

Gangguan spektrum autisme (*autism spectrum disorder*—ASD) merupakan gangguan perkembangan saraf yang memerlukan deteksi dini yang andal sekaligus mudah dijelaskan secara klinis. Tantangan utama penerapan *deep learning* dalam bidang medis adalah sifatnya yang bersifat *black box*, sehingga hasil deteksi sering kali sulit diinterpretasikan oleh praktisi klinis dan bahkan tidak dapat diterima oleh pasien atau pengguna, terlebih ketika kondisi kritis dalam diagnosis dan prognosis. Penelitian ini bertujuan mengembangkan model *explainable deep learning* berbasis arsitektur Xception dengan metode *Gradient-weighted Class Activation Mapping* (Grad-CAM) untuk mendeteksi ASD melalui citra wajah anak secara akurat dan interpretatif. Data berupa 2.940 citra wajah anak (usia 2–14 tahun) dari Autism_Image_Dataset (Kaggle) dibagi menjadi *training*, *validation*, dan *testing dataset* dengan rasio 80:10:10. Tahapan penelitian meliputi *face detection* MTCNN, *data preprocessing*, augmentasi yang melibatkan *horizontal flipping*, klasifikasi dua kelas (*Autistic* dan *Non-Autistic*) menggunakan pendekatan *transfer learning*, hingga interpretasi dan evaluasi *Explainable Artificial Intelligence* (XAI). Hasil pengujian menunjukkan performa terbaik pada skema *deeper fine-tuning* dengan akurasi 0,89, AUC 0,95, presisi 0,86, *recall* 0,92, dan F1-score 0,89 pada kelas *Autistic*. Aspek interpretabilitas dicapai melalui visualisasi *heatmap* Grad-CAM yang memperlihatkan fokus konsisten pada area *midface* (antara mata dan hidung) pada citra anak autistik, serta area perhatian yang lebih menyebar ke area antara mata, pangkal dan ujung hidung, bibir atas dan bawah, bahkan dahi pada anak non-autistik. Dengan kerangka evaluasi *Technology Acceptance Model* (TAM) dan *Human-Centred Artificial Intelligence* (HCAI), hasil evaluasi interpretabilitas oleh tiga praktisi medis menunjukkan bahwa visualisasi Grad-CAM mudah dipahami dengan masih diperlukannya peningkatan pada aspek relevansi klinis, kolaborasi dengan manusia, dan konsistensi penilaian antar evaluator (*inter-rater reliability*). Secara keseluruhan, penelitian ini menunjukkan bahwa kombinasi Xception–Grad-CAM yang disertai validasi praktisi medis ini menghadirkan keseimbangan model yang tidak hanya akurat, tetapi juga transparan, dapat dijelaskan, dan divalidasi dalam konteks klinis. Sistem *explainable deep learning* ini berpotensi digunakan sebagai alat bantu skrining awal ciri autistik, dengan arah pengembangan selanjutnya pada perluasan data primer, peningkatan interpretabilitas visual, serta validasi klinis berskala lebih luas untuk memperkuat penerimaan sistem di lingkungan medis.

Kata kunci : *Autism Spectrum Disorder*, Klasifikasi Citra Wajah, *Transfer Learning*, *Explainable AI*, Xception, Grad-CAM

ABSTRACT

Autism spectrum disorder (ASD) is a neurodevelopmental disorder which requires early detection that is both reliable and clinically interpretable. This study develops an explainable deep learning pipeline for two-class screening from children's facial images. A dataset of 2,940 images (ages 2–14) from the Autism_Image_Dataset (Kaggle) was split into training, validation, and testing sets with an 80:10:10 ratio. The pipeline comprises face detection using MTCNN, data preprocessing, augmentation (including horizontal flipping), model training and testing using a transfer learning approach, followed by interpretation with Grad-CAM and its evaluation. The best-performing model, deeper fine-tuning by unfreezing the last 80 layers, achieved an accuracy of 0.89, AUC 0.95, precision 0.86, recall 0.92, and F1-score 0.89 for the Autistic class. Grad-CAM consistently highlighted midface regions, including the interocular area and nasal bridge for Autistic cases, with more diffuse attention to eyes, nasal root and tip, upper and lower lips, and sometimes forehead for Non-Autistic cases. Interpretability evaluation was further conducted using the Technology Acceptance Model (TAM) and Human-Centered Artificial Intelligence (HCAI) frameworks by three medical practitioners, indicating that Grad-CAM was easy to interpret but required improvement in clinical relevance, trustworthiness, and inter-rater reliability. Overall, the combination of Xception–Grad-CAM with medical validation produced a balanced model that is not only accurate but also transparent and clinically explainable. The proposed explainable deep learning system has the potential to serve as an early screening tool for autistic traits, with future work focusing on expanding primary data, enhancing visualization interpretability, and conducting broader clinical validation to strengthen medical acceptance.

Keywords : Autism Spectrum Disorder, Facial Image Classification, Transfer Learning, Explainable AI, Xception, Grad-CAM