UNIVERSITAS
GADJAH MADA
**Deteksi Strabismus Berbasis Convolutional Neural Network dengan Mekanisme Attention dan Evaluasi
Interpretabilitas Menggunakan Gradient-Weighted Class Activation Mapping**
Novanda Havid Ramadani, Dr.Eng. Ir. Sunu Wibirama, S.T., M.Eng., IPM.;Dr.Eng. Ir. Igi Ardiyanto, S.T., M.Eng., IPM
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

INTISARI

Strabismus (juling) merupakan gangguan ketidaksejajaran sumbu visual. Kondisi ini memengaruhi 2–5% populasi umum secara global. Di Indonesia, prevalensi strabismus pada anak usia 4–10 tahun mencapai 2–4%, dengan eksotropia sebagai tipe yang paling umum ditemukan. Strabismus menimbulkan tantangan diagnostik yang signifikan akibat dampak klinisnya yang multidimensional serta distribusi tenaga spesialis oftalmologi yang tidak merata. Deteksi dini strabismus sangat penting karena keterlambatan penanganan dapat menyebabkan ambliopia permanen dan gangguan penglihatan binokular jangka panjang.

Pendekatan deteksi berbasis *computer vision* dan kecerdasan buatan dibutuhkan karena pemeriksaan manual oleh dokter spesialis bersifat subjektif, membutuhkan waktu, dan sulit diakses secara merata di daerah dengan keterbatasan tenaga ahli. Meskipun pendekatan berbasis *convolutional neural network* (CNN) telah menunjukkan akurasi yang menjanjikan dalam deteksi strabismus otomatis, dua keterbatasan utama masih belum teratasi. Pertama, sifat *black box* pada prediksi CNN membuat dokter tidak dapat mengetahui alasan di balik keputusan model, sehingga sulit dipercaya dalam konteks klinis. Kedua, evaluasi interpretabilitas selama ini bersifat kualitatif dan tidak dapat dibandingkan secara objektif antarsistem, sehingga menghambat validasi dan adopsi klinis. *Explainable Artificial Intelligence* (XAI) adalah pendekatan yang bertujuan membuat keputusan model AI dapat dipahami dan dijelaskan kepada pengguna, termasuk tenaga medis. Penelitian ini mengatasi kedua keterbatasan tersebut dengan mengusulkan kerangka terpadu yang membandingkan enam *backbone* CNN *pretrained* dengan mekanisme *channel attention* Squeeze-and-Excitation (SE), optimasi *hyperparameter* Bayesian (HPO), dan protokol evaluasi XAI kuantitatif yang komprehensif.

Enam arsitektur, yaitu EfficientNetB0, MobileNetV3Large, DenseNet121, ResNet50, VGG16, dan InceptionResNetV2, dipilih karena mewakili spektrum desain CNN yang beragam, mulai dari arsitektur ringan (*mobile-friendly*) hingga arsitektur dalam berbasis *residual* dan *dense connection*, sehingga memungkinkan perbandingan yang komprehensif antara efisiensi komputasi dan kapasitas representasi fitur. Penelitian ini menjalankan tiga kondisi eksperimen. Eksperimen I menggunakan konfigurasi *default* tanpa HPO, Eksperimen II menambahkan Bayesian HPO, dan Eksperimen III merupakan replikasi Eksperimen II untuk memverifikasi determinisme *pipeline*. Seluruh eksperimen menggunakan *cache* augmentasi statis yang identik dan pembagian data stratifikasi anti-*leakage* pada 450 citra mata klinis (105 Normal, 345 Strabismus). Interpretabilitas dievaluasi dengan menggunakan tiga metrik Grad-CAM pada layer `att_conv1x1`, yaitu *Deletion AUC* (DAUC), *Insertion AUC* (IAUC), dan *Localization Score* geometris yang mencakup *Peak-in-ROI* dan *Top-10%-in-ROI*.

DenseNet121 mencapai performa klasifikasi terbaik setelah Bayesian HPO, dengan akurasi 93,8%, sensitivitas 93,8%, dan spesifisitas 93,8%, serta AUC-ROC 0,977. EfficientNetB0 menunjukkan peningkatan akurasi terbesar (+15,6 pp), mengkonfirmasi bahwa Bayesian HPO paling efektif untuk arsitektur yang sensitif terhadap konfigurasi *default*, yakni pengaturan *hyperparameter* bawaan tanpa penyesuaian khusus. Dari sudut pandang interpretabilitas, ResNet50 secara konsisten menghasilkan *Localization Score* tertinggi di semua eksperimen, yaitu 0,533 pada Eksperimen I dan 0,650 setelah HPO pada Eksperimen II, yang mengindikasikan bahwa arsitektur *residual connection-*

UNIVERSITAS
GADJAH MADA
**Deteksi Strabismus Berbasis Convolutional Neural Network dengan Mekanisme Attention dan Evaluasi
Interpretabilitas Menggunakan Gradient-Weighted Class Activation Mapping**
Novanda Havid Ramadani, Dr.Eng. Ir. Sunu Wibirama, S.T., M.Eng., IPM.;Dr.Eng. Ir. Igi Ardiyanto, S.T., M.Eng., IPM
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

nya memfasilitasi fokus aktivasi yang lebih presisi secara anatomis. InceptionResNetV2, meskipun memiliki *cross-validation accuracy* yang kompetitif, menghasilkan *Localization Score* terendah di semua kondisi, yaitu 0,174 dan 0,248, dengan nilai *Peak-in-ROI* = 0,000 pada Eksperimen I, yang mengindikasikan potensi perilaku *shortcut learning*.

Analisis statistik menggunakan Friedman Test, Wilcoxon *Signed-Ranks* dengan koreksi Holm, dan McNemar's Test mengungkapkan pola diferensiasi yang bergantung pada kondisi eksperimen. Perbedaan akurasi mencapai signifikansi statistik hanya pada Eksperimen II ($\chi^2 = 14{,}24$, $p = 0{,}014$), sementara perbedaan *faithfulness* DAUC signifikan hanya pada Eksperimen I ($\chi^2 = 12{,}68$, $p = 0{,}027$). *Localization Score* menunjukkan daya diskriminasi terkuat dengan mencapai signifikansi secara konsisten di Eksperimen I dan II ($p < 0{,}0001$). Kesamaan sempurna antara Eksperimen II dan III mengkonfirmasi determinisme penuh *pipeline*. Berdasarkan temuan ini, DenseNet121 direkomendasikan sebagai arsitektur terbaik untuk klasifikasi strabismus, sementara ResNet50 direkomendasikan sebagai arsitektur dengan interpretabilitas *heatmap* terbaik. Secara keseluruhan, temuan ini menunjukkan bahwa akurasi klasifikasi dan kualitas interpretabilitas *heatmap* merupakan dua properti yang berbeda dan sebagian independen dari arsitektur CNN, sehingga pemilihan model untuk *deployment* klinis harus mempertimbangkan kedua dimensi tersebut secara bersamaan.

Kata Kunci: deteksi strabismus, *convolutional neural network*, *channel attention*, optimasi *hyperparameter* Bayesian, *explainable artificial intelligence*

UNIVERSITAS
GADJAH MADA
**Deteksi Strabismus Berbasis Convolutional Neural Network dengan Mekanisme Attention dan Evaluasi
Interpretabilitas Menggunakan Gradient-Weighted Class Activation Mapping**
Novanda Havid Ramadani, Dr.Eng. Ir. Sunu Wibirama, S.T., M.Eng., IPM.;Dr.Eng. Ir. Igi Ardiyanto, S.T., M.Eng., IPM
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

ABSTRACT

Strabismus is a visual axis misalignment disorder that affects 2–5% of the general population worldwide. In Indonesia, the prevalence of strabismus in children aged 4–10 years reaches 2–4%, with exotropia being the most commonly found type. Strabismus poses significant diagnostic challenges due to its multidimensional clinical impact and the uneven distribution of ophthalmology specialists. Early detection is critical, as delayed treatment can lead to permanent amblyopia and long-term binocular vision impairment.

Computer vision and artificial intelligence-based detection approaches are needed because manual examination by specialists is subjective, time-consuming, and difficult to access equitably in areas with limited expert availability. Although *convolutional neural network* (CNN)-based approaches have shown promising accuracy in automated strabismus detection, two major limitations remain unresolved. First, the *black box* nature of CNN predictions prevents clinicians from understanding the reasoning behind model decisions, making them difficult to trust in clinical settings. Second, interpretability evaluation has thus far been qualitative and cannot be objectively compared across systems, thereby hindering clinical validation and adoption. *Explainable Artificial Intelligence* (XAI) is an approach aimed at making AI model decisions understandable and explainable to users, including medical professionals.This study addresses both limitations by proposing a unified framework that compares six pre-trained backbones with a Squeeze-and-Excitation (SE) channel attention mechanism, Bayesian hyperparameter optimization (HPO), and a comprehensive quantitative XAI evaluation protocol.

Six architectures, namely EfficientNetB0, MobileNetV3Large, DenseNet121, ResNet50, VGG16, and InceptionResNetV2, were selected because they represent a diverse spectrum of CNN designs, ranging from lightweight (*mobile-friendly*) architectures to deep architectures based on *residual* and *dense connections*, enabling a comprehensive comparison between computational efficiency and feature representation capacity. This study ran three experimental conditions. Experiment I used the default configuration without HPO, Experiment II added Bayesian HPO, and Experiment III was a replication of Experiment II to verify full pipeline determinism. All experiments used an identical static augmentation cache and anti-leakage stratified data splitting on 450 clinical eye images (105 Normal, 345 Strabismus). Interpretability was evaluated using three Grad-CAM metrics on the `att_conv1x1` layer, namely *Deletion AUC* (DAUC), *Insertion AUC* (IAUC), and a geometric *Localization Score* encompassing *Peak-in-ROI* and *Top-10%-in-ROI*.

DenseNet121 achieved the best classification performance after Bayesian HPO, with accuracy, sensitivity, and specificity each at 93.8%, and an AUC-ROC of 0.977. EfficientNetB0 demonstrated the largest accuracy gain (+15.6 pp), confirming that Bayesian HPO is most effective for architectures sensitive to default configurations, that is, built-in hyperparameter settings without custom adjustment. From an interpretability standpoint, ResNet50 consistently produced the highest *Localization Score* across all experiments, reaching 0.533 in Experiment I and 0.650 after HPO in Experiment II, indicating that its *residual connection* architecture facilitates more anatomically precise activation focus. InceptionResNetV2, despite having competitive *cross-validation accuracy*, produced the lowest *Localization Score* across all conditions, at 0.174 and 0.248, with a *Peak-in-ROI* value of 0.000 in Experiment I, indicating potential *shortcut learning* behavior.

UNIVERSITAS
GADJAH MADA
**Deteksi Strabismus Berbasis Convolutional Neural Network dengan Mekanisme Attention dan Evaluasi
Interpretabilitas Menggunakan Gradient-Weighted Class Activation Mapping**
Novanda Havid Ramadani, Dr.Eng. Ir. Sunu Wibirama, S.T., M.Eng., IPM.;Dr.Eng. Ir. Igi Ardiyanto, S.T., M.Eng., IPM
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

Statistical analysis using the Friedman Test, Wilcoxon Signed-Ranks with Holm correction, and McNemar's Test revealed differentiation patterns that depend on the experimental condition. Accuracy differences reached statistical significance only in Experiment II ($\chi^2 = 14.24$, $p = 0.014$), while DAUC faithfulness differences were significant only in Experiment I ($\chi^2 = 12.68$, $p = 0.027$). The *Localization Score* demonstrated the strongest discriminative power, consistently reaching significance in both Experiment I and II ($p < 0.0001$). Perfect agreement between Experiment II and III confirmed full pipeline determinism. Based on these findings, DenseNet121 is recommended as the best architecture for strabismus classification, while ResNet50 is recommended as the architecture with the best *heatmap* interpretability. Overall, these findings indicate that classification accuracy and *heatmap* interpretability quality are two distinct and partially independent properties of CNN architectures, such that model selection for clinical deployment must consider both dimensions simultaneously.

Keywords: strabismus detection, *convolutional neural network*, *channel attention*, Bayesian hyperparameter optimization, *explainable artificial intelligence*