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/

INTISARI

EVALUASI PERFORMA DAN STABILITAS PEMBELAJARAN ARSITEKTUR CONVNEXT UNTUK KLASIFIKASI PERDARAHAN OTAK PADA DATASET CT SCAN TERBATAS

Oleh

Tengku Syaid Farhan
22/493998/PA/21234

Perdarahan intrakranial (*Intracranial Hemorrhage*, ICH) menyumbang sekitar 29% kejadian stroke dunia dan memerlukan diagnosis dalam *golden hour*, namun ketersediaan radiolog yang terbatas menjadi hambatan. Sistem kecerdasan buatan berpotensi menjadi alat bantu triase untuk mempercepat identifikasi kasus prioritas, bukan menggantikan peran radiolog. Namun, 89% penelitian klasifikasi ICH menggunakan dataset >5.000 sampel yang tidak sebanding dengan data yang dimiliki sebagian besar rumah sakit, dan tidak satupun melaporkan analisis stabilitas pembelajaran. *Gap* ini mendorong pertanyaan apakah arsitektur modern yang mengadopsi prinsip desain *Vision Transformer*, dengan komponen yang secara teoritis relevan untuk kondisi data terbatas, mampu memberikan performa yang lebih stabil dibanding arsitektur konvensional pada skala data realistis tingkat rumah sakit. Penelitian ini mengevaluasi *ConvNeXt-Tiny*, yang meskipun telah menunjukkan keberhasilan pada domain medis terbatas lain seperti deteksi *leukoencephalopathy* (Cernekova et al., 2024) dan klasifikasi Alzheimer berbasis MRI (Xia et al., 2025), belum pernah diuji pada klasifikasi biner ICH dengan dataset <500 sampel. Model ini dibandingkan *ResNet-50*, *EfficientNet-B0*, dan LightGBM dengan fitur *radiomics* pada dataset CQ500, serta divalidasi secara eksternal pada dataset RSNA untuk mengukur *domain shift*. Evaluasi mencakup empat skenario klasifikasi biner (Any-ICH, IPH, SDH, SAH) dengan metrik utama AUC-ROC dan analisis stabilitas melalui varians AUC antar-*epoch*. *ConvNeXt-Tiny* mengungguli LightGBM sebesar 45,88% secara relatif, melampaui VGG16 (Castro et al., 2019) sebesar 12,6% meski menggunakan porsi pelatihan lebih kecil, serta mempertahankan *domain gap* terkecil ($\Delta\text{AUC} = -0{,}1865$) dan AUC tertinggi 0,6949 pada validasi RSNA. Komponen desain *ConvNeXt* berkontribusi pada stabilitas pembelajaran, dengan AUC tertinggi 0,9141 pada subtipe SAH. Penelitian ini menegaskan bahwa evaluasi stabilitas dan generalisasi merupakan indikator kelayakan klinis yang tidak dapat diabaikan.

Kata kunci: perdarahan intrakranial, *ConvNeXt*, klasifikasi CT, dataset terbatas, stabilitas pembelajaran, generalisasi lintas domain

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/

ABSTRACT

Evaluation of Performance and Learning Stability of ConvNeXt Architecture for Intracranial Hemorrhage Classification on Limited CT Scan Datasets

By

Tengku Syaid Farhan
22/493998/PA/21234

Intracranial hemorrhage (ICH) accounts for approximately 29% of global stroke cases and requires diagnosis within the golden hour, yet the limited availability of radiologists remains a barrier. Artificial intelligence systems have potential as triage aids to accelerate the identification of priority cases, not to replace radiologists. However, 89% of ICH classification studies use datasets of more than 5,000 samples, which is not comparable to the data available at most hospitals, and none report learning stability analysis. This gap raises the question of whether a modern architecture adopting *Vision Transformer* design principles, with components theoretically suited for limited-data conditions, can achieve more stable performance than conventional architectures at realistic hospital-scale data volumes. This study evaluates *ConvNeXt-Tiny*, which has demonstrated success on other limited medical imaging domains such as leukoencephalopathy detection (Cernekova et al., 2024) and Alzheimer's classification from MRI (Xia et al., 2025), yet has never been tested on binary ICH classification with fewer than 500 samples. It is compared against *ResNet-50*, *EfficientNet-B0*, and LightGBM with radiomics features on the CQ500 dataset, with external validation on the RSNA dataset to measure domain shift. Evaluation covers four binary classification scenarios (Any-ICH, IPH, SDH, SAH) using AUC-ROC as the primary metric and learning stability analysis through per-epoch AUC variance. *ConvNeXt-Tiny* outperforms LightGBM by 45.88% relatively, surpasses VGG16 (Castro et al., 2019) by 12.6% despite using a smaller training split, and maintains the smallest domain gap ($\Delta\text{AUC} = -0.1865$) alongside the highest external AUC 0.6949 on RSNA validation. *ConvNeXt* design components contribute to learning stability, with a peak AUC of 0.9141 on the SAH subtype. This study confirms that stability and generalization evaluation are clinical readiness indicators that cannot be overlooked.

Keywords: intracranial hemorrhage, *ConvNeXt*, CT classification, limited dataset, learning stability, cross-domain generalization