UNIVERSITAS
GADJAH MADA
**Evaluasi Performa dan Stabilitas Pembelajaran Arsitektur ConvNeXt untuk Klasifikasi Perdarahan Otak
pada Dataset CT Scan Terbatas**
Tengku Syaid Farhan, Afiahayati, S.Kom., M.Cs., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

DAFTAR ISI

Halaman Judul ii

Halaman Pengesahan iii

Halaman Pernyataan iv

KATA PENGANTAR v

INTISARI viii

ABSTRACT ix

DAFTAR ISI x

DAFTAR TABEL xvi

DAFTAR GAMBAR xx

DAFTAR KODE xxii

I PENDAHULUAN 1

1.1 Latar Belakang 1
1.2 Rumusan Masalah 4
1.3 Batasan Masalah 4
1.4 Tujuan Penelitian 6
1.5 Manfaat Penelitian 7
1.6 Struktur Penelitian 7
1.7 Sistematika Penulisan 9

II TINJAUAN PUSTAKA 11

III LANDASAN TEORI 17

3.1 Generalisasi dalam Deep Learning 17
3.1.1 Bias-Variance Tradeoff 17
3.1.2 Overfitting dan Underfitting 18
3.1.3 Faktor-Faktor yang Mempengaruhi Generalisasi 19

UNIVERSITAS
GADJAH MADA
**Evaluasi Performa dan Stabilitas Pembelajaran Arsitektur ConvNeXt untuk Klasifikasi Perdarahan Otak
pada Dataset CT Scan Terbatas**
Tengku Syaid Farhan, Afiahayati, S.Kom., M.Cs., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

3.2 Stabilisasi dalam Deep Learning 19
3.2.1 Pengukuran Stabilitas 21
3.3 Tantangan Deep Learning pada Dataset Terbatas 21
3.3.1 Karakteristik Dataset Medis 21
3.3.2 Mengapa Deep Learning Membutuhkan Data Besar 22
3.3.3 Risiko Overfitting pada Dataset Kecil 23
3.4 Convolutional Neural Network 23
3.4.1 Definisi dan Arsitektur CNN 23
3.4.2 Komponen Utama CNN 24
3.5 ConvNeXt: Modernisasi Arsitektur CNN 25
3.5.1 Desain dan Motivasi 25
3.5.2 Layer Normalization 26
3.5.3 Stochastic Depth Regularization 27
3.5.4 Inverted Bottleneck 29
3.5.5 GELU Activation Function 30
3.6 Metrik Evaluasi Model 31
3.6.1 Confusion Matrix dan Metrik Dasar 31
3.6.2 Area Under the ROC Curve (AUC-ROC) 32
3.6.3 Area Under Precision-Recall Curve (AUC-PR) 32
3.6.4 Loss Function 33
3.7 Computed Tomography Scan dan Intracranial Hemorrhage 33
3.7.1 Prinsip CT Scan 33
3.7.2 Hounsfield Unit (HU) 34
3.7.3 Format DICOM 34
3.7.4 Windowing 35
3.8 Teknik Pelatihan 36
3.8.1 Early Stopping 36
3.8.2 Learning Rate Scheduling 36
3.9 Transfer Learning 37
3.9.1 Konsep Transfer Learning 37
3.10 Uji Statistik 40
3.10.1 Bootstrap Confidence Interval 40
3.10.2 McNemar Test 41

IV METODOLOGI 43

UNIVERSITAS
GADJAH MADA
**Evaluasi Performa dan Stabilitas Pembelajaran Arsitektur ConvNeXt untuk Klasifikasi Perdarahan Otak
pada Dataset CT Scan Terbatas**
Tengku Syaid Farhan, Afiahayati, S.Kom., M.Cs., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

4.1 Deskripsi Umum Sistem 43
4.2 Lingkungan 44
4.2.1 Spesifikasi Hardware 44
4.2.2 Spesifikasi Software 45
4.3 Dataset 46
4.3.1 Dataset CQ500 46
4.3.2 Dataset RSNA 48
4.3.3 Perbandingan Dataset 49
4.3.4 Skema Pelabelan Binary 50
4.3.5 Pelabelan Slice (Weak Supervision) 51
4.3.6 Subset RSNA 52
4.4 Pembagian Data 53
4.4.1 Stratified Split 53
4.4.2 Pencegahan Data Leakage 55
4.5 Preprocessing 57
4.5.1 Pembacaan File DICOM 57
4.5.2 Hounsfield Unit dan Brain Windowing 57
4.5.3 Normalisasi dan Resize 58
4.5.4 Strategi *Caching Preprocessing* 59
4.6 Arsitektur 60
4.6.1 Spesifikasi Arsitektur 60
4.7 Baseline Machine Learning 62
4.7.1 Modifikasi Classification Head 62
4.8 Protokol Pelatihan 64
4.8.1 Transfer Learning 64
4.8.2 Loss Function 64
4.8.3 Optimizer dan Learning Rate Schedule 65
4.8.4 Data Augmentation 67
4.8.5 Hyperparameter 67
4.8.6 Early Stopping 70
4.8.7 Protokol Replikasi Eksperimen 70
4.9 Analisis Sensitivitas Ukuran Dataset 70
4.9.1 Konfigurasi Subset 70
4.9.2 Protokol Eksperimen 71
4.9.3 Analisis Crossover Point 71

UNIVERSITAS
GADJAH MADA
Evaluasi Performa dan Stabilitas Pembelajaran Arsitektur ConvNeXt untuk Klasifikasi Perdarahan Otak pada Dataset CT Scan Terbatas
Tengku Syaid Farhan, Afiahayati, S.Kom., M.Cs., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

4.10 Pipeline Inferensi dan Agregasi 72
4.10.1 Prediksi Level Slice 72
4.10.2 Agregasi Kombinasi Top-K Mean dan Exponentially Weighted Mean 72
4.10.3 Optimasi Threshold 73
4.10.4 Klasifikasi Final 74
4.11 Metrik Evaluasi 74
4.11.1 Metrik Performa (RQ1) 75
4.11.2 Metrik Stabilitas (RQ2) 75
4.11.3 Metrik Generalisasi (RQ3) 76
4.11.4 Metrik Sensitivitas Data (RQ4) 77
4.11.5 Agregasi Metrik Lintas Eksperimen 77
4.11.6 Ringkasan Metrik Evaluasi 78
4.12 Analisis Statistik 78
4.12.1 Bootstrap Confidence Interval 78

V IMPLEMENTASI 80
5.1 Spesifikasi 80
5.1.1 Perangkat Keras 80
5.1.2 Perangkat Lunak 81
5.1.3 Inisialisasi *Library* 81
5.1.4 Data Path 82
5.2 Persiapan Data 82
5.2.1 Konstruksi Metadata DICOM 82
5.2.2 Konstruksi Label dengan Majority Voting 84
5.2.3 Definisi Eksperimen dan Kriteria Inklusi 86
5.2.4 Stratified Patient-Level Split 87
5.2.5 Verifikasi Data Leakage 89
5.3 Preprocessing 90
5.3.1 Pembacaan File DICOM dan Konversi Hounsfield Unit 90
5.3.2 Brain Windowing 90
5.3.3 Resize dan Normalisasi ImageNet 91
5.3.4 Caching Preprocessing 92
5.4 Arsitektur 94
5.4.1 Reproducibility dan Konfigurasi Global 94

UNIVERSITAS
GADJAH MADA
**Evaluasi Performa dan Stabilitas Pembelajaran Arsitektur ConvNeXt untuk Klasifikasi Perdarahan Otak
pada Dataset CT Scan Terbatas**
Tengku Syaid Farhan, Afiahayati, S.Kom., M.Cs., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

5.4.2 Dataset dan Data Augmentation 95
5.4.3 Arsitektur Model dan Classification Head 98
5.4.4 Layer-wise Learning Rate Decay dan Loss Function 100
5.5 Pelatihan 103
5.5.1 DataLoader dan Weighted Random Sampling 103
5.6 Pipeline dan Evaluasi 105
5.6.1 Eksekusi Pelatihan Multi-Backbone 105
5.6.2 Implementasi Fungsi Agregasi Prediksi Pasien 105
5.6.3 Evaluasi pada Test Set Internal 108
5.6.4 Evaluasi pada Test Set Internal dengan Bootstrap Confidence Interval 109
5.6.5 Evaluasi Generalisasi Eksternal pada Dataset RSNA 111
5.6.6 Analisis Stabilitas Validasi AUC Antar-Epoch 116
5.6.7 Ekstraksi Fitur *Radiomics* untuk Perbandingan *Machine Learning* 116

VI HASIL DAN PEMBAHASAN 119
6.1 Gambaran Umum Dataset dan Distribusi Kelas 119
6.2 Perbandingan Performa Arsitektur CNN pada Dataset CQ500 119
6.2.1 Eksperimen Any-ICH vs Normal 120
6.2.2 Eksperimen IPH vs Normal 122
6.2.3 Eksperimen SDH vs Normal 125
6.2.4 Eksperimen SAH vs Normal 128
6.2.5 Ringkasan Komparatif Lintas Eksperimen 132
6.3 Perbandingan *Deep Learning* versus *Machine Learning* Konvensional 133
6.4 Analisis *Learning Curve* dan Efisiensi Data 135
6.5 Validasi Generalisasi Eksternal pada Dataset RSNA 136
6.5.1 Konfigurasi Dataset Validasi Eksternal 137
6.5.2 Hasil Evaluasi pada Dataset RSNA 137
6.5.3 Kuantifikasi *Domain Gap* dan Indikator *Domain Shift* 139
6.5.4 Interpretasi *Domain Shift* dan Rekomendasi Teknis 139
6.6 Stabilitas Pelatihan dan Analisis Generalisasi Internal 141
6.6.1 Stabilitas Validasi AUC Antar-Epoch 141
6.6.2 *Train-Val Gap* sebagai Indikator *Overfitting* 143
6.6.3 Statistik Pelatihan dan Efisiensi Konvergensi 146

UNIVERSITAS
GADJAH MADA
**Evaluasi Performa dan Stabilitas Pembelajaran Arsitektur ConvNeXt untuk Klasifikasi Perdarahan Otak
pada Dataset CT Scan Terbatas**
Tengku Syaid Farhan, Afiahayati, S.Kom., M.Cs., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

6.6.4 Analisis *Bootstrap Confidence Interval* Lintas Eksperimen . . 147
6.7 Perbandingan dengan Penelitian Terdahulu 148
6.7.1 Batas Komparabilitas 148
6.7.2 Posisi Hasil Terhadap Literatur 148
6.7.3 Perbandingan Generalisasi Lintas Dataset 149
6.8 Makna Hasil dalam Konteks Medis 150
6.8.1 Makna Klinis per Subtipe ICH 150
6.8.2 AI sebagai Alat Bantu Triase, Bukan Pengganti Radiolog . . . 151
6.8.3 Stabilitas sebagai Syarat Keandalan Klinis 151
6.8.4 Keterbatasan Generalisasi dan Relevansinya untuk Indonesia . 152
6.8.5 Limitasi Penelitian 152

VII PENUTUP **154**
7.1 Kesimpulan 154
7.2 Saran 155

DAFTAR PUSTAKA **157**