewan Penguji
pada tanggal 02 Desember 2025

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Tonny Hidayat, S.Kom., M.Kom., Ph.D.

NIK. 190302182

Robert Marco, S.T., M.T., Ph.D.

NIK. 190302228

Prof. Arlef Setyanto, S.Si., M.T., Ph.D.

NIK. 190302036



Tesis ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Magister Komputer
Tanggal 02 Desember 2025

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusriani, M.Kom.

NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN TESIS

Yang bertandatangan di bawah ini,

Nama mahasiswa : Rudyanto
NIM : 22.55.1249
Konsentrasi : Business Intelligence

Menyatakan bahwa Tesis dengan judul berikut:

MODEL STACKING CNN UNTUK KLASIFIKASI CITRA BATUAN

Dosen Pembimbing Utama : Prof. Arief Setyanto, S.Si, M.T., Ph.D.
Dosen Pembimbing Pendamping : Kusnawi, S.Kom., M.Eng

1. Karya tulis ini adalah benar-benar **ASLI** dan **BELUM PERNAH** diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian **SAYA** sendiri, tanpa bantuan pihak lain kecuali arahan dari Tim Dosen Pembimbing
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab **SAYA**, bukan tanggung jawab Universitas AMIKOM Yogyakarta
5. Pernyataan ini **SAYA** buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka **SAYA** bersedia menerima **SANKSI AKADEMIK** dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi

Yogyakarta, 02 Desember 2025

Yang Menyatakan,



Rudyanto



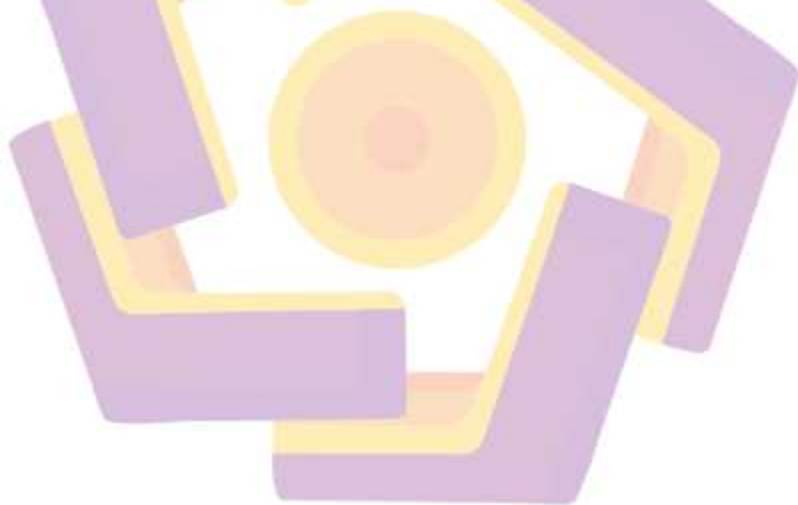
HALAMAN PERSEMBAHAN

Dengan segala kerendahan hati dan rasa syukur yang mendalam, karya ilmiah ini penulis persembahkan kepada:

- 1 Tuhan Yang Maha Esa, atas segala limpahan rahmat, petunjuk, kekuatan, dan kesempatan yang telah diberikan sehingga penulis dapat menyelesaikan tesis ini dengan sebaik-baiknya.
- 2 Almarhumah Ibu Asminah, sosok penuh kasih yang menjadi sumber inspirasi, semangat, dan doa dalam setiap langkah kehidupan penulis. Semoga Allah SWT senantiasa melampirkan tempat terbaik di sisi-Nya.
- 3 Istri tercinta, Susanti, S.KM, yang selalu memberikan dukungan, doa, dan pengorbanan tanpa henti, serta anak-anak tercinta, Naura Hasna Khoirunnisa dan Nufiza Febrina Nashwa, yang menjadi cahaya dan penguat semangat dalam setiap perjuangan.
- 4 Bapak Prof. Arief Setyanto, S.Si, M.T., Ph.D dan Bapak Kusnawi, S.Kom., M.Eng, selaku dosen pembimbing, atas segala bimbingan, ilmu, kesabaran, dan motivasi yang telah diberikan sepanjang proses penyusunan tesis ini.
- 5 Bapak Dr. Robert Marco, M.T, Bapak Dr. Andi Sunyoto, M.Kom, Tonny Hidayat, S.Kom., M.Kom., Ph.D, dan Bapak Prof. Arief Setyanto, S.Si, M.T., Ph.D, selaku dosen penguji tesis, atas saran, kritik konstruktif, dan masukan berharga yang telah memperkaya kualitas karya ilmiah ini.

- 6 Seluruh dosen AMIKOM Yogyakarta, yang telah membekali penulis dengan ilmu pengetahuan, wawasan, serta nilai-nilai akademik yang sangat berharga selama masa studi.
- 7 Rekan-rekan seangkatan tahun 2022, yang telah menjadi teman berbagi cerita, perjuangan, dan semangat dalam melewati setiap fase perkuliahan.
- 8 Sahabat-sahabat di Fakultas Geografi Universitas Muhammadiyah Surakarta, atas kebersamaan, dukungan, dan semangat yang telah diberikan selama ini.

Semoga karya ini dapat memberikan manfaat dan menjadi kontribusi positif dalam pengembangan ilmu pengetahuan.



KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas segala rahmat, taufik, dan hidayah-Nya, sehingga penulis dapat menyelesaikan tesis ini yang berjudul "MODEL *STACKING* CNN UNTUK KLASIFIKASI CITRA BATUAN" sebagai salah satu syarat untuk memperoleh gelar Magister Teknik Informatika di Universitas AMIKOM Yogyakarta.

Penyusunan tesis ini tentunya tidak lepas dari bimbingan, dukungan, serta doa dari berbagai pihak. Oleh karena itu, dengan segala kerendahan hati, penulis menyampaikan ucapan terima kasih dan penghargaan yang sebesar-besarnya kepada:

- 1 Prof. Dr. M. Suyanto, MM., selaku Rektor Universitas AMIKOM Yogyakarta, yang telah memberikan fasilitas dan suasana akademik yang kondusif selama penulis menempuh studi.
- 2 Prof. Dr. Kusriani, M.Kom., selaku Direktur Program Pascasarjana Universitas AMIKOM Yogyakarta, atas arahnya dalam pelaksanaan program pendidikan Magister Teknik Informatika.
- 3 Bapak Prof. Arief Setyanto, S.Si, M.T., Ph.D. dan Bapak Kusnawi, S.Kom., M.Eng., selaku dosen pembimbing yang telah dengan sabar memberikan arahan, bimbingan, motivasi, dan masukan berharga selama proses penyusunan tesis ini.
- 4 Bapak Dr. Robert Marco, M.T., Bapak Dr. Andi Sunyoto, M.Kom., Bapak Tonny Hidayat, S.Kom., M.Kom., Ph.D, dan Bapak Prof. Arief Setyanto, S.Si, M.T., Ph.D., selaku tim penguji tesis, yang telah memberikan masukan dan evaluasi yang membangun untuk penyempurnaan karya ini.

- 5 Seluruh dosen Program Magister Teknik Informatika Universitas AMIKOM Yogyakarta, atas ilmu dan wawasan yang telah diberikan selama masa studi.
- 6 Orang tua tercinta, terutama almarhumah ibu penulis, yang telah menjadi sumber doa, semangat, dan keteladanan hidup dalam setiap langkah perjuangan ini.
- 7 Rekan-rekan mahasiswa PJJ MTI angkatan 2022, atas kebersamaan, dukungan, dan semangat dalam menjalani setiap proses pembelajaran.
- 8 Rekan kerja di lingkungan Fakultas Geografi Universitas Muhammadiyah Surakarta, atas dukungan moril, semangat, dan kerja sama yang telah banyak membantu selama masa studi dan penyusunan tesis ini.

Penulis menyadari bahwa tesis ini masih jauh dari sempurna. Oleh karena itu, segala kritik dan saran yang membangun sangat penulis harapkan untuk perbaikan di masa mendatang. Semoga karya ini dapat memberikan manfaat bagi pengembangan ilmu pengetahuan, khususnya dalam bidang Teknik Informatika..

Yogyakarta, 02 Desember 2025

Penulis

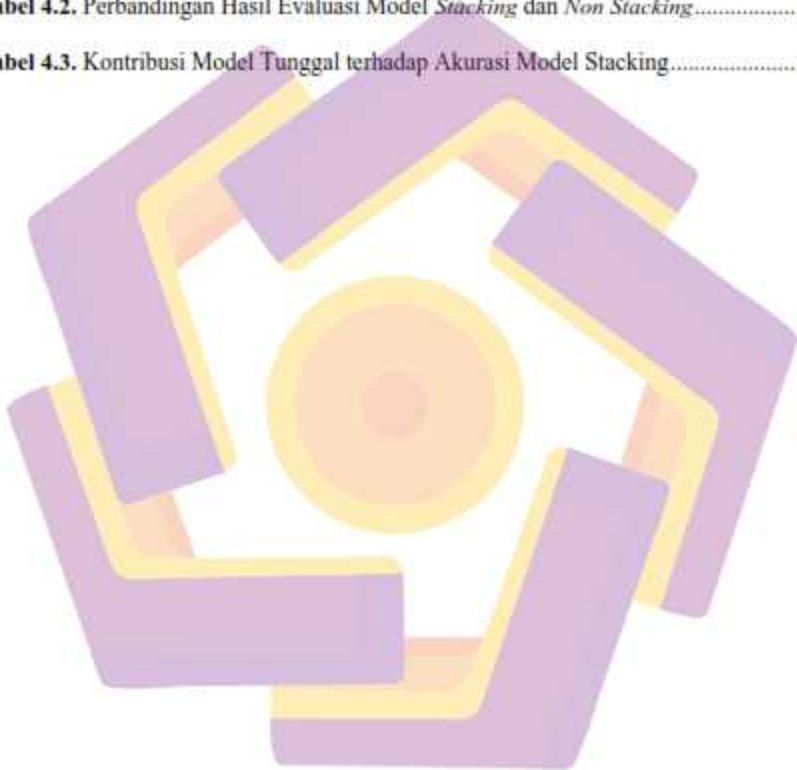
DAFTAR ISI

| | |
|---|------|
| HALAMAN JUDUL | i |
| HALAMAN PERSETUJUAN | ii |
| HALAMAN PENGESAHAN | iii |
| HALAMAN PERNYATAAN KEASLIAN TESIS Error! Bookmark not defined. | |
| HALAMAN PERSEMBAHAN | iv |
| KATA PENGANTAR | vii |
| DAFTAR ISI | ix |
| DAFTAR TABEL | xi |
| DAFTAR GAMBAR | xii |
| DAFTAR ISTILAH | xiii |
| BAB I PENDAHULUAN | 1 |
| 1.1. Latar Belakang Masalah | 1 |
| 1.2. Rumusan Masalah | 4 |
| 1.3. Batasan Masalah | 5 |
| 1.4. Tujuan Penelitian | 6 |
| 1.5. Manfaat Penelitian | 6 |
| 1.6. Hipotesis | 7 |
| BAB II TINJAUAN PUSTAKA | 8 |
| 2.1. Tinjauan Pustaka | 8 |
| 2.2. Keaslian Penelitian | 10 |
| 2.3. Landasan Teori | 19 |

| | |
|--|----|
| BAB III METODE PENELITIAN | 25 |
| 3.1. Jenis, Sifat, dan Pendekatan Penelitian | 25 |
| 3.2. Alat dan Perangkat Penelitian | 26 |
| 3.3. Data, Metode Pengumpulan dan Pengolahan Data | 26 |
| 3.4. Metode Analisis Data | 28 |
| 3.5. Alur Penelitian | 29 |
| BAB IV HASIL PENELITIAN DAN PEMBAHASAN | 36 |
| 4.1. <i>Exploratory Data Analysis (EDA)</i> | 36 |
| 4.2. Hasil Preprocessing Data | 38 |
| 4.3. Pembagian Dataset | 39 |
| 4.4. Hasil Akurasi Model Stacking | 40 |
| 4.5. Perbandingan Evaluasi Model <i>Stacking</i> dan <i>Non Stacking</i> | 44 |
| 4.6. Pembahasan | 59 |
| BAB V PENUTUP | 70 |
| 5.1. Kesimpulan | 71 |
| 5.2. Saran | 72 |
| DAFTAR PUSTAKA | 74 |

DAFTAR TABEL

| | |
|---|----|
| Tabel 2.1. Matriks Literatur Review dan Posisi Penelitian | 11 |
| Tabel 4.1. Hasil Akurasi pada Model <i>Stacking</i> | 41 |
| Tabel 4.2. Perbandingan Hasil Evaluasi Model <i>Stacking</i> dan <i>Non Stacking</i> | 45 |
| Tabel 4.3. Kontribusi Model Tunggal terhadap Akurasi Model <i>Stacking</i> | 64 |



DAFTAR GAMBAR

| | |
|--|----|
| Gambar 2.1. Struktur Dasar dari CNN..... | 21 |
| Gambar 2.2. Tabel <i>Confusion Matrix</i> | 24 |
| Gambar 3.1. Proses <i>Stratified K-Fold Cross Validation Data</i> | 30 |
| Gambar 3.2. Diagram Alir Penelitian | 33 |
| Gambar 4.1. Hasil <i>Exploratory Data Analisis Citra Batuan</i> | 36 |
| Gambar 4.2. Contoh <i>Dataset Batuan</i> | 37 |
| Gambar 4.3. Perbandingan Jumlah <i>Dataset</i> Sebelum dan Sesudah <i>Preprocessing</i> | 39 |
| Gambar 4.4. Perbandingan Hasil Akurasi Model <i>Stacking</i> | 43 |
| Gambar 4.5. Perbandingan Hasil Evaluasi Model <i>Stacking</i> dan <i>Non Stacking</i> | 47 |
| Gambar 4.6. Hasil <i>Confusion Matrix</i> pada Model 1(MobileNetV2 + EfficienNetB0). 50 | |
| Gambar 4.7. Hasil <i>Confusion Matrix</i> pada Model 2 (VGG16 + DenseNet121)..... | 51 |
| Gambar 4.8. Hasil <i>Confusion Matrix</i> pada Model 3 (ResNet50, InceptionV3)..... | 51 |
| Gambar 4.9. Hasil <i>Confusion Matrix</i> pada Model 4 (Resnet50, MobileNetV2) | 51 |
| Gambar 4.10. Hasil <i>Confusion Matrix</i> pada Resnet50..... | 54 |
| Gambar 4.11. Hasil <i>Confusion Matrix</i> pada InceptionV3 | 54 |
| Gambar 4.12. Hasil <i>Confusion Matrix</i> pada VGG16..... | 55 |
| Gambar 4.13. Hasil <i>Confusion Matrix</i> pada DenseNet121..... | 55 |
| Gambar 4.14. Hasil <i>Confusion Matrix</i> pada MobileNetV2 | 56 |
| Gambar 4.15. Hasil <i>Confusion Matrix</i> pada EfficienNetB0 | 56 |

DAFTAR ISTILAH

1. *CNN (Convolutional Neural Network)*: Arsitektur jaringan saraf tiruan yang dirancang untuk mengenali pola visual dalam data berbentuk grid, seperti citra digital, melalui proses konvolusi, pooling, dan klasifikasi.
2. *Stacking*: Teknik *ensemble learning* yang menggabungkan prediksi dari beberapa model dasar (*base learners*) dengan memanfaatkan model lain (*meta-learner*) untuk membuat prediksi akhir.
3. *Base Learner*: Model dasar atau arsitektur CNN yang digunakan untuk mengekstraksi fitur dari citra, misalnya: ResNet50, MobileNetV2, InceptionV3.
4. *Meta Learner*: Model yang digunakan pada lapisan kedua dalam stacking untuk menggabungkan output dari *base learners* dan menghasilkan prediksi akhir. Dalam penelitian ini digunakan SVM.
5. *Fine-Tuning*: Proses penyesuaian ulang (*retraining*) sebagian atau seluruh layer model pretrained agar lebih sesuai dengan dataset baru.
6. *Transfer Learning*: Teknik pembelajaran mesin yang memanfaatkan model yang telah dilatih sebelumnya pada dataset besar untuk diterapkan pada tugas klasifikasi baru dengan dataset yang lebih kecil.
7. *Augmentasi Data*: Teknik untuk memperluas variasi data dengan melakukan transformasi seperti rotasi, *flipping*, *shifting*, dan *zooming* guna mengurangi *overfitting* dan meningkatkan generalisasi model.

8. *Stratified K-Fold Cross Validation*: Teknik validasi silang yang membagi dataset menjadi k subset, memastikan distribusi kelas tetap seimbang pada setiap subset.
9. *F1-Score*: Metrik evaluasi yang menggabungkan nilai presisi dan *recall* untuk memberikan penilaian kinerja model yang seimbang.
10. *Confusion Matrix*: Matriks yang menunjukkan jumlah prediksi benar dan salah dari model klasifikasi terhadap setiap kelas dalam dataset.
11. *Feature Extraction*: Proses otomatis pengambilan ciri-ciri penting dari gambar oleh model CNN untuk keperluan klasifikasi.
12. ResNet50, InceptionV3, VGG16, DenseNet121, MobileNetV2, EfficientNetB0: Nama-nama arsitektur CNN pralatih yang digunakan dalam penelitian ini dengan karakteristik dan keunggulan masing-masing.
13. SVM (*Support Vector Machine*): Algoritma klasifikasi yang mencari *hyperplane* terbaik untuk memisahkan data dari berbagai kelas.

INTISARI

Penelitian ini bertujuan untuk (a) menilai perbedaan akurasi di antara beberapa model *stacking* dalam klasifikasi citra batuan, (b) mengevaluasi kinerja klasifikasi model *stacking* berdasarkan metrik akurasi, *precision*, *recall*, dan *F1-score* serta membandingkannya dengan model *non-stacking* (model tunggal), dan (c) menganalisis kontribusi masing-masing model *non-stacking* terhadap akurasi yang dihasilkan oleh model *stacking*. Penelitian dilakukan secara eksperimental dengan menguji enam model CNN pralatih ResNet50, InceptionV3, VGG16, DenseNet121, MobileNetV2, dan EfficientNetB0, serta empat konfigurasi model *stacking*, yakni model 1 (MobileNetV2 dan EfficientNetB0), model 2 (VGG16 dan DenseNet121), model 3 (ResNet50 dan InceptionV3), dan model 4 (ResNet50 dan MobileNetV2). Masing-masing konfigurasi terdiri atas dua model sebagai *base learner* dan satu *Support Vector Machine* (SVM) sebagai *meta-learner*. Untuk menjamin validitas hasil, digunakan teknik *Stratified K-Fold Cross Validation* pada data pelatihan. Hasil menunjukkan bahwa seluruh konfigurasi *stacking* mampu melampaui hasil model tunggal. Konfigurasi *stacking* model 4 (ResNet50 dan MobileNetV2) mencapai akurasi pengujian tertinggi sebesar 84%, diikuti oleh model *stacking* model 1 (MobileNetV2 dan EfficientNetB0) dan 3 (ResNet50 dan InceptionV3) dengan 83%, serta *stacking* model 2 (VGG16 dan DenseNet121) dengan 81%. Sebagai perbandingan, model individu terbaik, ResNet50, mencatat akurasi 80%, sementara MobileNetV2 dan EfficientNetB0 hanya mencapai 77%. Penerapan metode *stacking* ini mampu memberikan peningkatan akurasi sekitar 4%-11%. Jika melihat dari sisi efisiensi, *stacking* model 1 memiliki ukuran model terkecil, sehingga cocok diterapkan pada perangkat dengan sumber daya terbatas namun jika dilihat dari sisi akurasinya, maka *stacking* model 4 sangat cocok untuk diterapkan. Analisis lebih lanjut menunjukkan bahwa masing-masing model dalam *ensemble stacking* memiliki peran strategis yang saling melengkapi. Model bersitektur besar memberikan cakupan fitur mendalam, sedangkan model ringan berkontribusi dalam menangkap detail lokal. Kolaborasi ini memperkuat kemampuan generalisasi sistem terhadap data baru. Secara keseluruhan, teknik *ensemble stacking* terbukti efektif, adaptif, dan efisien untuk meningkatkan akurasi klasifikasi citra batuan, terutama pada dataset terbatas dan kelas dengan kemiripan visual tinggi.

Kata kunci: *stacking*, pembelajaran ensemble, citra batuan, CNN, SVM

ABSTRACT

This study aims to (a) assess the accuracy differences among several stacking models for rock image classification, (b) evaluate the classification hasilnce of stacking models based on accuracy, precision, recall, and F1-score, and compare it with non-stacking (single) models, and (c) analyze the contribution of each non-stacking model to the accuracy achieved by the stacking model. The research was conducted experimentally by testing six pre-trained CNN architectures ResNet50, InceptionV3, VGG16, DenseNet121, MobileNetV2, and EfficientNetB0 as well as four stacking configurations: Model 1 (MobileNetV2 and EfficientNetB0), Model 2 (VGG16 and DenseNet121), Model 3 (ResNet50 and InceptionV3), and Model 4 (ResNet50 and MobileNetV2). Each configuration consisted of two models serving as base learners and a Support Vector Machine (SVM) as the meta-learner. To ensure the validity of the results, the Stratified K-Fold Cross Validation technique was applied during training. The results indicate that all stacking configurations outperformed their corresponding single models. The Model 4 stacking configuration (ResNet50 and MobileNetV2) achieved the highest testing accuracy of 84%, followed by Model 1 (MobileNetV2 and EfficientNetB0) and Model 3 (ResNet50 and InceptionV3) with 83%, while Model 2 (VGG16 and DenseNet121) reached 81%. In comparison, the best-performing individual model, ResNet50, achieved 80%, whereas MobileNetV2 and EfficientNetB0 reached only 77%. These findings show that the stacking approach improved classification accuracy by approximately 4%–11%. In terms of efficiency, Model 1 stacking had the smallest model size, making it suitable for devices with limited computational resources. However, from an accuracy standpoint, Model 4 stacking proved to be the most optimal configuration. Further analysis revealed that each CNN architecture within the ensemble contributed complementary strengths: deeper architectures captured broad and abstract global features, while lightweight models effectively identified fine-grained local details. This collaboration enhanced the system's generalization ability toward unseen data. Overall, the ensemble stacking technique demonstrated to be an effective, adaptive, and efficient approach for improving rock image classification accuracy, particularly when dealing with limited datasets and visually similar rock classes.

Keywords: *stacking, ensemble learning, rock image, CNN, SVM*

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Secara umum, terdapat tiga jenis batuan utama di Indonesia, yakni batuan beku, batuan sedimen, dan batuan metamorf. Ketiga jenis batuan ini terbentuk melalui proses geologi yang berbeda. Batuan beku terbentuk dari pendinginan magma (Boulanger & France, 2023), batuan sedimen berasal dari proses sedimentasi akibat pelapukan dan erosi (Carter et al., 2023), sedangkan batuan metamorf terbentuk dari transformasi batuan beku atau sedimen akibat tekanan dan suhu tinggi (Hasria et al., 2022). Keanekaragaman warna, tekstur, dan bentuk batuan di lapangan sering kali menjadi kendala bagi surveyor dalam mengidentifikasi jenis batuan dengan akurat.

Kendala tersebut dapat diatasi dengan memanfaatkan teknologi klasifikasi citra yang mampu mengidentifikasi batuan secara otomatis. Klasifikasi citra merupakan salah satu teknik dalam pengolahan citra yang bertujuan untuk mengelompokkan gambar berdasarkan karakteristik visualnya. Saat ini, terdapat berbagai macam algoritma yang dapat dimanfaatkan untuk klasifikasi citra batuan antara lain: *Random Forest*, *Support Vector Machine (SVM)*, *K-Nearest Neighbors (KNN)*, *Decision Tree*, *Naive Bayes*, dan *Convolutional Neural Network (CNN)*.

CNN menjadi salah satu yang paling populer di kalangan peneliti karena kemampuannya dalam mengenali pola visual secara efektif (Khanam & Mehta, 2023). Banyaknya arsitektur CNN yang tersedia membuat pemilihan model terbaik untuk

klasifikasi batuan menjadi tantangan tersendiri, karena masing-masing memiliki kelebihan dan kekurangan.

Meskipun memiliki kemampuan ekstraksi fitur yang unggul, CNN juga memiliki sejumlah keterbatasan dalam implementasinya. Model ini umumnya memerlukan jumlah dataset yang besar agar mampu melakukan generalisasi dengan baik, serta membutuhkan sumber daya komputasi tinggi untuk proses pelatihan dan inferensi. Selain itu, waktu pelatihan yang lama sering menjadi kendala terutama pada arsitektur dengan jumlah parameter besar. Permasalahan tersebut menjadikan efisiensi dan efektivitas penggunaan CNN sebagai salah satu isu penting dalam penelitian klasifikasi citra, termasuk citra batuan.

Perbandingan beberapa arsitektur *Convolutional Neural Network (CNN)* menunjukkan variasi hasil yang dipengaruhi oleh tingkat kompleksitas dan efisiensi model. AlexNet menjadi pelopor CNN modern dengan struktur sederhana dan mudah dipahami, namun memiliki ukuran model besar (~240 MB) sehingga menimbulkan isu kebutuhan komputasi tinggi. MobileNetV2 lebih ringan (8,66 MB) dan efisien untuk perangkat dengan GPU terbatas, tetapi akurasi (79,4%) relatif lebih rendah. Sementara itu, EfficientNet menunjukkan keseimbangan antara akurasi (85,2%) dan ukuran model (12,38 MB), meskipun struktur yang kompleks menyulitkan modifikasi dan sensitif terhadap perubahan data (Darapaneni et al., 2020).

Model dengan arsitektur lebih dalam seperti ResNet-50 dan InceptionV3 memberikan akurasi tinggi masing-masing sebesar 92,04% dan 94,25%. ResNet-50 unggul karena adanya mekanisme *shortcut connection* yang mencegah degradasi informasi pada jaringan dalam, namun ukurannya yang besar (90,05 MB) menjadi isu

efisiensi. Sebaliknya, InceptionV3 menawarkan ukuran kecil (8,24 MB) dengan akurasi tertinggi melalui kombinasi berbagai ukuran filter dalam satu blok, tetapi struktur yang kompleks menyulitkan penyesuaian lebih lanjut. Secara umum, terdapat *trade-off* antara akurasi, ukuran model, dan kompleksitas arsitektur, yang menjadi pertimbangan utama dalam pemilihan model sesuai kebutuhan dan kapasitas sistem (Darapaneni et al., 2020).

Perbedaan karakteristik dari masing-masing arsitektur tersebut menunjukkan adanya *trade-off* antara akurasi, efisiensi, dan kompleksitas model, yang menjadi faktor penting dalam menentukan kinerja sistem klasifikasi citra batuan. Model dengan akurasi tinggi sering kali membutuhkan sumber daya besar, sedangkan model ringan cenderung mengalami penurunan akurasi. Oleh karena itu, pemilihan arsitektur CNN yang tepat menjadi krusial untuk mencapai keseimbangan antara hasil dan efisiensi sistem.

Namun demikian, sebagian besar penelitian sebelumnya hanya berfokus pada peningkatan akurasi tanpa memperhatikan aspek validitas hasil dan kemampuan generalisasi model. Banyak studi yang melaporkan akurasi tinggi meskipun proses pelatihannya tidak melibatkan teknik *cross-validation*, sehingga hasil model berpotensi tidak konsisten terhadap data baru. Selain itu, sebagian penelitian masih menggunakan satu arsitektur CNN tunggal tanpa mempertimbangkan integrasi antar-model yang dapat meningkatkan stabilitas dan ketepatan klasifikasi.

Berdasarkan kondisi tersebut, diperlukan pendekatan lanjutan yang tidak hanya mengejar akurasi tinggi, tetapi juga memastikan keandalan hasil melalui evaluasi berbasis *cross-validation* serta penggabungan keunggulan dari berbagai arsitektur CNN. Salah satu metode yang dapat diterapkan adalah *ensemble learning*, khususnya teknik *stacking model*, yang memungkinkan kombinasi beberapa arsitektur CNN dengan

karakteristik berbeda untuk menghasilkan prediksi yang lebih stabil dan akurat terhadap data batuan yang kompleks.

Penelitian ini mengimplementasikan enam arsitektur *Convolutional Neural Network (CNN)* sebagai model individu, yaitu ResNet50, InceptionV3, VGG16, DenseNet121, MobileNetV2, dan EfficientNetB0. Sementara itu, pendekatan *ensemble stacking* dikategorikan ke dalam empat skema berdasarkan kompleksitas model, yakni: model 1 (MobileNetV2 dan EfficientNetB0), model 2 (VGG16 dan DenseNet121), model 3 (ResNet50 dan InceptionV3), dan model 4 (ResNet50 dan MobileNetV2). Semua model-model tersebut berfungsi sebagai *base learner*, sementara *meta learner*-nya memanfaatkan *Support Vector Machine (SVM)*.

1.2. Rumusan Masalah

Berdasarkan pada latarbelakang permasalahan yang ada, maka dapat dirumuskan beberapa permasalahan sebagai berikut:

- a. Seberapa besar nilai akurasi dari beberapa model *stacking* dalam klasifikasi citra batuan?
- b. Apakah nilai akurasi, precision, recall, dan F1-score yang dihasilkan oleh model *stacking* lebih tinggi dibandingkan dengan model *non-stacking* (model tunggal)?
- c. Seberapa besar kontribusi setiap model *non-stacking* (model tunggal) terhadap akurasi yang dihasilkan oleh model *stacking*?

1.3. Batasan Masalah

Agar penelitian ini tetap fokus dan memiliki ruang lingkup yang jelas, diperlukan batasan masalah yang menghindari pembahasan di luar cakupan utama. Batasan ini juga bertujuan untuk memastikan hasil penelitian lebih terarah dan sesuai dengan tujuan yang telah ditetapkan. Adapun batasan masalah dalam penelitian ini adalah sebagai berikut:

- a. Model yang digunakan dalam penelitian ini terbatas pada model-model CNN yang sudah dikenal, seperti ResNet50, InceptionV3, VGG16, DenseNet121, MobileNetV2, dan EfficientNetB0.
- b. Teknik yang diterapkan dalam penelitian ini adalah metode *ensemble stacking* dengan pembagian kategori model menjadi ringan, kuat, dan kombinasi.
- c. Penelitian ini hanya menggunakan dataset batuan dari Brac University Mars Rover dan tersimpan di website KAGGLE, yang terdiri dari citra dalam format RGB dan tidak mencakup citra dengan kondisi khusus seperti citra multispektral atau data berbasis teks. Dataset dipilih karena merupakan dataset terbuka yang belum banyak dipakai untuk riset.
- d. Evaluasi model akan dilakukan menggunakan metrik akurasi keseluruhan seperti *Accuracy, F1-Score, Precision, Recall*.
- e. Penelitian ini terbatas pada penggunaan perangkat keras dengan kemampuan komputasi T4 GPU, yang mempengaruhi kecepatan proses *training* dan *inference*.

1.4. Tujuan Penelitian

Berdasarkan pada perumusan permasalahan yang ada, maka tujuan dari penelitian ini adalah sebagai berikut:

- a. Menilai besarnya perbedaan akurasi dari beberapa model *stacking* dalam klasifikasi citra batuan.
- b. Mengevaluasi hasil klasifikasi model *stacking* berdasarkan metrik akurasi, *precision*, *recall*, dan *F1-score*, serta membandingkannya dengan model *non-stacking* (model tunggal).
- c. Menilai besarnya kontribusi setiap model *non-stacking* (model tunggal) terhadap akurasi yang dihasilkan oleh model *stacking*.

1.5. Manfaat Penelitian

Manfaat yang dapat dihasilkan dari penelitian ini antara lain:

- a. Penelitian ini akan memperluas pemahaman tentang teknik *ensemble stacking*, khususnya dalam kombinasi berbagai arsitektur CNN yang memiliki karakteristik berbeda dalam hal akurasi dan efisiensi. Hasil penelitian ini dapat memperkaya literatur yang ada tentang metode *ensemble* di bidang pengolahan citra dan kecerdasan buatan.
- b. Penelitian ini memberikan kontribusi terhadap penelitian yang menggabungkan berbagai model CNN dalam satu sistem *ensemble* untuk menghasilkan klasifikasi citra yang lebih akurat dan efisien.

- c. Hasil dari penelitian ini dapat diimplementasikan dalam sistem klasifikasi citra batuan, yang memiliki potensi aplikasi dalam berbagai bidang seperti geologi, eksplorasi sumber daya alam, dan pengelolaan sumber daya mineral.
- d. Pemanfaatan model *stacking* dalam klasifikasi citra batuan dalam penelitian ini dapat meningkatkan kinerja aplikasi-aplikasi yang membutuhkan akurasi tinggi, seperti pemetaan geologi dan identifikasi jenis batuan di lapangan.
- e. Manfaat langsung dari penelitian ini bisa dirasakan dalam berbagai sektor, termasuk industri kreatif (pengeditan foto otomatis), keamanan (pengenalan wajah), atau media sosial (klasifikasi foto berdasarkan kategori).

1.6. Hipotesis

- a. Terdapat perbedaan akurasi dari beberapa model *stacking* dalam klasifikasi citra batuan.
- b. Model *stacking* memiliki akurasi klasifikasi citra batuan yang lebih baik dibandingkan dengan model *non stacking* (model tunggal).
- c. Setiap model *non-stacking* (model tunggal) memberikan kontribusi yang bervariasi terhadap hasil akurasi klasifikasi citra batuan yang dihasilkan dari model *stacking*.

BAB II

TINJAUAN PUSTAKA

2.1. Tinjauan Pustaka

2.1.1. Penelitian klasifikasi batuan

Penelitian klasifikasi jenis batuan berbasis citra telah banyak dilakukan seiring dengan berkembangnya teknologi pengolahan citra dan pembelajaran mesin. Beberapa studi menggunakan pendekatan konvensional seperti metode ekstraksi fitur manual yang dikombinasikan dengan algoritma klasifikasi seperti SVM atau *K-Nearest Neighbor* (Fauzi et al., 2022). Namun, pendekatan ini sering kali kurang akurat karena keterbatasan dalam mengenali pola kompleks dari tekstur batuan. Studi yang dilakukan oleh (Ali et al., 2023), juga menunjukkan bahwa metode klasifikasi berbasis tekstur tradisional menghasilkan akurasi yang bervariasi tergantung pada kualitas dataset citra dan pencahayaan.

2.1.2. Penggunaan CNN dalam klasifikasi citra batuan

Munculnya *Convolutional Neural Network (CNN)* berdampak pada meningkatnya hasil klasifikasi batuan secara signifikan. CNN mampu mengekstraksi fitur spasial secara otomatis tanpa perlu proses ekstraksi manual, sehingga lebih adaptif terhadap variasi bentuk dan tekstur batuan. Beberapa penelitian telah menerapkan CNN seperti AlexNet, VGG, dan ResNet dalam klasifikasi batuan dengan hasil yang cukup menjanjikan. Meskipun demikian, model CNN klasik seperti AlexNet memiliki keterbatasan pada efisiensi dan skala parameter yang besar, sementara model seperti

ResNet cenderung memerlukan daya komputasi tinggi untuk mencapai akurasi optimal (Darapaneni et al., 2020).

2.1.3. Ensemble learning dalam klasifikasi citra

Teknik ensemble learning telah banyak diterapkan dalam bidang komputer vision untuk meningkatkan akurasi dan ketahanan model klasifikasi dan prediksi. Salah satu teknik ensemble yang efektif adalah stacking, yaitu menggabungkan prediksi dari beberapa model dasar (*base learners*) yang berbeda melalui meta-learner (Yildiz et al., 2023). Pada konteks klasifikasi citra, stacking dapat membantu mengintegrasikan kelebihan model ringan dan model dalam (*deep*) untuk menghasilkan hasil yang lebih stabil dan akurat. Penelitian yang dilakukan oleh (Ouedraogo et al., 2023) menunjukkan bahwa kombinasi CNN dan EfficientNet dalam skema stacking dapat meningkatkan akurasi klasifikasi citra limbah TPA hingga 2–4% dibandingkan model tunggal.

2.1.4. Kekosongan riset (Gap)

Meskipun berbagai model CNN telah berhasil digunakan dalam klasifikasi citra batuan, penggunaan teknik *ensemble learning*, khususnya metode *stacking* antara berbagai jenis arsitektur CNN (ResNet50, InceptionV3, VGG16, DenseNet121, MobileNetV2, dan EfficientNetB0) sebagai *base learner* dan SVM sebagai *meta learner* masih belum banyak dieksplorasi. Padahal, gabungan ini berpotensi meningkatkan akurasi klasifikasi jenis batuan secara signifikan, terutama dalam kondisi citra yang bervariasi seperti jumlah data yang sedikit, pencahayaan alami atau latar belakang yang tidak homogen. Selain itu riset ini mencoba dengan dataset terbatas, mampu menghasilkan akurasi yang baik. Oleh karena itu, riset ini hadir untuk mengisi

celah tersebut dengan menguji efektivitas teknik *stacking* dalam pengklasifikasian jenis batuan menggunakan citra foto sebagai input utama.

2.2. Keaslian Penelitian

Penelitian ini menawarkan pendekatan inovatif dalam klasifikasi citra batuan dengan menerapkan *ensemble learning* berbasis model *stacking* CNN. Pendekatan ini berbeda dari mayoritas penelitian sebelumnya yang cenderung mengandalkan satu jenis arsitektur CNN secara individual. Metode *stacking* yang dilakukan dalam penelitian ini adalah dengan menggabungkan beberapa arsitektur CNN yang memiliki karakteristik berbeda, seperti: MobileNetV2 yang ringan, EfficientNet yang efisien dan akurat, serta ResNet yang kuat dan dalam. Setiap model dipilih berdasarkan keunggulannya masing-masing, yang saling melengkapi dalam proses klasifikasi. Penggabungan berbagai arsitektur CNN dalam bentuk *stacking* untuk klasifikasi citra batuan masih jarang dijelajahi, sehingga penelitian ini memiliki potensi kontribusi signifikan terhadap pengembangan sistem klasifikasi citra yang lebih akurat, efisien, dan adaptif terhadap variasi data. Keaslian pendekatan ini dapat dilihat secara lebih rinci pada Tabel 2.1.

Tabel 2.1. Matriks Literatur Review dan Posisi Penelitian Model *Stacking* CNN untuk Klasifikasi Jenis Batuan

| No | Judul Penelitian | Nama Peneliti, Tahun, dan Index | Metode Penelitian | Hasil | Keunggulan dan Kelemahan | Perbandingan |
|----|--|---------------------------------------|--|---|--|--|
| 1 | <i>Rock Image Classification using Deep Residual Neural Network with Transfer Learning</i> | (Chen et al.) 2023. Q2 Scopus | <i>Transfer learning</i> dengan arsitektur <i>Google Net</i> | Penelitian ini berhasil membuat model klasifikasi batuan berbasis ResNet34 dan <i>transfer learning</i> dengan akurasi (82%). | Keunggulan: Mencapai akurasi tinggi melalui ResNet34 dan <i>transfer learning</i> , tetap stabil pada berbagai kondisi lapangan, serta dapat dijalankan secara offline pada perangkat <i>embedded</i> . Kelemahan: Dataset asli berukuran kecil (315 gambar dari 7 kelas) sehingga kemampuan generalisasi terbatas, dan hasil cenderung menurun pada gambar | Penelitian ini menggunakan dataset berjumlah 1.632 gambar yang terbagi ke dalam 5 kelas, menerapkan metode <i>ensemble stacking</i> dengan 4 model, serta menggunakan teknik validasi silang |

| No | Judul Penelitian | Nama Peneliti, Tahun, dan Index | Metode Penelitian | Hasil | Keunggulan dan Kelemahan | Perbandingan |
|----|--|--|---|--|--|--|
| | | | | | beresolusi rendah | pada data pelatihan |
| 2 | <i>Deep Learning for Lithological Classification of Carbonate Rock Micro-CT Images</i> | (dos Anjos et al.). 2021. Q2 Scopus | Penelitian menggunakan <i>Convolutional Neural Networks (CNN)</i> | Penelitian ini menghasilkan model terbaik (Model 2 dengan gambar yang di-resize) mencapai: 75.54% akurasi pada evaluasi per gambar, dan 81.33% akurasi pada evaluasi per <i>plug</i> | Keunggulan: Mampu mengotomatisasi klasifikasi litologi secara cepat dan non-destruktif menggunakan micro-CT dan CNN, sehingga mengurangi subjektivitas dan mempercepat analisis yang biasanya memakan waktu. Kelemahan: Keterbatasan dataset yang kecil dan hanya mencakup tiga jenis litologi, sehingga model kurang mampu | Penelitian ini menggunakan dataset berjumlah 1.632 gambar yang terbagi ke dalam 5 kelas, menerapkan metode <i>ensemble stacking</i> dengan 4 model, serta menggunakan teknik validasi silang |

| No | Judul Penelitian | Nama Peneliti, Tahun, dan Index | Metode Penelitian | Hasil | Keunggulan dan Kelemahan | Perbandingan |
|----|--|---------------------------------------|--|--|---|---|
| | | | | (satu sampel utuh). | melakukan generalisasi pada variasi batuan yang lebih kompleks | pada data pelatihan |
| 3 | <i>Impact of Dataset Size and Convolutional Neural Network Architecture on Transfer Learning for Carbonate Rock Classification</i> | (dawson et, al). 2023. Scopus Q1 | Penelitian ini menggunakan <i>deep learning</i> berbasis CNN dengan pendekatan <i>transfer learning</i> dan <i>fine-tuning</i> | VGG19 terbaik (81,3%), hasil naik seiring ukuran dataset | Keunggulan: akurat, cepat, mengurangi subjektivitas, efektif untuk dataset sedang-besar. Kelemahan: Butuh dataset besar, rawan bias, <i>overfitting</i> pada dataset kecil | Penelitian ini menggunakan dataset berjumlah 1.632 gambar yang terbagi ke dalam 5 kelas, menerapkan metode <i>ensemble stacking</i> dengan 4 model, serta menggunakan |

| No | Judul Penelitian | Nama Peneliti, Tahun, dan Index | Metode Penelitian | Hasil | Keunggulan dan Kelemahan | Perbandingan |
|----|--|---------------------------------------|---|---|--|---|
| | | | | | | teknik validasi silang pada data pelatihan |
| 4 | <i>Lithology Classification of Whole Core CT Scans using Convolutional Neural Networks</i> | Chawshin et al., 2021. Q2 Scopus | Penelitian menggunakan <i>Convolutional Neural Network (CNN)</i> dengan arsitektur VGG16 yang telah dilakukan | Model CNN mampu mengklasifikasikan mineral lempung dengan akurasi tinggi, termasuk membedakan jenis mineral yang tampak sangat mirip secara visual. Hasil | Keunggulan: Metode <i>deep learning</i> mampu mengklasifikasikan mineral lempung secara akurat dan konsisten, bahkan pada jenis yang sangat mirip. CNN dapat mengekstraksi fitur kompleks tanpa proses manual, sehingga analisis menjadi lebih cepat, objektif, dan efisien dibandingkan identifikasi | Penelitian ini menggunakan dataset berjumlah 1.632 gambar yang terbagi ke dalam 5 kelas, menerapkan metode <i>ensemble stacking</i> dengan 4 model, serta menggunakan |

| No | Judul Penelitian | Nama Peneliti, Tahun, dan Index | Metode Penelitian | Hasil | Keunggulan dan Kelemahan | Perbandingan |
|----|------------------|---------------------------------------|--|---|---|---|
| | | | <p><i>transfer learning</i> dan <i>fine-tuning</i></p> | <p>menunjukkan bahwa pendekatan <i>deep learning</i> ini efektif dalam memproses citra tipis (<i>thin section</i>) dan memberikan hasil konsisten pada dataset uji.</p> | <p>petrografi tradisional</p> <p>Kelemahan: Ukuran dataset yang kecil sehingga berisiko menimbulkan <i>overfitting</i> dan membatasi generalisasi model. Selain itu, kualitas citra sangat memengaruhi akurasi, dan sifat <i>black-box</i> CNN membuat hasil klasifikasi sulit untuk dijelaskan secara geologis.</p> <p>Penelitian ini juga membutuhkan lebih banyak variasi data untuk meningkatkan</p> | <p>teknik validasi silang pada data pelatihan</p> |

| No | Judul Penelitian | Nama Peneliti, Tahun, dan Index | Metode Penelitian | Hasil | Keunggulan dan Kelemahan | Perbandingan |
|----|---|---------------------------------------|--|---|--|---|
| | | | | | <i>robust-ness</i> model. | |
| 5 | <i>Evaluation of machine learning methods for lithology classification using geophysical data</i> | Bressan et al., 2020. Q2 Scopus | Penelitian ini menerapkan empat algoritma <i>machine learning</i> (<i>Decision Tree</i> , <i>Random Forest</i> , dan <i>SVM</i>) | <i>Random Forest</i> memberikan hasil terbaik pada sebagian besar template dengan akurasi di atas 80% | Keunggulan: Metode <i>machine learning</i> terbukti mampu mempercepat dan meningkatkan akurasi identifikasi litologi sehingga membantu geolog dalam interpretasi data secara lebih efisien. Kekurangan: Hasil model sangat dipengaruhi oleh variasi kualitas data | Penelitian ini menggunakan dataset berjumlah 1.632 gambar yang terbagi ke dalam 5 kelas, menerapkan metode <i>ensemble stacking</i> dengan 4 model, serta menggunakan |

| No | Judul Penelitian | Nama Peneliti, Tahun, dan Index | Metode Penelitian | Hasil | Keunggulan dan Kelemahan | Perbandingan |
|----|--|---------------------------------------|--|--|---|--|
| | | | | | setiap ekspedisi, sehingga beberapa algoritma menunjukkan hasil kurang stabil dan kemampuan generalisasi antar wilayah geologi masih terbatas. | teknik validasi silang pada data pelatihan |
| 6 | <i>Enhancing the Classification and Identification of Natural Rocks using SwinTransformer Architecture</i> | Bera, 2022. Master's thesis, | Penelitian ini menggunakan empat model <i>deep learning</i> (VGG-19, Inception-V3, | Model <i>Swin-Transformer</i> memberikan hasil terbaik dengan akurasi 73.10%, <i>precision</i> 0.8036, dan <i>recall</i> | Kelebihan: Model <i>Swin-Transformer</i> memberikan akurasi tertinggi dalam klasifikasi citra batuan karena mampu mengekstraksi fitur secara lebih efektif dibandingkan CNN konvensional Kekurangan: Hasil model masih | Penelitian ini menggunakan dataset berjumlah 1.632 gambar yang terbagi ke dalam 5 kelas, menerapkan metode <i>ensemble</i> |

| No | Judul Penelitian | Nama Peneliti, Tahun, dan Index | Metode Penelitian | Hasil | Keunggulan dan Kelemahan | Perbandingan |
|----|------------------|---------------------------------------|--|---|---|---|
| | | | Custom CNN, dan <i>Swin-Transformer</i>) | 0,6430, jauh mengungguli Inception-V3, VGG- 19, dan Custom CNN yang gagal mencapai akurasi tinggi akibat keterbatasan dataset | dibatasi oleh ukuran dataset yang kecil sehingga mengurangi kemampuan generalisasi model ke data batuan yang lebih beragam | <i>stacking</i> dengan 4 model, serta menggunakan teknik validasi silang pada data pelatihan |

2.3. Landasan Teori

2.3.1. Pengolahan Citra Digital

Pengolahan citra digital merupakan teknik yang digunakan untuk mengolah dan menganalisis gambar dalam format digital untuk memperoleh informasi yang lebih bermakna. Secara umum, pengolahan citra digital terdiri dari beberapa tahapan utama yang saling berhubungan. Tahapan-tahapan tersebut meliputi *preprocessing*, ekstraksi fitur, segmentasi, dan klasifikasi. Masing-masing tahapan ini memiliki peran penting dalam memastikan bahwa informasi yang dihasilkan dari citra digital dapat diinterpretasikan dengan baik.

Tahapan *preprocessing* bertujuan untuk meningkatkan kualitas citra sebelum dilakukan analisis lebih lanjut. Langkah-langkah dalam *preprocessing* meliputi penyesuaian warna, normalisasi, serta augmentasi data untuk meningkatkan variasi dalam dataset. Teknik augmentasi, seperti rotasi, *flipping*, dan perubahan kontras, digunakan untuk memperkaya dataset dan membantu model dalam mengenali pola dengan lebih baik.

Setelah citra melalui tahap *preprocessing*, langkah berikutnya adalah ekstraksi fitur. Ekstraksi fitur merupakan proses mengambil karakteristik penting dari citra yang dapat digunakan untuk membedakan satu objek dengan objek lainnya. Beberapa metode yang digunakan dalam ekstraksi fitur antara lain deteksi tepi, analisis tekstur, serta transformasi domain frekuensi.

Tahap selanjutnya adalah segmentasi data, yakni dengan membagi citra menjadi beberapa bagian yang lebih kecil agar objek atau fitur tertentu dapat dikenali dengan lebih baik. Teknik segmentasi yang umum digunakan meliputi *thresholding*, *edge*

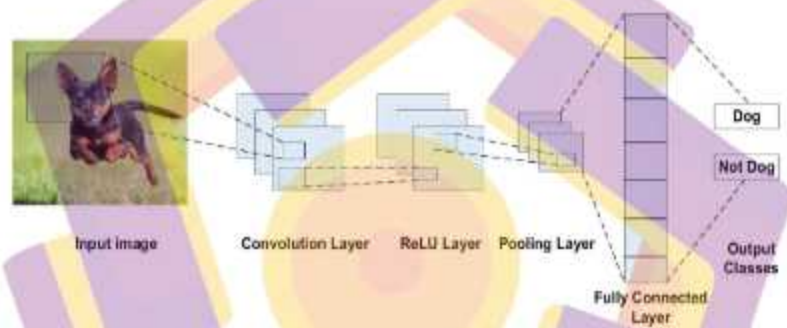
detection, dan algoritma berbasis *region-growing*. Pada klasifikasi batuan, segmentasi digunakan untuk memisahkan bagian batuan dari latar belakang agar fitur batuan lebih mudah dianalisis.

Tahapan akhir dalam pengolahan citra ini adalah proses klasifikasi, yaitu proses pengelompokan citra berdasarkan fitur yang telah diekstraksi sebelumnya. Algoritma klasifikasi yang sering digunakan meliputi metode berbasis pembelajaran mesin tradisional, seperti *Random Forest* dan *Support Vector Machine (SVM)*, serta metode berbasis *deep learning*, seperti *Convolutional Neural Network (CNN)*. CNN lebih unggul dalam klasifikasi citra karena mampu secara otomatis mengekstrak fitur dari gambar melalui lapisan konvolusi dan *pooling*.

2.3.2. Convolutional Neural Network (CNN)

Convolutional Neural Networks (CNN) merupakan jenis jaringan saraf tiruan yang dirancang khusus untuk mengolah data berbentuk grid, seperti gambar (Xu et al., 2022). CNN bekerja dengan mengaplikasikan operasi konvolusi pada data input untuk menghasilkan fitur yang relevan. Proses ini melibatkan penggunaan filter atau *kernel* yang bergerak di atas gambar input untuk mendeteksi pola-pola dasar seperti tepi, sudut, dan tekstur. Lapisan konvolusi diikuti oleh lapisan *pooling* yang mereduksi dimensi fitur untuk mengurangi kompleksitas dan meningkatkan efisiensi model. Akhirnya, lapisan *fully connected* digunakan untuk mengklasifikasikan fitur yang telah diekstraksi ke dalam kategori tertentu. Terdapat 4 komponen utama dalam arsitektur CNN. Secara visual komponen tersebut dapat dilihat pada Gambar 2.1 dan penjelasan dari masing-masing komponen tersebut adalah sebagai berikut:

- a. Lapisan konvolusi: menerapkan filter untuk mendeteksi fitur spesifik.
- b. Lapisan *pooling*: mengurangi ukuran peta fitur dengan teknik seperti *max pooling* atau *average pooling*.
- c. Lapisan normalisasi: menstabilkan proses pelatihan dengan menormalkan aktivasi neuron.
- d. Lapisan *fully connected*: menghubungkan semua neuron dari lapisan sebelumnya untuk klasifikasi akhir.



Gambar 2.1. Struktur Dasar dari CNN

Sumber: (Alzubaidi et al., 2021)

2.3.3. Arsitektur CNN

AlexNet merupakan salah satu arsitektur CNN yang membawa perubahan besar dalam pengolahan citra digital. Struktur Alexnet yang tergolong sederhana, menjadikan model yang cukup mudah dipahami dan diimplementasikan. Namun, ukuran model yang besar membuatnya kurang efisien untuk dijalankan pada sistem dengan keterbatasan sumber daya komputasi (H. C. Chen et al., 2022). Keterbatasan ini menjadi perhatian utama ketika aplikasi *deep learning* mulai diarahkan untuk penggunaan di perangkat *edge* atau *mobile*.

Sebagai respons terhadap kebutuhan akan model yang lebih ringan dan efisien, dikembangkanlah model MobileNetV1 dan MobileNetV2. Kedua model ini dirancang untuk memberikan efisiensi komputasi tinggi dengan ukuran parameter yang jauh lebih kecil, sehingga sangat ideal digunakan pada perangkat seluler atau sistem dengan GPU terbatas (Sousa et al., 2020). Meskipun akurasi tidak setinggi model-model besar, kompromi antara kecepatan, ukuran, dan ketepatan menjadikan MobileNet salah satu arsitektur yang banyak diadopsi dalam konteks aplikasi nyata dengan keterbatasan perangkat keras.

Selanjutnya, arsitektur seperti EfficientNet, ResNet-50, dan GoogLeNet (InceptionV3) menunjukkan hasil yang baik dalam meningkatkan kinerja tanpa mengorbankan efisiensi (Geetha et al., 2023). EfficientNet menawarkan keseimbangan antara akurasi dan jumlah parameter dengan pendekatan scaling yang sistematis. ResNet-50 memperkenalkan konsep *residual connection* untuk memungkinkan pelatihan jaringan yang sangat dalam tanpa mengurangi akurasi. GoogLeNet menggunakan strategi multi-skala dengan berbagai ukuran filter dalam satu blok jaringan untuk menangkap fitur secara lebih menyeluruh. Masing-masing arsitektur ini memberikan kontribusi penting terhadap pengembangan metode klasifikasi citra yang lebih kuat, efisien, dan adaptif terhadap kompleksitas data.

2.3.4. Ensemble Learning dan Stacking

Ensemble learning adalah pendekatan dalam pembelajaran mesin yang menggabungkan beberapa model untuk meningkatkan hasil prediksi (Ouedraogo et al., 2023). Salah satu teknik *ensemble* yang efektif adalah *stacking*, yaitu metode yang memadukan prediksi dari beberapa model dasar (*base learners*) dan

menggabungkannya melalui model *meta-learner* (Ture et al., 2024). Teknik ini memungkinkan sistem pembelajaran memanfaatkan kelebihan setiap arsitektur sambil mengurangi kelemahannya.

Penerapan metode *stacking* dalam klasifikasi jenis batuan berbasis citra foto, menjadi solusi strategis karena jenis batu memiliki ciri visual yang sangat beragam, seperti corak, tekstur, dan warna. Kombinasi antara CNN konvensional yang andal dalam ekstraksi fitur dasar dan EfficientNet yang efisien serta akurat memungkinkan sistem untuk bekerja lebih *robust*. Penelitian yang dilakukan oleh (Ouedraogo et al., 2023) menunjukkan bahwa metode *stacking* dapat meningkatkan akurasi 2-3% klasifikasi citra limbah TPA dibandingkan penggunaan model tunggal, terutama ketika data dilatih menggunakan citra lapangan dengan *noise* dan pencahayaan yang bervariasi.

2.3.5. Evaluasi Model

Evaluasi model dilakukan dengan menggunakan metrik seperti akurasi, *precision*, *recall*, dan *F1 score* untuk menilai kinerja model pada subset data pengujian. Adapun rumus yang dipakai untuk menghitung adalah sebagai berikut:

- Accuracy* merupakan rasio prediksi benar (positif dan negatif) dengan keseluruhan data. $Accuracy = (TP + TN) / (TP + FP + FN + TN)$
- Precision* merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif. $Precision = (TP) / (TP + FP)$
- Recall* merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif. $Recall = (TP) / (TP + FN)$

d. *F1 Score* merupakan perbandingan rata-rata presisi dan *recall* yang dibobotkan.

$$F1\ Score = 2 * (Recall * Precision) / (Recall + Precision)$$

Confusion matrix juga digunakan untuk menganalisis distribusi kesalahan klasifikasi dan memahami kinerja model dalam mengenali setiap jenis batuan. *Confusion matrix* merupakan sebuah informasi mengenai prediksi dan aktual dari suatu sistem klasifikasi yang memiliki data jawaban benar (*supervised*). *Confusion matrix* menggunakan data dalam bentuk matriks (Zeng, 2020). Terdapat 4 jenis isi yang terdapat dalam tabel *confusion matrix* antara lain:

1. *True Positive (TP)* adalah kondisi dimana model memprediksi data sebagai ya (*TRUE*) dan jawaban aktualnya adalah ya (*TRUE*).
2. *True Negative (TN)* adalah kondisi dimana model memprediksi data sebagai tidak (*FALSE*) dan jawaban aktualnya adalah tidak (*FALSE*).
3. *False Positive (FP)* adalah kondisi dimana model memprediksi data sebagai ya (*TRUE*) dan jawaban aktualnya adalah tidak (*FALSE*).
4. *False Negative (FN)*, kondisi dimana model memprediksi data sebagai tidak (*FALSE*) dan jawaban aktualnya adalah ya (*TRUE*).

| | | Actual Values | |
|------------------|--------------|---------------|--------------|
| | | Positive (1) | Negative (0) |
| Predicted Values | Positive (1) | TP | FP |
| | Negative (0) | FN | TN |

Gambar 2.2. Tabel *Confusion Matrix*

Sumber: (Atmajaya, et all., 2023)

BAB III

METODE PENELITIAN

3.1. Jenis, Sifat, dan Pendekatan Penelitian

Penelitian ini merupakan penelitian eksperimental yang bertujuan untuk mengevaluasi hasil klasifikasi citra batuan menggunakan pendekatan *ensemble learning* berbasis model *stacking* CNN. Pendekatan ini mengintegrasikan berbagai jenis arsitektur CNN yang terbagi menjadi beberapa kategori model *stacking*, yakni *stacking* model 1 (MobileNetV2 dan EfficientNetB0), model 2 (VGG16 dan DenseNet121), model 3 (ResNet50 dan InceptionV3), dan *stacking* model 4 (ResNet50 dengan MobileNetV2). Eksperimen dilakukan untuk menilai efektivitas masing-masing kategori model *stacking* dalam meningkatkan akurasi dan efisiensi sistem klasifikasi citra serta membandingkannya dengan model atau arsitektur CNN tunggal.

Penelitian ini bersifat deskriptif, eksplanatif, dan komparatif. Deskriptif karena menggambarkan penerapan berbagai arsitektur CNN dalam proses klasifikasi citra batuan. Eksplanatif karena menjelaskan bagaimana kontribusi masing-masing model, baik yang ringan, kuat, maupun kombinasi terhadap kinerja akhir dari sistem *ensemble*. Komparatif karena membandingkan hasil klasifikasi antara model tunggal dan model *stacking* untuk menentukan pendekatan yang paling efektif.

Pendekatan penelitian ini melibatkan serangkaian eksperimen yang mencakup pelatihan model CNN secara individual, penerapan metode *stacking* dalam 4 (empat) konfigurasi model, serta evaluasi hasil klasifikasi menggunakan metrik akurasi, *precision*, *recall*, dan *F1-score*. Hasil dari model *stacking* dibandingkan dengan model

tunggal untuk menilai sejauh mana teknik *ensemble stacking* ini mampu mengatasi keterbatasan arsitektur tertentu dan meningkatkan hasil akurasi klasifikasi secara keseluruhan.

3.2. Alat dan Perangkat Penelitian

Penelitian ini dilaksanakan menggunakan platform Google Colab sebagai lingkungan komputasi berbasis *cloud* yang mendukung pemrosesan *deep learning*. Bahasa pemrograman yang digunakan adalah Python dengan dukungan *framework TensorFlow* dan *Keras* untuk membangun serta melatih model *Convolutional Neural Network (CNN)* dan teknik *ensemble stacking*. Lingkungan *runtime* Colab berjalan pada sistem operasi Linux (*Ubuntu* berbasis *cloud*) dengan spesifikasi perangkat keras meliputi GPU NVIDIA Tesla T4 berkapasitas 16 GB VRAM, RAM tipe High-RAM dengan kapasitas sekitar 25-51 GB, serta CPU virtual Intel Xeon.

3.3. Data, Metode Pengumpulan dan Pengolahan Data

Penelitian ini memanfaatkan data citra batuan yang diperoleh dari tim riset Brac University Mars Rover dan diakses secara terbuka melalui *platform* Kaggle (Hossain et al., 2021). Dataset tersebut berisi 1.632 gambar yang dikategorikan ke dalam lima jenis batuan, yaitu sandstone (326 gambar), slate (300 gambar), marble (308 gambar), basalt (404 gambar), dan granite (294 gambar). Proses pengumpulan data dilakukan dengan mengunduh dataset tersebut, kemudian dilanjutkan dengan eksplorasi awal melalui *Exploratory Data Analysis (EDA)* untuk menilai sebaran kelas dan kualitas gambar.

Tahapan selanjutnya adalah pra-pemrosesan data, yang mencakup pembersihan gambar dari data duplikat, rusak, atau tanpa label. Pemeriksaan dilakukan untuk memastikan bahwa semua gambar memiliki kualitas dan label yang benar. Seluruh gambar kemudian diubah ukurannya menjadi 224×224 piksel agar seragam dan sesuai dengan input yang dibutuhkan oleh model CNN. Dataset dibagi menjadi dua bagian utama, yaitu 80% untuk pelatihan (*training*) dan 20% untuk pengujian (*testing*). Data pelatihan tersebut selanjutnya digunakan dalam proses pelatihan model dengan menerapkan *Stratified K-Fold Cross-Validation* sebanyak 5 fold, yang memungkinkan evaluasi model dilakukan secara lebih menyeluruh dan adil terhadap seluruh data pelatihan.

Untuk meningkatkan variasi data dan mengurangi risiko *overfitting*, dilakukan proses augmentasi khusus pada data pelatihan. Augmentasi ini diterapkan secara dinamis menggunakan *ImageDataGenerator* dengan teknik transformasi seperti rotasi acak hingga 10 derajat, pergeseran horizontal dan vertikal hingga 10% dari ukuran gambar, zoom acak sebesar 5%, serta pembalikan gambar secara horizontal. Setiap transformasi dilengkapi dengan pengisian area kosong menggunakan metode *nearest neighbor*. Proses augmentasi ini dilakukan hanya pada data training di setiap fold, sedangkan data validasi dan data uji dibiarkan tanpa augmentasi agar evaluasi hasil model tetap objektif. Dengan pendekatan ini, model diharapkan mampu belajar dari data yang lebih beragam tanpa mengorbankan validitas evaluasi.

3.4. Metode Analisis Data

Analisis data pada penelitian ini dimulai dengan proses pra-pemrosesan, yaitu mengubah ukuran citra batuan menjadi 224×224 piksel agar sesuai dengan kebutuhan berbagai arsitektur CNN, seperti ResNet50, MobileNetV2, InceptionV3, VGG16, dan EfficientNetB0. Untuk memperluas variasi data dan mengurangi kemungkinan *overfitting*, diterapkan teknik augmentasi citra berupa rotasi acak, pergeseran, *zoom*, dan pembalikan horizontal. Selanjutnya, dataset dilatih menggunakan pendekatan *Stratified K-Fold Cross-Validation* sebanyak 5 *fold*, yang membagi data pelatihan menjadi lima subset dengan distribusi label yang seimbang. Teknik ini memastikan bahwa setiap data digunakan sebagai data validasi sebanyak satu kali dan sebagai data latihan sebanyak empat kali, sehingga evaluasi model menjadi lebih menyeluruh dan *robust*.

Selanjutnya, dilakukan ekstraksi fitur dari citra menggunakan beberapa model CNN yang telah dilatih sebelumnya (*pretrained*), lalu hasil fitur dari dua model dalam setiap konfigurasi digabungkan melalui metode *stacking*. Konfigurasi *stacking* ini dikelompokkan menjadi 4 kategori model. Hasil penggabungan fitur digunakan sebagai masukan untuk model *Support Vector Machine (SVM)* sebagai klasifikator utama.

Evaluasi hasil klasifikasi model dilakukan dengan menghitung akurasi, *precision*, *recall*, dan *F1-score*, serta divisualisasikan menggunakan *confusion matrix*. Analisis ini membantu mengevaluasi kemampuan model dalam mengenali setiap jenis litologi, sekaligus membandingkan efektivitas masing-masing konfigurasi *stacking*.

3.5. Alur Penelitian

3.5.1. Pengumpulan dataset

Tahap awal dimulai dengan pengumpulan dataset citra batuan yang diperoleh dari tim riset Brac University Mars Rover dan tersedia secara publik di platform KAGGLE (Hossain et al., 2021). Dataset ini dikategorikan menjadi 5 (lima) kelas batuan, yaitu: basalt, granite, marble, slate, dan sandstone. Setiap citra disimpan dalam struktur direktori sesuai label masing-masing. Untuk memahami sebaran data, dilakukan proses *Exploratory Data Analysis (EDA)* guna mengevaluasi proporsi dan keberagaman jumlah data pada tiap kelas.

3.5.2. Pra-pemrosesan dan pengolahan data

a. *Import library*

Beberapa *library* yang digunakan antara lain: TensorFlow, yang digunakan untuk membangun dan melatih model CNN; Scikit-learn, yang dimanfaatkan untuk evaluasi model, penerapan algoritma *machine learning* seperti SVM sebagai *meta-learner* pada arsitektur *stacking*. Selain itu, Matplotlib dan Seaborn digunakan untuk visualisasi grafik, seperti *training curves*, dan *confusion matrix*.

b. Konfigurasi dataset dan parameter

Parameter utama dalam proses pelatihan model meliputi: CLEAN_PATH: path direktori dataset, IMG_SIZE: 224x224 pixel, BATCH_SIZE: 32, SEED: 42, EPOCHS: 10,

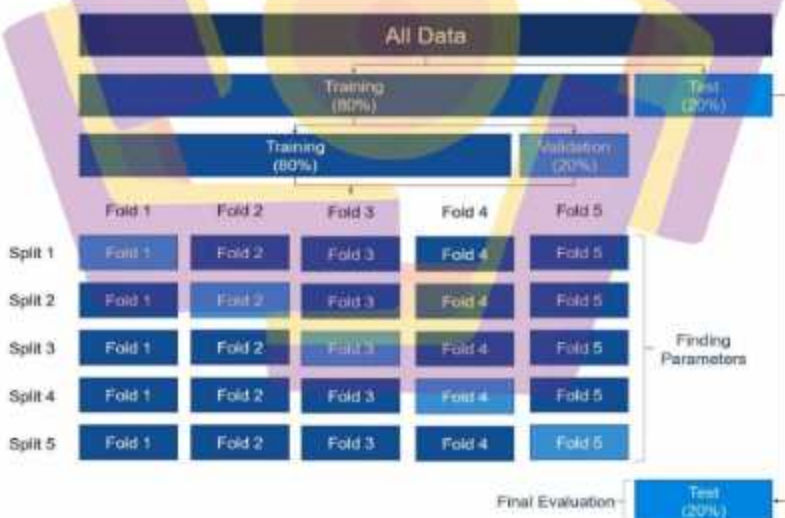
c. Augmentasi data

Augmentasi data yang dilakukan meliputi: rotasi acak hingga 10 derajat, pergeseran horizontal dan vertikal hingga 10% dari ukuran gambar, *zoom* acak sebesar

5%, serta membalikkan gambar secara horizontal. Setiap transformasi dilengkapi dengan pengisian area kosong menggunakan metode *nearest neighbor*. Metode augmentasi menggunakan augmentasi *on-the-fly* saat training.

d. *Load label* dan generator

Melalui *flow_from_directory*, sistem secara otomatis memuat label berdasarkan struktur direktori. Namun, alih-alih menggunakan *validation_split*, proses pelatihan dan validasi kini dilakukan menggunakan teknik *Stratified K-Fold Cross Validation*, yang membagi data pelatihan menjadi lima lipatan (*fold*) dengan distribusi label yang seimbang pada setiap *fold*. Pendekatan ini memungkinkan evaluasi model yang lebih stabil dan mengurangi bias terhadap pembagian data tertentu. Secara detail alur pembagiannya dapat dilihat pada Gambar 3.1.



Gambar 3.1. Proses *Stratified K-Fold Cross Validation* Data

Sumber: (Dawson et al., 2023)

3.5.3. Proses modeling dan analisis data

1. Model tunggal

Model tunggal ini memanfaatkan 6 (enam) arsitektur CNN *pretrained* yang populer, yaitu: ResNet50, InceptionV3, VGG16, DenseNet121, MobileNetV2, dan EfficientNetB0. Masing-masing model diterapkan dalam dua tahap:

- a) *Transfer learning*: hanya melatih layer atas (*fully connected*), sedangkan *backbone* dibekukan.
- b) *Fine-tuning*: membuka sebagian layer bawah untuk dilatih ulang agar lebih sesuai dengan dataset.
- c) Model dievaluasi menggunakan: *Classification report* (akurasi, presisi, *recall*, *f1-score*), dan *confusion matrix*

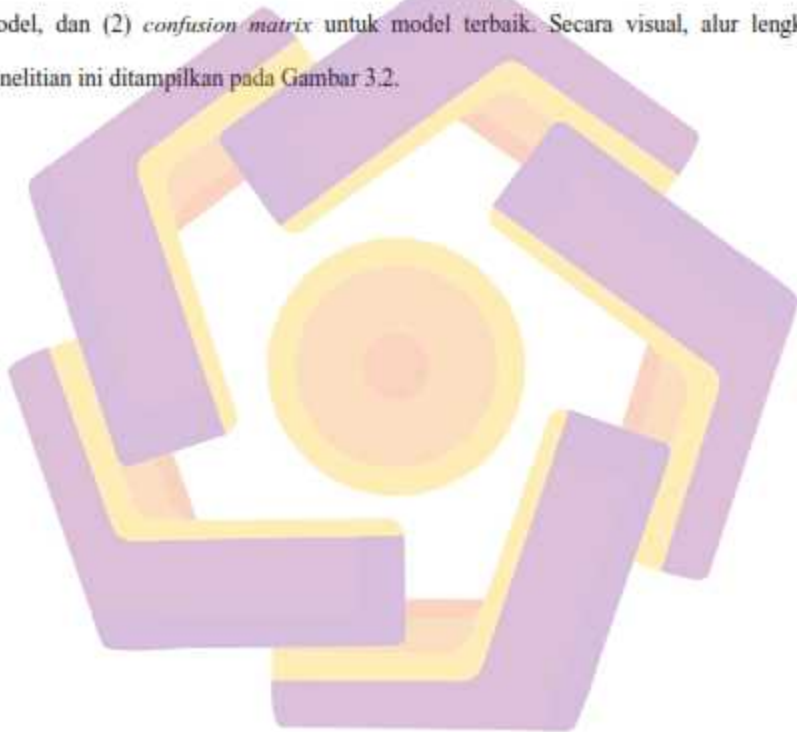
2. Model *stacking*

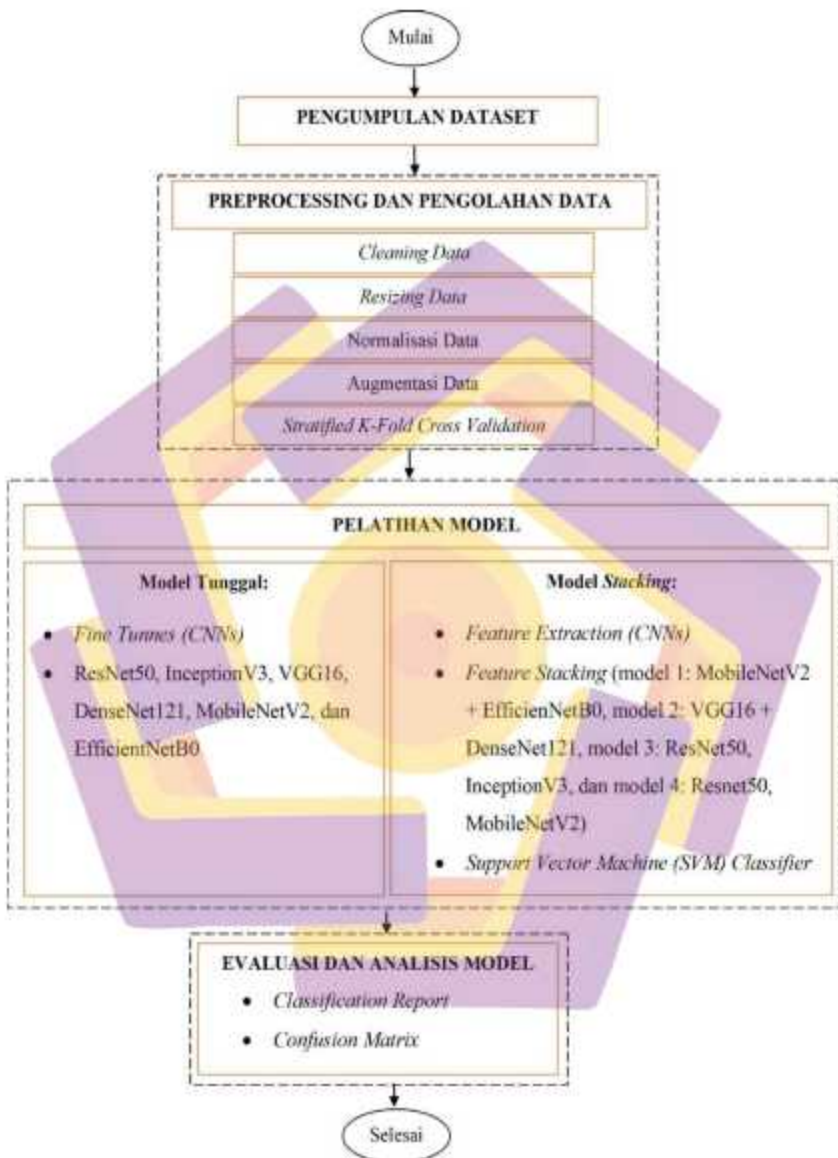
Pendekatan *stacking* dilakukan dengan tahapan berikut:

- a) Ekstraksi fitur menggunakan *pretrained CNN* tanpa *top-layer*, lalu hasil *feature map* diringkas menggunakan *GlobalAveragePooling2D*
- b) Gabungan fitur dengan menggabungkan beberapa fitur dari dua model CNN per konfigurasi (model 1, model 2, model 3, model 4).
- c) Standardisasi menggunakan *StandardScaler* untuk menormalkan vektor fitur.
- d) Klasifikasi menggunakan algoritma *Support Vector Machine (SVM)* dengan *kernel RBF* sebagai *meta-learner* dalam arsitektur *stacking*.
- e) Evaluasi dilakukan terhadap hasil prediksi dari setiap konfigurasi *meta-learner*, mencakup: *classification report*, dan *confusion matrix*.

3.4.4. Evaluasi dan perbandingan akhir

Pada tahap ini, seluruh model *stacking* dan *non-stacking* dibandingkan menggunakan metrik akurasi, precision, recall, F1-score, dan *confusion matrix*. Model dengan nilai akurasi tertinggi ditetapkan sebagai model terbaik untuk tugas klasifikasi citra batuan. Visualisasi hasil akhir meliputi: (1) *bar chart* perbandingan akurasi antar model, dan (2) *confusion matrix* untuk model terbaik. Secara visual, alur lengkap penelitian ini ditampilkan pada Gambar 3.2.





Gambar 3.2. Diagram Alir Penelitian

Sumber: Peneliti, 2025

Berdasarkan pada Gambar 3.2, maka penelitian ini diawali dengan pengumpulan dataset citra batuan yang diperoleh dari tim riset Brac University Mars Rover dan tersedia secara publik di *platform Kaggle* (Hossain et al., 2021). Dataset tersebut terdiri atas lima kelas batuan, yaitu *basalt*, *granite*, *marble*, *slate*, dan *sandstone*. Untuk mempermudah proses pengolahan, citra-citra batuan tersebut diklasifikasikan ke dalam direktori berdasarkan masing-masing kelas.

Langkah berikutnya adalah tahap *preprocessing*, yang meliputi lima prosedur utama. Pertama, dilakukan pembersihan data untuk menghapus citra yang rusak, duplikat, atau tidak relevan. Kedua, seluruh citra diubah ukurannya menjadi 224×224 piksel agar sesuai dengan format input arsitektur CNN. Ketiga, dilakukan normalisasi nilai piksel ke skala 0–1 untuk mempercepat proses pembelajaran model. Keempat, citra diperluas secara buatan melalui augmentasi data seperti rotasi, *flipping*, *zooming*, dan *shifting* guna meningkatkan keberagaman data serta mengurangi risiko *overfitting*. Kelima, dilakukan pembagian data menggunakan *Stratified K-Fold Cross Validation*, agar distribusi kelas tetap seimbang di setiap *fold*.

Kegiatan eksperimen dalam penelitian ini memanfaatkan model CNN tunggal dan model *ensemble stacking*. Pada pendekatan CNN tunggal, digunakan beberapa arsitektur *deep learning pretrained*, yaitu ResNet50, MobileNetV2, InceptionV3, VGG16, dan EfficientNetB0. Semua model dilakukan *fine-tuning* dengan mengaktifkan kembali seluruh layer (*requires_grad=True*) dan menyesuaikan layer klasifikasi terakhir dengan jumlah kelas (5 kelas). Proses pelatihan dilakukan menggunakan *optimizer Adam* dengan *learning rate* sebesar 0.0001, serta fungsi *loss CrossEntropyLoss* yang sesuai untuk klasifikasi multikategori.

Sementara itu, pendekatan *ensemble stacking* dalam penelitian ini dibagi menjadi empat kategori berdasarkan kombinasi arsitektur CNN. Ke-empat model tersebut antara lain model 1 (MobileNetV2 dan EfficientNetB0), model 2 (VGG16 dan DenseNet121), model 3 (ResNet50 dan InceptionV3), dan model 4 (ResNet50 dan MobileNetV2). Fitur-fitur hasil ekstraksi dari kombinasi arsitektur CNN tersebut kemudian diklasifikasikan menggunakan algoritma *Support Vector Machine (SVM)* sebagai *meta-learner*.

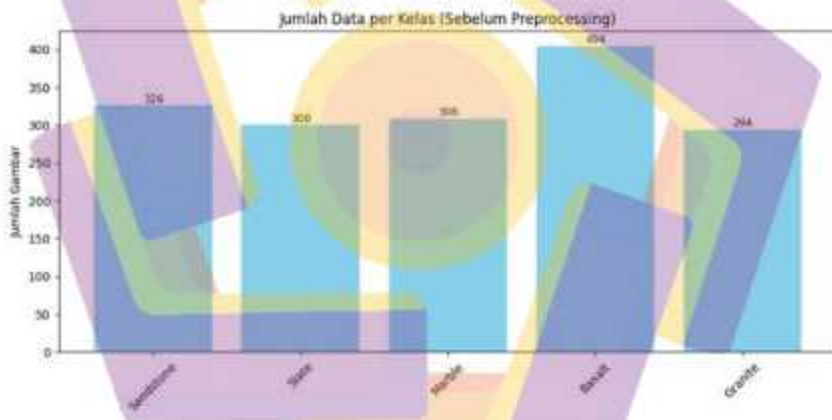
Evaluasi hasil klasifikasi model dilakukan menggunakan metrik evaluasi umum seperti akurasi, *precision*, *recall*, dan *F1-score*, serta divisualisasikan menggunakan *confusion matrix*. Evaluasi dilakukan pada data pelatihan dan data pengujian untuk memperoleh gambaran menyeluruh terhadap kinerja masing-masing model. Hasil evaluasi tersebut kemudian dianalisis secara deskriptif kuantitatif dan komparatif untuk membandingkan efektivitas pendekatan *ensemble stacking* terhadap model tunggal. Analisis ini menjadi dasar dalam menentukan model terbaik yang mampu memberikan hasil klasifikasi akurat dan stabil dalam klasifikasi citra batuan.

BAB IV

HASIL PENELITIAN DAN PEMBAHASAN

4.1. Exploratory Data Analisis (EDA)

Eksplorasi data awal dilakukan untuk memahami karakteristik suatu dataset citra batuan sebelum dilakukan pemodelan. Dataset yang digunakan dalam penelitian ini meliputi lima kelas batuan, yakni batuan *sandstone*, *slate*, *marble*, *basalt*, dan *granite*. Setiap kelas diperiksa jumlah citranya untuk menilai keseimbangan data. Adapun hasil dari eksplorasi data dapat dilihat pada Gambar 4.1.



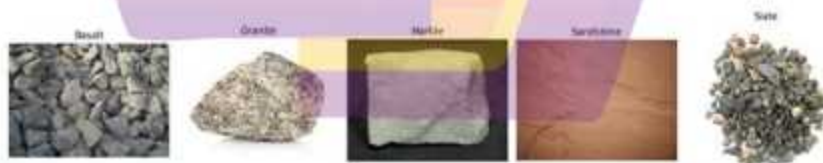
Gambar 4.1. Hasil Exploratory Data Analisis Citra Batuan

Gambar 4.1 menunjukkan bahwa terdapat ketidakseimbangan data citra batuan antar kelas, yakni kelas batuan *Granite* memiliki jumlah citra yang jauh lebih sedikit dan batuan *Quartzite* memiliki jumlah citra batuan yang lebih besar dibandingkan

dengan kelas lainnya. Hal ini menjadi dasar dilakukannya proses augmentasi data pada kelas minor agar distribusi antar kelas menjadi lebih seimbang.

Selain menganalisis distribusi jumlah citra, kegiatan eksplorasi data juga melakukan visualisasi contoh citra dari masing-masing kelas. Visualisasi ini memberikan gambaran awal mengenai perbedaan tekstur dan warna antar kelas batuan, yang sangat penting dalam proses klasifikasi berbasis jenis citra batuan. Citra-citra ini memiliki variasi ukuran dan resolusi, sehingga pada tahap preprocessing seluruh citra diubah ke resolusi yang seragam, yaitu 224x224 piksel, agar sesuai dengan arsitektur model CNN yang digunakan.

Untuk mendukung proses normalisasi, nilai dari statistik pixel dari citra juga dianalisis. Rata-rata nilai intensitas pixel (mean) berada dalam rentang wajar 0.3–0.6 (dalam skala 0–1 setelah normalisasi), dan tidak ditemukan outlier yang signifikan setelah pembersihan data dilakukan. Berdasarkan hasil eksplorasi data yang telah dilakukan, maka dapat disimpulkan bahwa dataset siap untuk digunakan dalam tahap pelatihan model klasifikasi dengan terlebih dahulu dilakukan *preprocessing* terhadap dataset batuan tersebut.



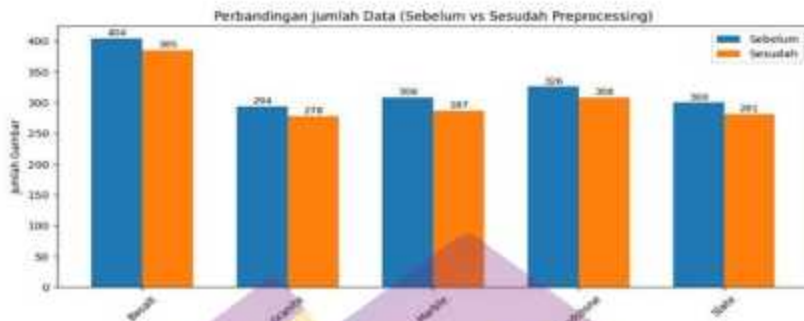
Gambar 4.2. Contoh Dataset Batuan

4.2. Hasil Preprocessing Data

Tahap *preprocessing* data dalam penelitian ini difokuskan pada upaya peningkatan kualitas data sebelum digunakan dalam pelatihan model. Langkah pertama yang dilakukan adalah pembersihan dataset, yang mencakup proses identifikasi dan penghapusan citra yang rusak serta citra yang duplikat. Gambar yang dikategorikan rusak biasanya tidak dapat dibuka oleh sistem akibat kerusakan file atau format yang tidak lengkap. Sedangkan gambar duplikat dideteksi melalui pencocokan nilai *hash* (MD5) dari setiap file, sehingga hanya citra unik yang dipertahankan untuk proses selanjutnya.

Setelah data berhasil dibersihkan, seluruh citra kemudian diubah ukurannya menjadi 224×224 piksel, agar sesuai dengan format input standar pada arsitektur *deep learning* seperti MobileNetV2 dan EfficientNetB0. Selanjutnya dilakukan normalisasi nilai piksel ke dalam skala $[0,1]$, yang bertujuan untuk mempercepat proses pelatihan dan membantu model mencapai konvergensi yang lebih stabil.

Distribusi jumlah citra per kelas sebelum dan sesudah tahap *preprocessing* divisualisasikan dalam Gambar 4.3. Visualisasi ini memperlihatkan proporsi data yang tersedia untuk masing-masing kelas setelah proses pembersihan dilakukan. Meskipun pada tahap ini belum dilakukan augmentasi, hasil pembersihan ini menjadi dasar penting untuk menjaga kualitas dataset dan memastikan bahwa model tidak dilatih dengan data yang redundan atau bermasalah.



Gambar 4.3. Perbandingan Jumlah Dataset Sebelum dan Sesudah *Preprocessing*

Kegiatan augmentasi data dalam penelitian ini tidak termasuk dalam tahap *preprocessing*, melainkan dilakukan secara dinamis (*on-the-fly*) selama proses pelatihan khusus pada data *training*. Teknik augmentasi yang digunakan mencakup rotasi acak, *flipping horizontal*, dan *jitter* warna, yang diterapkan secara acak pada setiap *batch* data. Pendekatan ini bertujuan untuk meningkatkan keragaman data tanpa memperbesar ukuran dataset secara permanen, serta membantu mengurangi *overfitting* dan meningkatkan generalisasi model.

4.3. Pembagian Dataset

Pembagian data dalam penelitian ini menggunakan pendekatan *Stratified K-Fold Cross Validation*, yaitu metode validasi silang yang memastikan proporsi kelas tetap seimbang pada setiap *fold*. Teknik ini membagi data ke dalam beberapa lipatan (*fold*), di mana masing-masing subset bergantian berperan sebagai data pelatihan dan validasi. Dengan demikian, seluruh data digunakan secara menyeluruh dalam proses pelatihan dan evaluasi, sehingga hasil yang diperoleh lebih representatif dan tidak bias terhadap

kelas tertentu. Pendekatan ini sangat bermanfaat untuk menjaga keberagaman data dan kestabilan hasil klasifikasi model.

Berbeda dengan metode pembagian data tetap seperti 80:20, *cross validation* memungkinkan model untuk dilatih dan divalidasi pada variasi data yang lebih luas, meningkatkan generalisasi model. Rincian jumlah citra untuk masing-masing kelas setelah tahap preprocessing disajikan dalam Gambar 4.3. Sebagai ilustrasi, kelas *basalt* memiliki total 385 citra yang seluruhnya akan digunakan secara bergilir dalam proses pelatihan dan validasi sesuai jumlah *fold* yang ditetapkan.

Untuk memperkuat pemahaman terhadap distribusi awal data, visualisasi dalam bentuk diagram batang (*bar chart*) disajikan pada Gambar 4.3. Grafik ini menampilkan jumlah citra per kelas sebelum dibagi dalam skema *K-Fold*, dan menunjukkan bahwa distribusi data telah seimbang sejak awal. Hal ini memastikan bahwa setiap kelas tetap terwakili secara proporsional di setiap tahap pelatihan dan evaluasi model.

4.4. Hasil Akurasi Model Stacking

Model *stacking* dalam penelitian ini terbagi menjadi empat kategori, yaitu model 1 (MobileNetV2 dan EfficientNetB0), model 2 (VGG16 dan DenseNet121), model 3 (ResNet50 dan InceptionV3), dan model 4 (ResNet50 dan MobileNetV2). Keempat model tersebut berperan sebagai *base learner*, sementara *Support Vector Machine* (SVM) digunakan sebagai meta-learner yang menggabungkan representasi fitur dari masing-masing arsitektur untuk menghasilkan keputusan akhir. Pendekatan *stacking* seperti ini diharapkan mampu meningkatkan kemampuan generalisasi melalui kombinasi fitur yang lebih kaya dan beragam.

Untuk menilai efektivitas setiap kombinasi dari 4 (empat) model *stacking* yang dikembangkan, dilakukan pengukuran akurasi menggunakan skema *cross-validation* dan pengujian pada data uji terpisah, serta pencatatan ukuran file model yang dihasilkan. Ketiga aspek ini digunakan sebagai indikator awal untuk memahami konsistensi hasil prediksi, kemampuan generalisasi, dan tingkat kompleksitas masing-masing model. Secara detail mengenai hasil akurasi dan ukuran file yang dihasilkan model dapat dilihat pada Tabel 4.1. Sementara itu untuk visualisasi akurasi pada model *stacking* dapat dilihat pada Gambar 4.4.

Tabel 4.1. Hasil Akurasi pada Model *Stacking*

| Model | <i>Cross Validation Accuracy</i> | <i>Test Accuracy</i> | Ukuran File (mb) |
|--|----------------------------------|----------------------|------------------|
| Model 1: <i>base learner</i> (MobileNetV2 + EfficientNetB0), <i>meta learner</i> (SVM) | 0,80 | 0,83 | 22,18 |
| Model 2: <i>base learner</i> (VGG16 + DenseNet121), <i>meta learner</i> (SVM) | 0,84 | 0,81 | 13,08 |
| Model 3: <i>base learner</i> (ResNet50, InceptionV3), <i>meta learner</i> (SVM) | 0,80 | 0,83 | 35,92 |
| Model 4: <i>base learner</i> (Resnet50, MobileNetV2), <i>meta learner</i> (SVM) | 0,81 | 0,84 | 28,89 |

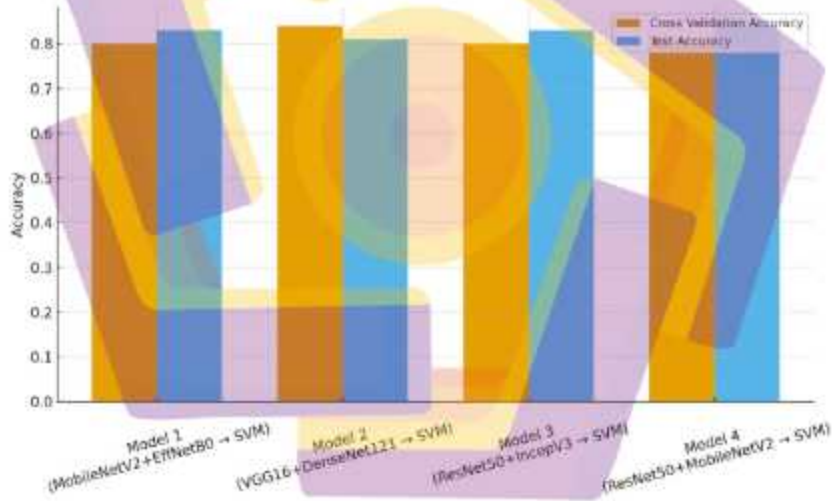
Sumber: Pengolahan Data, 2025

Berdasarkan pada Tabel 4.1, dapat diketahui bahwa terdapat variasi menarik antara nilai akurasi *cross-validation* dan akurasi pengujian pada empat model *stacking* yang dibangun. Model 1 dan Model 3 masing-masing memperoleh akurasi *cross-validation* sebesar 0,80 dan menghasilkan akurasi pengujian sebesar 0,83. Pola ini mengindikasikan bahwa kedua model tersebut memiliki kemampuan generalisasi yang baik, dimana hasil akurasi justru meningkat pada data uji. Fenomena serupa juga terlihat pada Model 4, yang mencapai akurasi *cross-validation* sebesar 0,81 dan meningkat menjadi 0,84 pada data uji. Model 4 menunjukkan akurasi pengujian tertinggi serta konsistensi hasil yang baik, sehingga dapat dinilai sebagai model *stacking* yang paling unggul dalam penelitian ini. **Berbeda dengan tiga model tersebut, Model 2** menunjukkan pola yang berlawanan, yaitu akurasi *cross-validation* paling tinggi (0,84) tetapi menurun pada data uji menjadi 0,81. Penurunan ini mengindikasikan adanya potensi *overfitting*, di mana model terlalu menyesuaikan diri dengan data pelatihan sehingga kurang mampu menangani variasi baru pada data uji.

Berdasarkan pada aspek ukuran filenya, keempat model menunjukkan perbedaan kompleksitas yang cukup mencolok. Model 2 memiliki ukuran file terkecil, yaitu 13,08 MB, sehingga menjadi model yang paling efisien dari sisi penggunaan sumber daya. Namun, efisiensi ini tidak sejalan dengan kestabilan hasil akurasi pada data uji. Sebaliknya, Model 3 memiliki ukuran file terbesar, yaitu 35,92 MB, yang menunjukkan kompleksitas tinggi pada arsitektur *base learner*-nya. Walaupun akurasinya cukup baik, ukuran file yang besar dapat menjadi kendala untuk implementasi pada perangkat dengan keterbatasan memori. Model 1 dan Model 4 berada pada kategori ukuran

menengah, masing-masing sebesar 22,18 MB dan 28,89 MB, dan keduanya menunjukkan akurasi uji yang lebih stabil dibanding Model 2.

Secara keseluruhan, hasil ini memperlihatkan bahwa ukuran file model tidak selalu berkorelasi langsung dengan tingkat akurasi, sehingga pemilihan model terbaik perlu mempertimbangkan kombinasi antara akurasi, kemampuan generalisasi, dan kompleksitas arsitektur. Model 4 (ResNet50 + MobileNetV2) memberikan keseimbangan terbaik dan menjadi kandidat paling layak untuk tugas klasifikasi citra batuan.



Gambar 4.4. Perbandingan Hasil Akurasi Model *Stacking*

Berdasarkan pada Gambar 4.4, dapat diketahui bahwa akurasi *cross-validation* dan akurasi pengujian untuk keempat model berada pada kisaran 0,80-0,84. Nilai ini memberikan gambaran mengenai kemampuan masing-masing model dalam menangani

data baru secara konsisten. Model 4 (ResNet50 + MobileNetV2) memiliki akurasi pengujian tertinggi, yaitu 0,84, sehingga dapat dianggap sebagai model dengan generalisasi paling baik di antara konfigurasi yang diuji. Pada sisi lain, Model 2 menunjukkan selisih yang lebih besar antara akurasi *cross-validation* (0,84) dan akurasi pengujian (0,81), yang mengindikasikan kecenderungan hasil yang kurang stabil pada data uji. Sementara itu, Model 1 dan Model 3 memperlihatkan tingkat konsistensi yang lebih seimbang antara kedua metrik tersebut. Secara keseluruhan, grafik ini menunjukkan bahwa setiap model memiliki kekuatan masing-masing, namun kombinasi arsitektur pada Model 4 memberikan hasil paling meyakinkan dalam mengklasifikasikan citra batuan pada data pengujian.

4.5. Perbandingan Evaluasi Model *Stacking* dan *Non Stacking*

4.5.1. Perbandingan evaluasi model dengan *cross validation accuracy*, *test accuracy*, *precision*, *recall*, *f1-score* dan ukuran model

Kinerja seluruh model dievaluasi dengan membandingkan empat model *stacking* dan enam model *non-stacking* melalui beberapa indikator penting, yaitu akurasi *cross-validation*, akurasi pada data uji, ukuran file model, nilai *precision*, *recall*, dan *F1-score*. Kumpulan metrik ini memberikan gambaran yang lebih lengkap mengenai bagaimana model belajar dari data latih, seberapa baik model mampu mengenali pola baru pada data uji, dan sejauh mana kemampuan model dalam melakukan klasifikasi tetap merata pada tiap kelas dalam konteks klasifikasi multikelas. Selain itu, informasi mengenai ukuran file disertakan agar dapat menilai kebutuhan memori ketika model diterapkan pada perangkat dengan sumber daya terbatas.

Perbandingan antara model *stacking* dan model tunggal menjadi penting untuk mengetahui apakah pendekatan *ensemble* benar-benar memberikan peningkatan kualitas prediksi. Melalui analisis terhadap seluruh metrik, dapat terlihat model mana yang paling stabil, paling akurat, sekaligus paling efisien dari sisi ukuran. Hasil lengkap penilaian terhadap masing-masing model dapat dilihat pada Tabel 4.2.

Tabel 4.2. Perbandingan Hasil Evaluasi Model *Stacking* dan *Non Stacking*

| Model | Cross Validation Accuracy | Test Accuracy | Precision | Recall | F1- Score | Ukuran File (mb) |
|----------------|--|--------------------------|------------------|---------------|----------------------|-----------------------------|
| Model 1 | 0,80 | 0,83 | 0,84 | 0,82 | 0,83 | 22,18 |
| Model 2 | 0,84 | 0,81 | 0,82 | 0,80 | 0,81 | 13,08 |
| Model 3 | 0,80 | 0,83 | 0,84 | 0,83 | 0,83 | 35,92 |
| Model 4 | 0,81 | 0,84 | 0,85 | 0,84 | 0,84 | 28,89 |
| ResNet50 | 0,80 | 0,80 | 0,80 | 0,80 | 0,80 | 205,03 |
| InceptionV3 | 0,76 | 0,80 | 0,80 | 0,80 | 0,80 | 169,16 |
| VGG16 | 0,70 | 0,77 | 0,78 | 0,77 | 0,78 | 110,26 |
| DenseNet121 | 0,75 | 0,78 | 0,78 | 0,78 | 0,78 | 40,59 |
| MobileNetV2 | 0,72 | 0,80 | 0,81 | 0,80 | 0,80 | 23,49 |
| EfficientNetB0 | 0,76 | 0,80 | 0,81 | 0,80 | 0,80 | 41,46 |

Sumber: Hasil Pengolahan Data, 2025

Berdasarkan Tabel 4.2, terlihat adanya perbedaan yang cukup jelas antara hasil evaluasi model *stacking* dan *non-stacking*, baik dari sisi akurasi, metrik makro (*precision*, *recall*, *F1-score*), maupun ukuran file model. Secara umum, model *stacking*

memberikan kinerja yang relatif lebih konsisten dan stabil dibandingkan model *non-stacking*. Hal ini terlihat dari nilai *precision*, *recall*, dan *F1-score*-nya bahwa model *stacking* yang berada pada rentang 0,81-0,85, yang lebih tinggi dibandingkan sebagian besar model tunggal. Berdasarkan pada temuan tersebut menjadi indikasi positif, bahwa teknik penggabungan fitur dari dua arsitektur CNN mampu menghasilkan representasi visual yang lebih kaya sehingga meningkatkan kualitas prediksi jenis batuan.

Namun, beberapa temuan menarik lain juga muncul dari perbandingan hasil akurasi model. Model 2 (VGG16 + DenseNet121) menunjukkan anomali, karena memiliki nilai *cross-validation* tertinggi (0,84) namun justru mengalami penurunan pada akurasi uji (0,81). Pola ini berbeda dari tiga model *stacking* lainnya yang justru cenderung stabil atau meningkat pada data uji. Anomali ini mengisyaratkan bahwa kombinasi VGG16 dan DenseNet121 berpotensi mengalami *overfitting*, sehingga meskipun tampak unggul pada proses pelatihan, model tersebut tidak mampu mempertahankan akurasi yang sama ketika diuji pada data baru. Dampaknya adalah model tersebut menjadi kurang dapat diandalkan untuk generalisasi.

Sebaliknya, Model 4 (Resnet50 + MobileNetV2) memberikan hasil paling positif, dengan akurasi uji tertinggi (0,84) dan metrik *F1-Score* sebesar 0,84, menjadikannya model *stacking* paling efektif. Hal ini menunjukkan sinergi yang baik antara ResNet50 dan MobileNetV2, di mana kombinasi fitur dalam tingkatan kedalaman berbeda mampu memperkuat kemampuan klasifikasi.

Berbeda dengan model *stacking*, beberapa model *non stacking*, seperti VGG16 dan DenseNet121, memperoleh skor akurasi yang lebih rendah, terutama pada data uji (0,77-0,78). Selain itu, model CNN besar seperti ResNet50 dan InceptionV3 memiliki

ukuran file yang sangat besar (169-205 MB), sehingga meskipun akurasi mendekati 0,80, model-model ini kurang efisien untuk implementasi di perangkat dengan keterbatasan memori. Sebaliknya, model *stacking* memiliki ukuran file jauh lebih kecil (13-35 MB), sehingga lebih praktis tanpa mengorbankan akurasi.

Adanya perbedaan hasil tersebut, memperkuat kesimpulan bahwa model *stacking* tidak hanya unggul dalam metrik evaluasi, tetapi juga lebih efisien secara komputasi. Pada sisi lain, anomali pada Model 2 (VGG16 + DenseNet121) mengingatkan bahwa kombinasi arsitektur CNN tidak selalu menghasilkan peningkatan akurasi, dan diperlukan pemilihan *base-learner* yang tepat agar *stacking* benar-benar optimal. Secara visual mengenai perbandingan hasil evaluasi model *stacking* dan *non stacking* dapat dilihat pada Gambar 4.5.



Gambar 4.5. Perbandingan Hasil Evaluasi Model *Stacking* dan *Non Stacking*

Berdasarkan pada Gambar 4.5, terlihat bahwa model yang dibangun menggunakan teknik *stacking* secara konsisten menunjukkan kualitas hasil prediksi yang lebih unggul pada seluruh metrik penilaian. Hal ini tampak dari nilai *Cross Validation Accuracy*, *Test Accuracy*, *Precision*, *Recall*, hingga *F1-Score* yang rata-rata lebih tinggi pada Model 1 (MobileNetV2 + EfficientNetB0), Model 2 (VGG16 + DenseNet121), Model 3 (ResNet50, InceptionV3), dan Model 4 (Resnet50, MobileNetV2) dibandingkan dengan model CNN tunggal seperti ResNet50, InceptionV3, VGG16, serta DenseNet121.

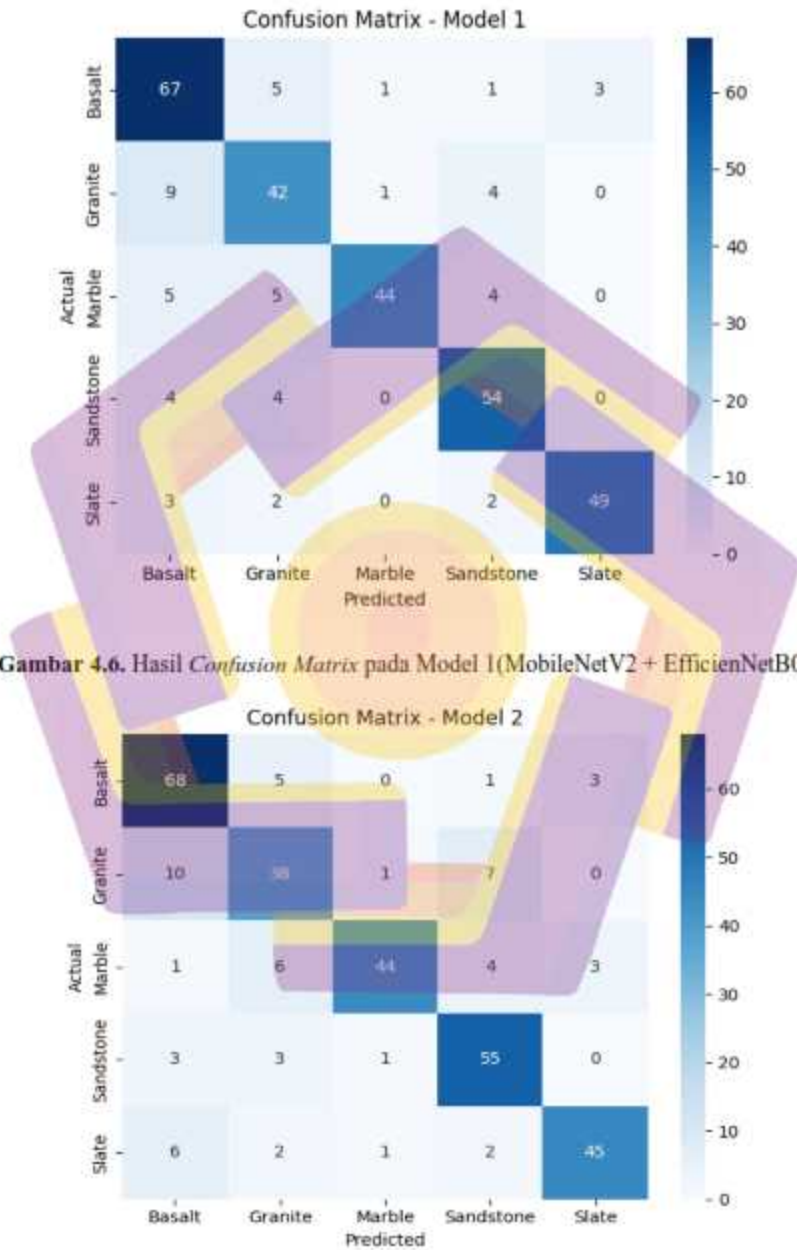
Kecenderungan tersebut menunjukkan bahwa pendekatan *stacking* mampu menggabungkan kekuatan berbagai arsitektur CNN dalam mengekstraksi informasi penting dari citra batuan. Tekstur batuan yang memiliki pola dan struktur kompleks umumnya sulit dipahami dengan baik oleh satu model saja. Melalui *stacking*, setiap model berperan saling melengkapi dalam mengenali karakteristik visual batuan, sehingga menghasilkan prediksi yang lebih stabil dan akurat. Kondisi ini juga tampak dari nilai *Precision*, *Recall*, dan *F1-Score* yang seimbang, menandakan bahwa kemampuan model dalam membedakan tiap kelas batuan cukup baik dan tidak hanya unggul pada satu kelas tertentu.

Secara khusus, Model 4 (Resnet50, MobileNetV2) menjadi model dengan hasil klasifikasi paling unggul, terutama pada data pengujian. Artinya, model tersebut mampu beradaptasi secara baik terhadap data baru yang belum pernah dilihat sebelumnya. Sebaliknya, beberapa model *non-stacking* seperti VGG16 dan DenseNet121 menunjukkan nilai yang cenderung lebih rendah, menandakan adanya keterbatasan dalam menyajikan representasi fitur secara menyeluruh.

Hasil analisis tersebut menunjukkan bahwa dalam konteks klasifikasi citra batuan, penggunaan teknik *ensemble* seperti *stacking* tidak hanya meningkatkan nilai evaluasi secara kuantitatif, tetapi juga memberikan keuntungan praktis.

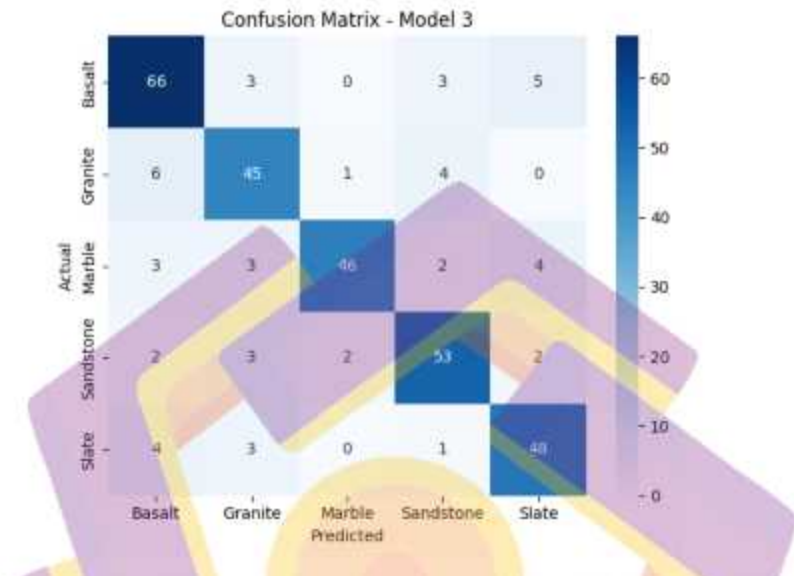
4.5.2. Perbandingan evaluasi model dengan *confusion matrix*

Confusion matrix digunakan dalam penelitian ini karena mampu memberikan informasi detail mengenai kemampuan model dalam mengklasifikasikan setiap kelas batuan secara lebih spesifik. Melalui perbandingan *confusion matrix* antar model, dapat diketahui kelas mana yang telah dikenali dengan tepat serta kelas mana yang masih sering mengalami kekeliruan. Analisis ini juga dijadikan dasar untuk melihat sejauh mana pendekatan *stacking* dapat mengatasi keterbatasan yang masih muncul pada model individual dalam proses prediksi. Adapun secara lengkap hasil *confusion matrix* model *stacking* dapat dilihat pada Gambar 4.6, Gambar 4.7, Gambar 4.8, Gambar 4.9, dan Gambar 4.10, sementara *non stacking* Gambar 4.11, Gambar 4.12, Gambar 4.13, Gambar 4.14, Gambar 4.15, dan Gambar 4.16.

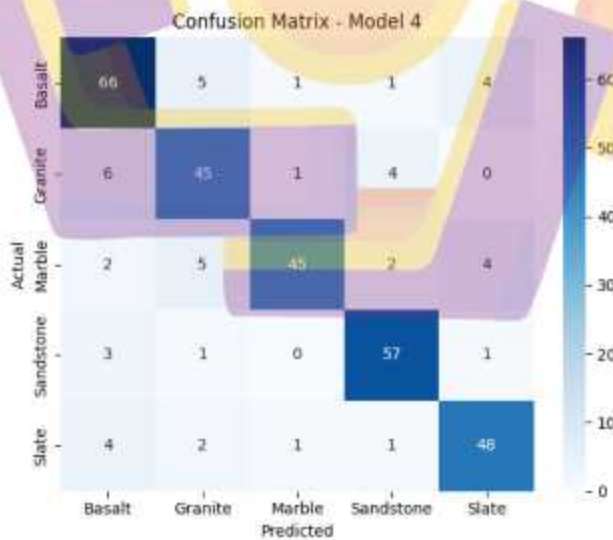


Gambar 4.6. Hasil *Confusion Matrix* pada Model 1 (MobileNetV2 + EfficientNetB0)

Gambar 4.7. Hasil *Confusion Matrix* pada Model 2 (VGG16 + DenseNet121)



Gambar 4.8. Hasil *Confusion Matrix* pada Model 3 (ResNet50, InceptionV3)



Gambar 4.9. Hasil *Confusion Matrix* pada Model 4 (Resnet50, MobileNetV2)

Berdasarkan hasil *confusion matrix* pada Gambar 4.6, Gambar 4.7, Gambar 4.8, dan Gambar 4.9, dapat diketahui bahwa seluruh model *stacking* menunjukkan kemampuan klasifikasi yang cukup baik pada setiap kelas batuan. Meskipun demikian, terdapat variasi ketepatan prediksi antar model dan antar kelas yang perlu dianalisis lebih lanjut secara kuantitatif.

Pada Model 1 (MobileNetV2 + EfficientNetB0), kelas Basalt memperoleh prediksi benar tertinggi yaitu sebanyak 67 sampel, dengan kesalahan tidak lebih dari 5 data. Hasil klasifikasi yang cukup baik juga terlihat pada kelas Sandstone dengan 54 prediksi benar, sedangkan kelas Slate berhasil dikenali sebesar 49 data, meskipun masih ditemukan kesalahan prediksi ke kelas Sandstone sebanyak 2 kasus.

Model 2 (VGG16 + DenseNet121) menunjukkan peningkatan pada kelas Basalt yang mencapai 68 prediksi benar, dan jumlah kesalahan yang lebih rendah dibandingkan Model 1 (MobileNetV2 + EfficientNetB0). Sandstone juga tetap stabil dengan 55 sampel tepat, hanya 3 data yang salah dipetakan ke kelas lain. Namun demikian, kelas Granite mengalami penurunan hasil klasifikasi karena hanya 38 sampel berhasil terklasifikasi benar, dengan sejumlah data terbagi ke Basalt dan Marble.

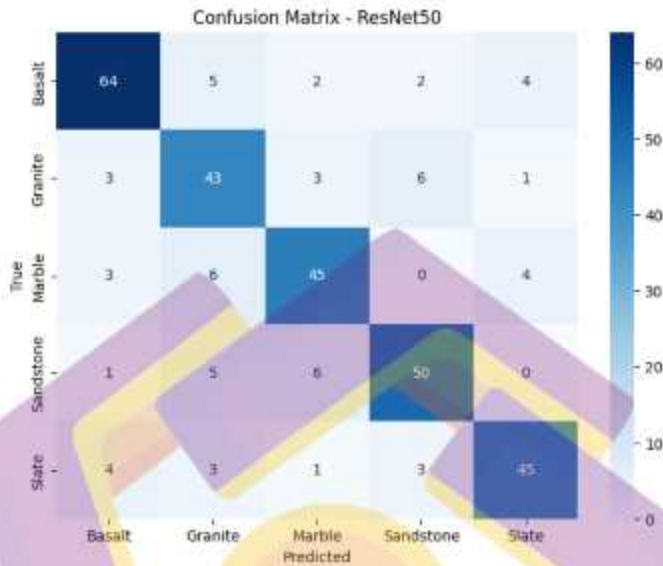
Model 3 (ResNet50, InceptionV3) memperlihatkan hasil klasifikasi yang cenderung mendekati Model 1 (MobileNetV2 + EfficientNetB0). Basalt berada pada angka 66 prediksi benar, Granite 43, Marble 46, Sandstone 53, dan Slate 48. *Misclass* masih ditemui terutama pada Granite dan Marble dengan tingkat kesalahan sekitar 4-5 data pada kelas dengan karakteristik visual yang serupa.

Sementara itu, Model 4 (Resnet50, MobileNetV2) menjadi model *stacking* dengan hasil paling stabil dan akurat. Seluruh kelas batuan memperoleh prediksi benar

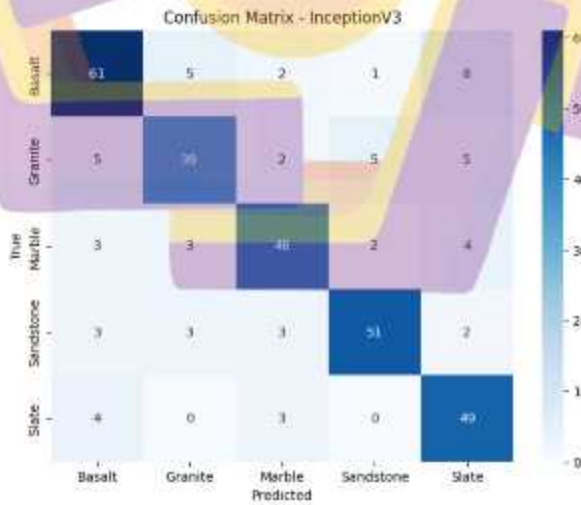
tertinggi dibandingkan ketiga model sebelumnya: Basalt (66), Granite (45), Marble (49), Sandstone (57), dan Slate (48). Jumlah kesalahan pada tiap kelas hanya berkisar antara 1-3 sampel, sehingga mengindikasikan kemampuan generalisasi yang unggul.

Secara lebih luas, pola hasil *confusion matrix* menunjukkan bahwa kelas Basalt dan Sandstone merupakan kelas yang paling mudah dikenali oleh semua model, karena memiliki karakteristik visual yang cukup tegas. Sebaliknya, kelas Granite dan Marble masih sering tertukar, mencapai hingga 10 kesalahan klasifikasi pada beberapa model, yang wajar mengingat keduanya memiliki tekstur yang sangat mirip secara geologis.

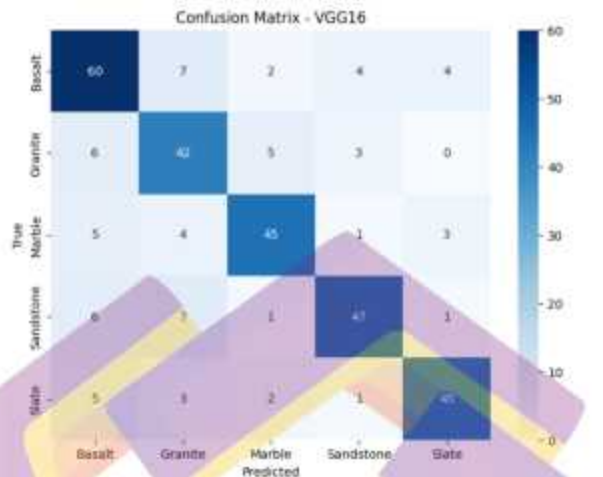
Berdasarkan pada pertimbangan hasil akurasi setiap kelas dan konsistensinya, maka dapat disimpulkan bahwa Model 4 (Resnet50, MobileNetV2) merupakan model *stacking* terbaik dalam penelitian ini. Penggabungan beragam model CNN pada teknik *stacking* terbukti memperkaya ekstraksi fitur, sehingga mampu mengurangi error yang sebelumnya muncul pada model individual. Oleh karena itu, pendekatan *stacking* tidak hanya meningkatkan akurasi keseluruhan, tetapi juga menjadikan sistem klasifikasi lebih tangguh dalam menangani variasi citra batuan.



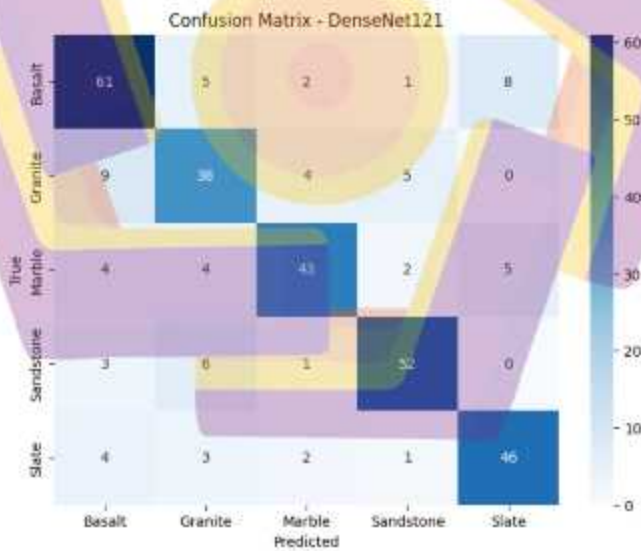
Gambar 4.10. Hasil *Confusion Matrix* pada ResNet50



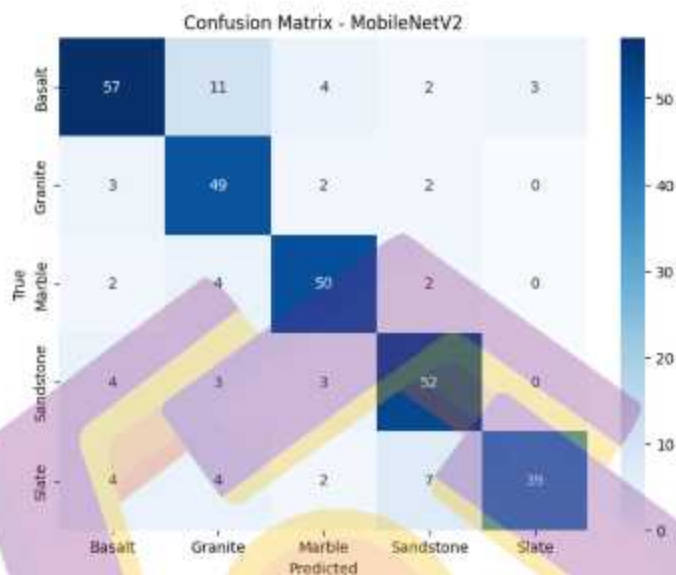
Gambar 4.11. Hasil *Confusion Matrix* pada InceptionV3



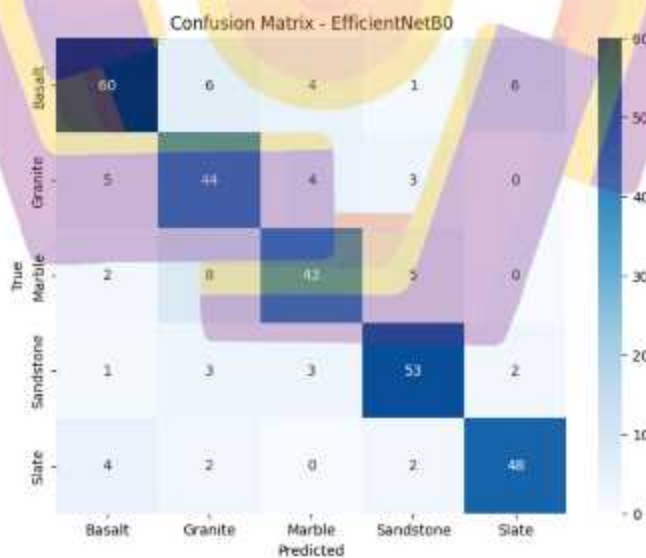
Gambar 4.12. Hasil *Confusion Matrix* pada VGG16



Gambar 4.13. Hasil *Confusion Matrix* pada DenseNet121



Gambar 4.14. Hasil *Confusion Matrix* pada MobileNetV2



Gambar 4.15. Hasil *Confusion Matrix* pada EfficientNetB0

Berdasarkan hasil *confusion matrix* pada Gambar 4.10, Gambar 4.11, Gambar 4.12, Gambar 4.13, Gambar 4.14, dan Gambar 4.15, dapat diketahui bahwa seluruh model CNN *non-stacking* telah mampu melakukan klasifikasi citra batuan dengan cukup baik, namun masih terdapat perbedaan tingkat akurasi antar kelas dan antar model.

Pada model ResNet50, kelas Basalt mencapai 64 prediksi benar, namun masih terdapat 8 kesalahan klasifikasi pada Granite dan Marble. Sandstone menunjukkan kinerja yang stabil dengan 50 prediksi tepat, sementara kesalahan pada Slate masih terjadi sebanyak 5 kasus, yang umumnya salah dialihkan ke kelas Basalt.

Model InceptionV3 memperlihatkan pola serupa. Basalt berhasil diklasifikasikan 61 kali dengan benar, sedangkan Granite masih menampilkan kerentanan dengan hanya 39 prediksi benar dan sebagian besar salah dipetakan ke Marble. Slate memperlihatkan hasil yang cukup baik dengan 49 prediksi benar, meskipun masih ada data yang tidak tepat sasaran.

Model VGG16 menunjukkan peningkatan pada kelas Granite yang memperoleh 42 prediksi benar, tetapi kelas Basalt sedikit menurun dengan 60 sampel yang tepat. Kelas Sandstone memiliki 47 prediksi benar, namun masih terdapat jumlah kesalahan yang cukup tinggi pada Slate dan Basalt sebagai kelas alternatif.

Pada DenseNet121, Granite kembali menjadi kelas yang paling sering salah diprediksi dengan hanya 38 data yang benar. Sementara itu, Basalt dan Sandstone masing-masing memperoleh 61 dan 52 prediksi benar, menunjukkan bahwa kedua kelas ini lebih mudah dikenali dibandingkan kelas lain.

Model MobileNetV2 memperlihatkan penurunan pada Basalt dengan 57 prediksi tepat dan kesalahan terbesar mencapai 11 kasus pada Granite yang salah dialihkan ke Basalt. Sandstone tetap berada pada tingkat yang stabil (52 prediksi benar), namun Slate mengalami penurunan menjadi 39 prediksi benar.

Terakhir, model EfficientNetB0 menunjukkan kinerja yang lebih merata di seluruh kelas. Sandstone memperoleh 53 prediksi benar, sementara Slate mencatat 48 prediksi tepat. Akan tetapi, Granite masih menjadi tantangan dengan 44 prediksi benar dan *misclassification* yang sebagian besar mengarah ke Basalt dan Marble.

Jika ditinjau dari pola keseluruhan, terdapat beberapa kesimpulan yang dapat dicatat: (a) Basalt dan Sandstone adalah kelas yang paling mudah dikenali oleh seluruh model *non-stacking*, (b) kedua kelas ini konsisten memperoleh nilai prediksi benar tertinggi di setiap model, (c) Granite dan Marble cenderung saling tertukar, menunjukkan kesamaan tekstur yang membuat model sulit membedakan keduanya. Kesalahan dapat mencapai 8-12 sampel pada beberapa model, dan (d) Slate menunjukkan kinerja yang fluktuatif, dengan prediksi benar berkisar antara 39-49 data, serta sering terklasifikasi sebagai Sandstone.

Secara keseluruhan, dapat disimpulkan bahwa meskipun model *non-stacking* memiliki kemampuan dasar yang baik dalam pengenalan pola batuan, namun masih terdapat kelemahan pada kelas-kelas dengan karakteristik visual yang berdekatan. Hasil ini menjadi dasar perbandingan yang kuat, bahwa teknik *stacking* diperlukan untuk meningkatkan ketelitian klasifikasi, terutama pada kelas yang secara geologis memiliki kemiripan struktur seperti Granite dan Marble.

4.6. Pembahasan

4.6.1. Perbandingan kinerja model tunggal dan model *stacking*

Hasil penelitian ini menunjukkan bahwa pendekatan *ensemble stacking* mampu menghasilkan kinerja yang lebih unggul dibandingkan dengan penggunaan model tunggal. Model *stacking* terbaik pada penelitian ini, yaitu Model 4 (Resnet50, MobileNetV2), menghasilkan nilai *test accuracy* sebesar 0,84 dengan *precision* 0,85, *recall* 0,84, dan *F1-score* 0,84. Hasil ini lebih tinggi dibandingkan dengan model CNN tunggal seperti ResNet50, InceptionV3, MobileNetV2, dan EfficientNetB0 yang masing-masing hanya mencapai akurasi sekitar 0,80. Temuan ini sejalan dengan penelitian sebelumnya yang menunjukkan bahwa pemanfaatan model *deep learning* berbasis *transfer learning* mampu memberikan hasil klasifikasi yang baik, seperti penelitian oleh Chen et al. (2023), yang memperoleh akurasi 82% menggunakan ResNet34. Namun, penelitian tersebut masih menggunakan dataset yang relatif kecil, sedangkan penelitian ini menggunakan dataset yang lebih besar dengan 1.632 citra dan lima kelas batuan, sehingga kemampuan generalisasi model menjadi lebih baik.

Apabila dibandingkan dengan penelitian oleh dos Anjos et al. (2021), yang memperoleh akurasi sebesar 75,54% pada evaluasi per gambar dan 81,33% pada evaluasi per plug menggunakan CNN pada citra micro-CT, hasil penelitian ini menunjukkan hasil yang lebih optimal. Perbedaan ini dipengaruhi oleh penggunaan pendekatan *ensemble stacking* yang mampu menggabungkan keunggulan beberapa model secara simultan, sedangkan penelitian tersebut masih mengandalkan satu model CNN tunggal. Selain itu, penelitian Dawson et al (2023) juga melaporkan bahwa arsitektur VGG19 menghasilkan akurasi terbaik sebesar 81,3%, yang kembali

menunjukkan bahwa hasil penelitian ini dengan akurasi 0,84 berada di atas model CNN tunggal pada penelitian terdahulu.

Penelitian oleh Chawshin et al. (2021) yang menggunakan arsitektur VGG16 juga menunjukkan bahwa CNN mampu mengklasifikasikan mineral dengan baik, namun masih menghadapi keterbatasan pada ukuran dataset yang kecil dan risiko *overfitting*. Hal yang sama juga terlihat pada penelitian Bressan et al. (2020), di mana *Random Forest* mampu mencapai akurasi di atas 80%, tetapi stabilitas model sangat dipengaruhi oleh kualitas data antar wilayah. Kondisi ini sejalan dengan temuan penelitian ini, bahwa meskipun model tunggal mampu mencapai akurasi yang cukup baik, namun kinerja masih belum melampaui model *stacking* yang mampu mengurangi ketergantungan terhadap satu arsitektur tertentu.

Sementara itu, penelitian oleh Bera (2022) dengan pendekatan Swin-Transformer hanya memperoleh akurasi sebesar 73,10% akibat keterbatasan ukuran dataset. Hal ini memperkuat temuan pada penelitian ini bahwa kualitas dan jumlah data sangat berpengaruh terhadap hasil akurasi, presisi, *recall*, dan *f1-score* model, selain juga dipengaruhi oleh strategi pemodelan yang digunakan. Melalui kombinasi beberapa arsitektur CNN melalui pendekatan *stacking* serta menggunakan dataset yang lebih besar dan validasi silang, penelitian ini berhasil mencapai hasil akurasi, presisi, *recall*, dan *f1-score* yang lebih stabil dan lebih tinggi dibandingkan sebagian besar penelitian terdahulu.

Secara keseluruhan, hasil penelitian ini tidak hanya mendukung temuan-temuan sebelumnya yang menyatakan bahwa CNN efektif untuk klasifikasi citra batuan, tetapi juga menunjukkan bahwa pendekatan *ensemble stacking* mampu memberikan

peningkatan kinerja yang lebih optimal dibandingkan penggunaan satu model tunggal. Akhirnya, penelitian ini memberikan kontribusi tambahan dalam pengembangan metode klasifikasi citra batuan berbasis *deep learning* yang lebih akurat dan *robust*.

4.6.2. Kontribusi model *non stacking* terhadap akurasi model *stacking*

Penelitian ini mengimplementasikan pendekatan *stacking* dengan mengombinasikan beberapa arsitektur *Convolutional Neural Network (CNN) pre-trained* guna menganalisis kontribusinya terhadap peningkatan akurasi klasifikasi citra batuan. Berdasarkan visualisasi hasil evaluasi pada metrik akurasi, *precision*, *recall*, dan *f1-score* yang ditampilkan pada Gambar 4.5, terlihat bahwa beberapa model tunggal (*non-stacking*) seperti ResNet50, InceptionV3, dan EfficientNetB0 telah menunjukkan kinerja yang cukup baik secara individual.

Hasil pengujian terhadap model tunggal menunjukkan bahwa ResNet50 menghasilkan akurasi pengujian tertinggi sebesar 0,80, diikuti oleh DenseNet121, MobileNetV2, dan VGG16 dengan rentang akurasi antara 0,76 hingga 0,77. Rentang nilai ini menunjukkan bahwa meskipun model tunggal mampu melakukan klasifikasi dengan cukup baik, masih terdapat keterbatasan dalam mengenali kompleksitas variasi tekstur dan pola citra batuan.

Setelah diterapkan pendekatan *stacking*, terjadi peningkatan akurasi yang konsisten pada seluruh konfigurasi model gabungan. Kombinasi ResNet50 dan MobileNetV2 dengan *meta-learner Support Vector Machine (SVM)* menghasilkan akurasi tertinggi sebesar 0,84. Sementara itu, kombinasi MobileNetV2 dan EfficientNetB0 serta ResNet50 dan InceptionV3 masing-masing menghasilkan akurasi sebesar 0,83. Selain itu, kombinasi VGG16 dan DenseNet121 juga menunjukkan

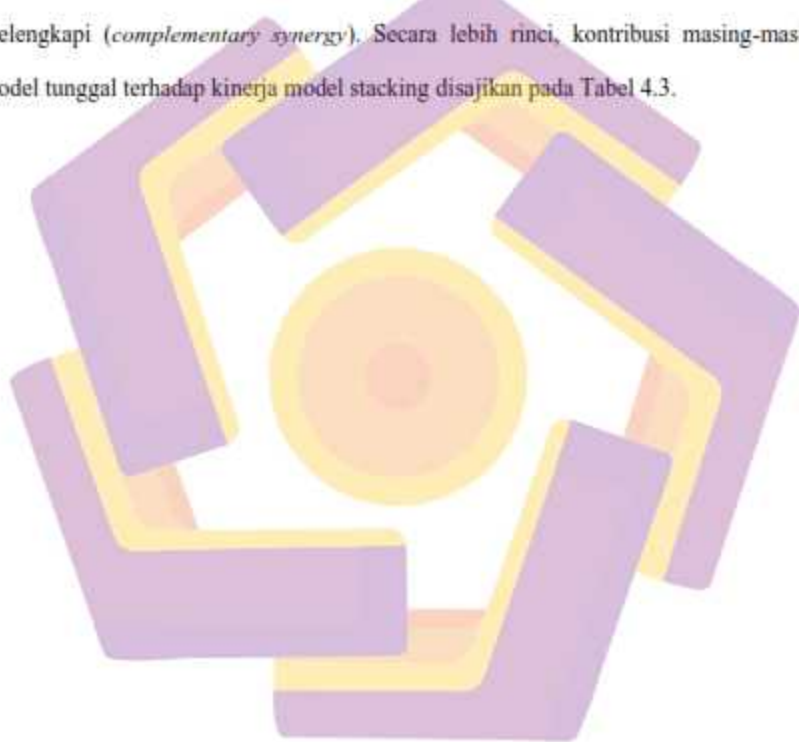
peningkatan hingga 0,81, yang lebih tinggi dibandingkan dengan hasil akurasi masing-masing model tunggal. Hasil ini menegaskan bahwa pendekatan *stacking* mampu meningkatkan kinerja klasifikasi melalui penguatan keunggulan antar model sekaligus mengompensasi kelemahan masing-masing arsitektur.

Keunggulan utama pendekatan *stacking* terletak pada kemampuannya dalam mengintegrasikan representasi fitur dari berbagai model CNN dengan karakteristik yang berbeda. Model dengan kompleksitas tinggi seperti ResNet50 unggul dalam mengekstraksi pola global dan struktur berskala besar, sedangkan model ringan seperti MobileNetV2 lebih efisien dalam mengenali fitur-fitur lokal yang halus. Ketika kedua tipe model tersebut digabungkan dalam satu sistem *stacking*, diperoleh representasi fitur yang lebih komprehensif sehingga meningkatkan ketepatan dan stabilitas hasil klasifikasi.

Secara keseluruhan, hasil penelitian menunjukkan bahwa pendekatan *stacking* merupakan strategi yang efektif dan adaptif dalam meningkatkan akurasi hasil klasifikasi citra batuan. Penggabungan beberapa model CNN melalui mekanisme meta-learning tidak hanya meningkatkan akurasi, tetapi juga memperbaiki kemampuan generalisasi model serta mengurangi risiko *overfitting*. Setelah terbukti bahwa model *stacking* memberikan hasil klasifikasi yang lebih unggul dibandingkan model tunggal, analisis selanjutnya difokuskan pada kontribusi masing-masing model individual sebagai *base learner* terhadap kinerja akhir sistem *stacking*.

Setiap arsitektur CNN memiliki karakteristik dan kekuatan yang berbeda dalam mengekstraksi fitur citra, baik dari sisi kedalaman jaringan, jumlah parameter, maupun kompleksitas pola visual. Arsitektur seperti ResNet50 dan InceptionV3 unggul dalam

menangkap pola global dan struktur berskala besar, sedangkan MobileNetV2 dan EfficientNetB0 lebih efisien dalam mendeteksi detail lokal. Melalui penggabungan model-model tersebut, sistem *stacking* memperoleh keseimbangan antara kemampuan representasi mendalam dan efisiensi komputasi. Kelebihan satu arsitektur mampu menutupi keterbatasan arsitektur lainnya sehingga tercipta sinergi yang saling melengkapi (*complementary synergy*). Secara lebih rinci, kontribusi masing-masing model tunggal terhadap kinerja model *stacking* disajikan pada Tabel 4.3.



Tabel 4.3. Kontribusi Model Tunggal terhadap Akurasi Model Stacking

| Model CNN | Akurasi Model Individual (<i>Non Stacking</i>) | Peran dalam <i>Stacking</i> | Akurasi Model <i>Stacking</i> | Estimasi Kontribusi (%) | Dasar Penilaian |
|----------------|---|--|----------------------------------|----------------------------|--|
| ResNet50 | 0,80 | <i>Stacking</i> model 3 dan model 4 | 0,83-0,84 | 30-32 | Memberi peningkatan 3-4% dan menjadi komponen dominan |
| InceptionV3 | 0,78 | <i>Stacking</i> model 3 | 0,83 | 20-22 | Menambah 3-5% kinerja ketika dikombinasikan dengan ResNet50 |
| VGG16 | 0,76 | <i>Stacking</i> model 2 | 0,81 | 10-12 | Meningkatkan 5% dibanding akurasi individunya |
| DenseNet121 | 0,77 | <i>Stacking</i> model 2 | 0,81 | 12-14 | Berkontribusi stabilitas dan konektivitas antar fitur |
| MobileNetV2 | 0,77 | <i>Stacking</i> model 1 dan model 4 | 0,83-0,84 | 14-16 | Kombinasi dengan ResNet50 menghasilkan peningkatan signifikan |
| EfficientNetB0 | 0,79 | <i>Stacking</i> model 1 | 0,83 | 10-12 | Menambah generalisasi dan efisiensi model ringan |

Berdasarkan pada Tabel 4.3, dapat diketahui bahwa setiap arsitektur CNN memberikan kontribusi berbeda terhadap peningkatan akurasi *stacking*, bergantung pada kompleksitas lapisan dan kapasitas ekstraksi fiturnya. Model ResNet50 menjadi kontributor paling dominan dengan estimasi kontribusi 30-32%, berkat kemampuannya mengenali pola global dan fitur mendalam. Peningkatan akurasi sebesar 3-4% dibanding model tunggal tertinggi memperkuat posisinya sebagai komponen utama dalam konfigurasi *stacking* model 3 (ResNet50 dan InceptionV3) maupun model 4 (ResNet50 dan MobileNetV2).

Model InceptionV3 memberikan kontribusi cukup besar (20-22%) melalui mekanisme *multi-filter* yang menangkap fitur dengan berbagai ukuran dan kedalaman. Kombinasinya dengan ResNet50 menghasilkan akurasi hingga 0,83, menunjukkan hubungan komplementer yang kuat antarrepresentasi fitur.

Model VGG16 dan DenseNet121, meskipun memiliki akurasi dasar lebih rendah (0,76-0,77), tetap berperan penting dalam menjaga kestabilan sistem dengan kontribusi 10-14%. DenseNet121 secara khusus membantu memperdalam propagasi informasi antar lapisan, meningkatkan konektivitas fitur, dan memperkaya konteks representasi.

Model MobileNetV2 juga memiliki peran signifikan (14-16%) terutama pada konfigurasi *stacking* model 1 (MobileNetV2 + EfficientNetB0) dan model 4 (ResNet50 dan MobileNetV2). Kombinasinya dengan ResNet50 memberikan hasil terbaik (0,84) menunjukkan bahwa model ringan dapat memberikan dampak besar bila dikombinasikan dengan model yang lebih dalam.

Adapun EfficientNetB0 memberikan tambahan kontribusi 10-12%, dengan keunggulan dalam efisiensi parameter dan kemampuan generalisasi. Kombinasinya

dalam konfigurasi *stacking* model 1 (MobileNetV2 + EfficientNetB0) ringan mampu mencapai akurasi tinggi (0,83) dengan kebutuhan komputasi minimal, menjadikannya ideal untuk sistem dengan sumber daya terbatas.

Secara keseluruhan, hasil ini memperlihatkan bahwa setiap arsitektur CNN memiliki peran strategis yang saling melengkapi dalam sistem *stacking*. Model berat seperti ResNet50 dan InceptionV3 berfungsi sebagai penangkap pola global, sementara model ringan seperti MobileNetV2 dan EfficientNetB0 memperkuat efisiensi dan generalisasi. Kolaborasi keduanya menciptakan sinergi optimal yang tercermin pada peningkatan akurasi, stabilitas, dan efisiensi sistem klasifikasi citra batuan secara keseluruhan.

4.6.2. Analisis hasil model berdasarkan kemampuan klasifikasi per kelas batuan

Hasil unggul dari model seperti ResNet50 dapat dikaitkan dengan arsitekturnya yang dalam dan stabil, yang memanfaatkan *residual connections* untuk mengatasi kendala *vanishing gradient*. Fitur ini memungkinkan informasi penting tetap mengalir secara efektif di setiap lapisan, sehingga model mampu menangkap pola-pola kompleks yang terdapat dalam citra batuan. Hal ini sangat krusial untuk membedakan batuan yang memiliki kemiripan visual tinggi, seperti *granite* dan *sandstone*, yang membutuhkan representasi fitur yang mendetail dan presisi tinggi.

Sementara itu, InceptionV3 menunjukkan hasil yang kompetitif berkat pendekatannya yang mengintegrasikan berbagai ukuran filter dalam satu modul. Strategi ini membuat model mampu mengenali fitur baik pada skala mikro maupun makro, menjadikannya fleksibel dalam menangkap variasi visual antar jenis batuan. Pendekatan

multiskala ini memungkinkan model menyesuaikan diri terhadap karakteristik batuan yang bervariasi dari segi tekstur dan bentuk.

Keunggulan pendekatan *stacking* khususnya pada konfigurasi seperti *Stacking* model 3 (ResNet50 + InceptionV3) dan *Stacking* model 4 (ResNet50 + MobileNetV2) terletak pada kemampuannya menggabungkan fitur dari model dengan karakteristik berbeda. ResNet50 menyumbang kedalaman analisis dan kestabilan prediksi, sementara MobileNetV2 atau InceptionV3 memberikan fleksibilitas tinggi dalam mendeteksi pola lokal maupun global. Kombinasi tersebut menciptakan sistem yang lebih tangguh terhadap keragaman data dan lebih presisi dalam menghasilkan prediksi.

Model ringan seperti MobileNetV2 dan EfficientNetB0, meskipun unggul dalam efisiensi komputasi, cenderung memiliki keterbatasan dalam mengenali fitur kompleks saat digunakan secara mandiri. Hal ini karena keduanya dirancang untuk efisiensi dan kecepatan, bukan eksplorasi fitur yang mendalam. Namun, ketika dikombinasikan dalam sistem *stacking*, kelemahan ini dapat dikompensasi oleh model yang lebih kuat, menghasilkan sistem klasifikasi yang lebih seimbang dan akurat.

Berdasarkan hasil *confusion matrix* dari pendekatan *stacking* menggunakan SVM sebagai *meta-learner*, terlihat variasi kinerja model dalam mengenali masing-masing kelas batuan. *Basalt* merupakan kelas yang paling mudah dikenali oleh seluruh model, dengan prediksi benar berkisar antara 72-76 sampel, menandakan tekstur dan pola visualnya cukup khas. Kelas *Granite* juga menunjukkan konsistensi tinggi, dengan jumlah klasifikasi benar antara 52-54 sampel, mengindikasikan bahwa karakteristik visualnya cukup menonjol dan tidak banyak mengalami tumpang tindih dengan kelas lain.

Untuk kelas *Marble*, sebagian besar model mampu mengidentifikasi dengan akurasi baik (sekitar 52-53 data benar), meskipun terdapat beberapa kasus *misclass* ke kelas *Sandstone*, yang kemungkinan besar disebabkan oleh kemiripan tekstur. Kesalahan serupa juga terjadi pada *Sandstone*, terutama pada model 1 (MobileNetV2 dan EfficientNetB0), yang cenderung salah mengklasifikasikan ke *Basalt*. Sebaliknya, pendekatan *stacking*, khususnya pada konfigurasi model 3 (ResNet50 + InceptionV3) dan model 4 (ResNet50 + MobileNetV2), menunjukkan perbaikan signifikan dalam mengurangi jenis kesalahan ini. Kelas *Slate* menjadi salah satu yang paling konsisten dikenali oleh semua model, dengan tingkat klasifikasi benar yang tinggi (sekitar 53-54 data), menunjukkan bahwa ciri visualnya sangat khas dan mudah dipisahkan dari kelas lain.

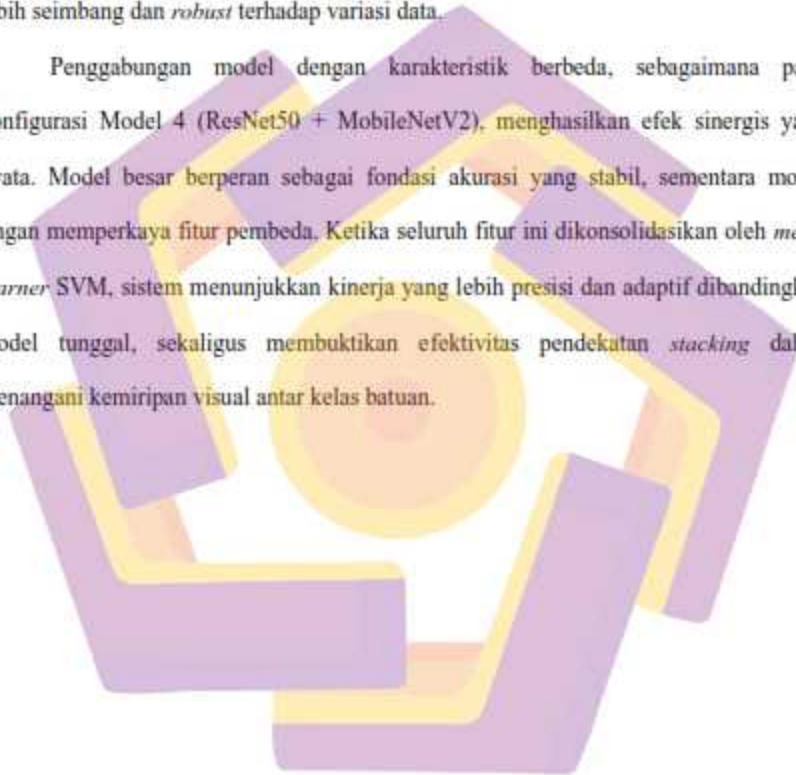
Secara keseluruhan, analisis ini menggarisbawahi bahwa karakteristik arsitektur model sangat berpengaruh terhadap hasil klasifikasi per kelas. Model *stacking* yang menggabungkan kelebihan dari berbagai CNN mampu meningkatkan ketahanan dan akurasi sistem terhadap variasi fitur antar kelas batuan. Sementara itu, model 1 (MobileNetV2 dan EfficientNetB0) tetap relevan dalam konteks efisiensi, tetapi memiliki keterbatasan dalam menangani batuan dengan kemiripan visual tinggi menegaskan nilai strategis dari pendekatan *ensemble* dalam klasifikasi citra geospasial yang kompleks seperti ini.

4.6.3. Implikasi praktis dari model terbaik untuk klasifikasi batuan

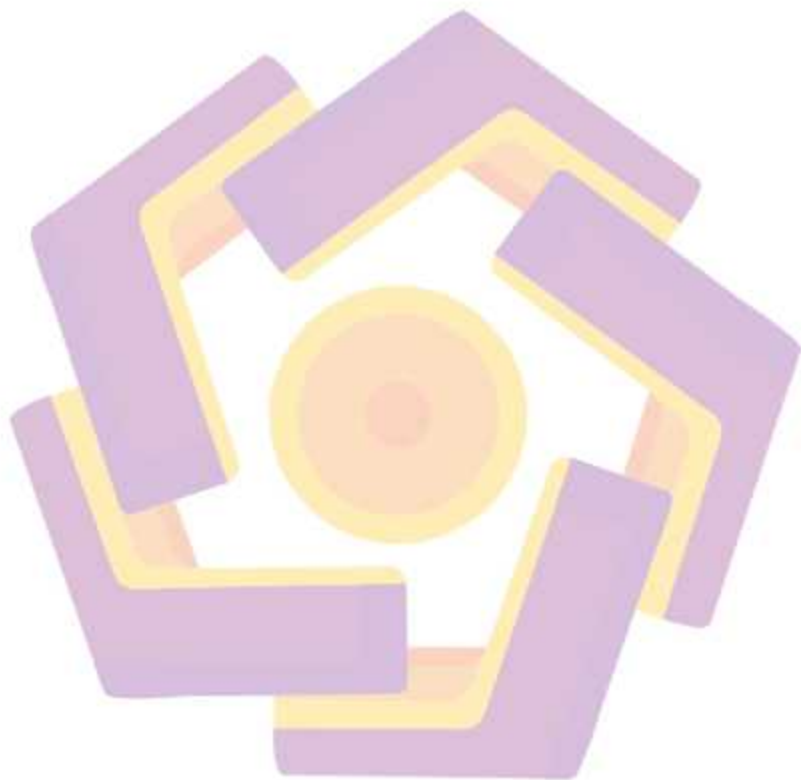
Setiap arsitektur CNN dalam pendekatan *ensemble stacking* memiliki peran yang saling melengkapi dalam meningkatkan akurasi klasifikasi citra batuan. Model berkapasitas besar seperti ResNet50 dan InceptionV3 unggul dalam mengekstraksi fitur

kompleks berupa pola tekstur dan struktur mendalam, sedangkan model ringan seperti MobileNetV2 dan EfficientNetB0 lebih efisien secara komputasi serta peka terhadap detail lokal. Perbedaan karakteristik ini memungkinkan sistem *stacking* memadukan representasi fitur global dan lokal secara optimal sehingga keputusan klasifikasi menjadi lebih seimbang dan *robust* terhadap variasi data.

Penggabungan model dengan karakteristik berbeda, sebagaimana pada konfigurasi Model 4 (ResNet50 + MobileNetV2), menghasilkan efek sinergis yang nyata. Model besar berperan sebagai fondasi akurasi yang stabil, sementara model ringan memperkaya fitur pembeda. Ketika seluruh fitur ini dikonsolidasikan oleh *meta-learner* SVM, sistem menunjukkan kinerja yang lebih presisi dan adaptif dibandingkan model tunggal, sekaligus membuktikan efektivitas pendekatan *stacking* dalam menangani kemiripan visual antar kelas batuan.



BAB V
PENUTUP



5.1. Kesimpulan

Penelitian ini berhasil mencapai seluruh tujuan yang telah ditetapkan, yaitu menilai perbedaan kinerja antar beberapa konfigurasi model *stacking*, membandingkan hasil akurasi, presisi, *recall*, dan *f1-score* model *stacking* dengan model *non-stacking*, serta menilai kontribusi masing-masing model *non-stacking* terhadap hasil akhir sistem *stacking*. Adapun kesimpulan yang diperoleh adalah sebagai berikut:

- a) Berdasarkan perbandingan beberapa konfigurasi model *stacking*, diperoleh perbedaan kinerja yang nyata antar model. Model *stacking* terbaik, yaitu kombinasi ResNet50 dan MobileNetV2 dengan *meta-learner Support Vector Machine* (SVM), menghasilkan akurasi tertinggi sebesar 84%, sedangkan konfigurasi lainnya berada pada kisaran 81%-83%. Hasil ini menunjukkan bahwa pemilihan kombinasi model berpengaruh signifikan terhadap kinerja akhir sistem *stacking*.
- b) Jika dibandingkan dengan model *non-stacking*, model *stacking* terbukti memberikan kinerja yang lebih unggul. Model tunggal terbaik, yaitu ResNet50, hanya mencapai akurasi sekitar 80%, sementara model *stacking* mampu meningkatkan akurasi hingga 84%, disertai peningkatan nilai *precision*, *recall*, dan *F1-score* yang lebih stabil. Hal ini membuktikan bahwa pendekatan *stacking* lebih efektif dibandingkan penggunaan satu model CNN tunggal.

- c) Setiap model *non-stacking* memiliki kontribusi yang saling melengkapi dalam sistem *stacking*. Model berarsitektur besar seperti ResNet50 dan InceptionV3 berperan dalam menangkap pola global dan tekstur mendalam, sedangkan model ringan seperti MobileNetV2 dan EfficientNetB0 lebih peka terhadap detail lokal. Sinergi antar model tersebut memungkinkan *meta-learner* SVM menghasilkan prediksi yang lebih presisi, stabil, dan memiliki kemampuan generalisasi yang lebih baik.
- d) Sebagai kesimpulan akhir, maka pendekatan *ensemble stacking* CNN terbukti efektif, fleksibel, dan adaptif dalam meningkatkan kinerja klasifikasi citra batuan multikategori, serta mampu mengatasi keterbatasan model tunggal dalam mengenali kompleksitas variasi tekstur, warna, dan pola visual batuan.

5.2. Saran

Hasil penelitian ini memperlihatkan bahwa teknik *ensemble stacking* mampu memberikan peningkatan dalam akurasi klasifikasi citra batuan. Namun demikian, keterbatasan jumlah data latih menjadi salah satu faktor yang menghambat hasil optimal model. Jumlah sampel yang sedikit dapat mengurangi kemampuan model dalam mengenali keragaman bentuk, tekstur, serta pola antar jenis batuan yang cenderung mirip. Oleh karena itu, studi lanjutan disarankan untuk menggunakan dataset yang lebih besar, seimbang, dan mencakup beragam karakteristik batuan agar fitur yang dipelajari menjadi lebih representatif.

Selain itu, teknik augmentasi yang digunakan masih bersifat dasar dan dapat ditingkatkan. Pengembangan augmentasi berbasis konteks geologi misalnya manipulasi

visual yang menyerupai kondisi pencahayaan alami atau orientasi khas batuan di lapangan dapat menjadi pendekatan yang lebih relevan dan efektif. Eksplorasi metode pembelajaran dengan sedikit label, seperti *semi-supervised learning* atau *self-supervised learning*, juga dapat menjadi alternatif untuk memaksimalkan pembelajaran dari data yang terbatas.

Ke depan, penggabungan metode *stacking* dengan arsitektur *deep learning* terbaru seperti Vision Transformer (ViT) atau model prelatih seperti *Segment Anything Model (SAM)* berpotensi memberikan peningkatan dalam kualitas fitur yang diekstraksi. Pengembangan aplikasi klasifikasi yang efisien dan dapat dioperasikan secara *real-time* di lapangan juga merupakan arah penelitian yang penting. Selain itu, uji coba model pada data dari berbagai lokasi atau kondisi geologis diperlukan untuk memastikan kemampuan generalisasi sistem dalam menghadapi data yang lebih kompleks dan bervariasi.

DAFTAR PUSTAKA

- Ali, M., Kumar, S., Pal, R., Singh, M. K., & Saini, D. (2023). Graph- and Machine-Learning-Based Texture Classification. *Electronics* (Switzerland). <https://doi.org/10.3390/electronics12224626>
- Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaria, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*. <https://doi.org/10.1186/s40537-021-00444-8>
- Atmajaya, Dedy & Febrianti, Annisa & Darwis, Herdianti. (2023). Metode SVM dan Naive Bayes untuk Analisis Sentimen ChatGPT di Twitter. *Indonesian Journal of Computer Science*. 12. 10.33022/ijcs.v12i4.3341.
- Bera, S. (2022). *Enhancing the Classification and Identification of Natural Rocks using Swin-Transformer Architecture* (Master's thesis, National College of Ireland). Available at <https://norma.ncirl.ie/6086/1/subhashreebera.pdf>
- Bressan, T. S., de Souza, M. K., Girelli, T. J. & Junior, F. C. Evaluation of machine learning methods for lithology classification using geophysical data. *Comput. Geosci.* **139**, 104475 (2020).
- Bhujel, S., & Shakya, S. (2022). Rice Leaf Diseases Classification Using Discriminative Fine Tuning and CLR on EfficientNet. *Journal of Soft Computing Paradigm*. <https://doi.org/10.36548/jscp.2022.3.006>
- Boiger, R., Churakov, S. V., Ballester Llagaria, I., Kosakowski, G., Wüst, R., & Prasianakis, N. I. (2024). Direct mineral content prediction from drill core

- images via transfer learning. *Swiss Journal of Geosciences*, 117(1), 8. <https://doi.org/10.1186/s00015-024-00458-3>
- Chen, H. C., Widodo, A. M., Wisnujati, A., Rahaman, M., Lin, J. C. W., Chen, L., & Weng, C. E. (2022). AlexNet Convolutional Neural Network for Disease Detection and Classification of Tomato Leaf. *Electronics (Switzerland)*, <https://doi.org/10.3390/electronics11060951>.
- Chen, W., Su, L., Chen, X., & Huang, Z. (2023). Rock image classification using deep residual neural network with transfer learning. *Frontiers in Earth Science*, 10, <https://doi.org/10.3389/feart.2022.1079447>
- K. Chawshin, C.F. Berg, D. Varagnolo, O. Lopez. 2021. Lithology classification of whole core CT scans using convolutional neural networks *SN Appl. Sci.*, 3 (2021), p. 668
- Darapaneni, N., Krishnamurthy, B., & Paduri, A. R. (2020). Convolution Neural Networks: A Comparative Study for Image Classification. 2020 IEEE 15th International Conference on Industrial and Information Systems, ICIIS 2020 - Proceedings. <https://doi.org/10.1109/ICIIS51140.2020.9342667>
- Dawson, H. L., Dubrule, O., & John, C. M. (2023). Impact of dataset size and convolutional neural network architecture on transfer learning for carbonate rock classification. *Computers & Geosciences*, 171, 105284. <https://doi.org/10.1016/j.cageo.2022.105284>
- dos Anjos, C.E.M., Avila, M.R.V., Vasconcelos, A.G.P. et al. Deep learning for lithological classification of carbonate rock micro-CT images. *Comput Geosci* 25, 971–983 (2021). <https://doi.org/10.1007/s10596-021-10033-6>

- Fauzi, M. D., Adhinata, F. D., Ramadhan, N. G., & Tanjung, N. A. F. (2022). A Hybrid DenseNet201-SVM for Robust Weed and Potato Plant Classification. *Jurnal Ilmiah Teknik Elektro Komputer Dan Informatika*. <https://doi.org/10.26555/jiteki.v8i2.23886>
- Geetha, S., Sharmila, V., Sasikala, S., Balamurugan, S. A. A., & Balamurugan, N. M. (2023). Detecting Gastro-Intestinal Cancer from Wireless Capsule Endoscopy Images using Efficient Net Model. 2023 14th International Conference on Information and Communication Technology and System, ICTS 2023. <https://doi.org/10.1109/ICTS58770.2023.10330886>
- Hossain, S., Uddin, J., Nahin, R., & Ibne Eunos, S. (2021). Rock Classification Dataset. <https://doi.org/10.34740/kaggle/ds/1293628>
- Huang, Z., Su, L., Wu, J., & Chen, Y. (2023). Rock Image Classification Based on EfficientNet and Triplet Attention Mechanism. *Applied Sciences*, 13(5), 3180. <https://doi.org/10.3390/app13053180>
- Khanam, R., & Mehta, G. (2023). Application of Deep CNN for image-based identification and classification of plant diseases. *Proceedings of the 10th International Conference on Signal Processing and Integrated Networks, SPIN 2023*. <https://doi.org/10.1109/SPIN57001.2023.10116967>
- Liu, X., & Song, H. (2020). Automatic identification of fossils and abiotic grains during carbonate microfacies analysis using deep convolutional neural networks. *Sedimentary Geology*, 410, 105790. <https://doi.org/10.1016/j.sedgeo.2020.105790>

- Ouedraogo, A. S., Kumar, A., & Wang, N. (2023). Landfill Waste Segregation Using Transfer and Ensemble Machine Learning: A Convolutional Neural Network Approach. *Energies*. <https://doi.org/10.3390/en16165980>
- Patro, S., Jhariya, D. C., Sahu, M., Dewangan, P., & Dhekne, P. Y. (2022). Igneous rock classification using Convolutional neural networks (CNN). *IOP Conference Series: Earth and Environmental Science*. <https://doi.org/10.1088/1755-1315/1032/1/012045>
- Sousa, J. V. M., Almeida, V. R. de, Saraiva, A. A., Santos, D. B. S., Pimentel, P. M. C., & Sousa, L. L. de. (2020). Classification of Pneumonia images on mobile devices with Quantized Neural Network. *Research, Society and Development*. <https://doi.org/10.33448/rsd-v9i10.8382>
- Ture, B. A., Akbulut, A., Zaim, A. H., & Catal, C. (2024). Stacking-based ensemble learning for remaining useful life estimation. *Soft Computing*. <https://doi.org/10.1007/s00500-023-08322-6>
- Yildiz, G., Ulu, A., Dizdaroglu, B., & Yildiz, D. (2023). Hybrid Image Improving and CNN (HIICNN) Stacking Ensemble Method for Traffic Sign Recognition. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2023.3292955>
- Zeng, G. (2020). On the confusion matrix in credit scoring and its analytical properties. *Communications in Statistics - Theory and Methods*, 49(9), 2080–2093. <https://doi.org/10.1080/03610926.2019.1568485>

Lampiran 1. Link Script Model Stacking dan Non Stacking

Model Stacking:

<https://colab.research.google.com/drive/IPNxqB5Os1FdNo9ZNGWhO30PjsK6sQTjD?usp=sharing>

Model Non Stacking:

<https://colab.research.google.com/drive/1qFD4aCTiv-ijNSfcatncM88ZwqQyRb8n?usp=sharing>

Dataset:

Preprocessing:

<https://colab.research.google.com/drive/1mtAYIKtBULu51w7cBsxxRIjsTVrtT1eS?usp=sharing>

Data Bersih:

https://drive.google.com/drive/folders/1w31_Fb2nLCvym8oyVDzAh58iNKyFyaez?usp=sharing

Data Testing:

https://drive.google.com/drive/folders/1owHGzbGONWVCwf78K_wYOpKWqx5F9IbR?usp=sharing