

Performansi Seleksi Fitur pada Metode Multi Klasifikasi untuk Deteksi Dini Autisme Berbasis Citra Wajah Anak

Gede Angga Pradipta¹, Putu Desiana Wulaning Ayu²

Institut Teknologi dan Bisnis STIKOM Bali

e-mail: ¹angga_pradipta@stikom-bali.ac.id, ²wulaning_ayu@stikom-bali.ac.id

Diajukan: 26 Januari 2024; Direvisi: 26 Februari 2024; Diterima: 1 Maret 2024

Abstrak

Deteksi dini terhadap gangguan spektrum autisme (ASD) pada anak sangat penting untuk memberikan intervensi dan terapi tepat waktu. Deteksi dini secara tepat dapat membantu meningkatkan kualitas hidup anak yang terindikasi ASD. Metode pendekatan deteksi dapat dilakukan dengan observasi klinis dan kuesioner psikologi, tetapi metode ini sering kali subjektif dan membutuhkan waktu dalam mengetahui hasilnya. Sehingga dengan melihat permasalahan yang ada, maka penelitian ini bertujuan bertujuan untuk mengembangkan dan mengevaluasi model klasifikasi berbasis multi-klasifikasi dan metode seleksi serta jumlah fitur pada citra wajah untuk mendeteksi secara dini pada ASD. Hasil pengujian menunjukkan perpaduan metode klasifikasi logistik regresi dengan seleksi fitur ANOVA dengan menggunakan 150 fitur menghasilkan performansi terbaik dari sisi akurasi sebesar 0.9688, presisi sebesar 0.9687, dan recall sebesar 0.9688, dibandingkan dengan penggunaan metode seleksi fitur Information Gain. Hasil pengujian menunjukkan metode Logistic Linear Regression memiliki keunggulan dalam melakukan klasifikasi pada kelas biner dengan fitur yang terbatas.

Kata kunci: Autisme, Anak, Deteksi, Machine learning, Multi-klasifikasi.

Abstract

Early detection of autism spectrum disorder (ASD) in children is crucial for providing timely intervention and therapy. Accurate early detection can help improve the quality of life for children indicated with ASD. Detection approaches can be carried out through clinical observation and psychological questionnaires, but these methods are often subjective and time-consuming in obtaining results. Therefore, addressing these issues, this research aims to develop and evaluate a multi-classification-based model and selection methods along with the number of features in facial images to detect ASD early. The testing results show that the combination of logistic regression classification methods with ANOVA feature selection using 150 features yields the best performance in terms of accuracy (0.9688), precision (0.9687), and recall (0.9688), compared to the use of Information Gain feature selection methods. The testing results indicate that the logistic linear method has advantages in performing classification on binary classes with limited features.

Keywords: Autism, Child, Detection, Machine learning, Multi classifier.

1. Pendahuluan

Gangguan spektrum autisme (ASD) adalah kondisi perkembangan neurologis yang kompleks, mempengaruhi kemampuan individu untuk berkomunikasi dan berinteraksi secara sosial, serta ditandai oleh perilaku terbatas dan repetitif [1]. Deteksi dini ASD sangat penting untuk memberikan intervensi dan terapi yang tepat waktu, yang dapat membantu meningkatkan kualitas hidup anak yang terindikasi ASD. Metode dalam mendeteksi atau mendiagnosa ASD dapat dilakukan dengan pendekatan observasi secara klinis dan kuesioner psikologis, tetapi metode ini sering kali subjektif dan membutuhkan waktu dalam mengetahui hasilnya [2]. Oleh karena itu, diperlukan pendekatan otomatis yang dapat mendeteksi atau mendiagnosis ASD secara lebih efisien dan akurat.

Dalam beberapa tahun terakhir, kemajuan dalam bidang *machine learning* dan pengolahan citra telah membuka peluang baru untuk deteksi dini ASD. Salah satu pendekatan yang menjanjikan adalah penggunaan metode multi-klasifikasi (Multi-classifier) dan seleksi fitur pada citra wajah anak [3]. Metode ini melibatkan penggunaan berbagai algoritma *machine learning* untuk mengklasifikasikan gambar wajah

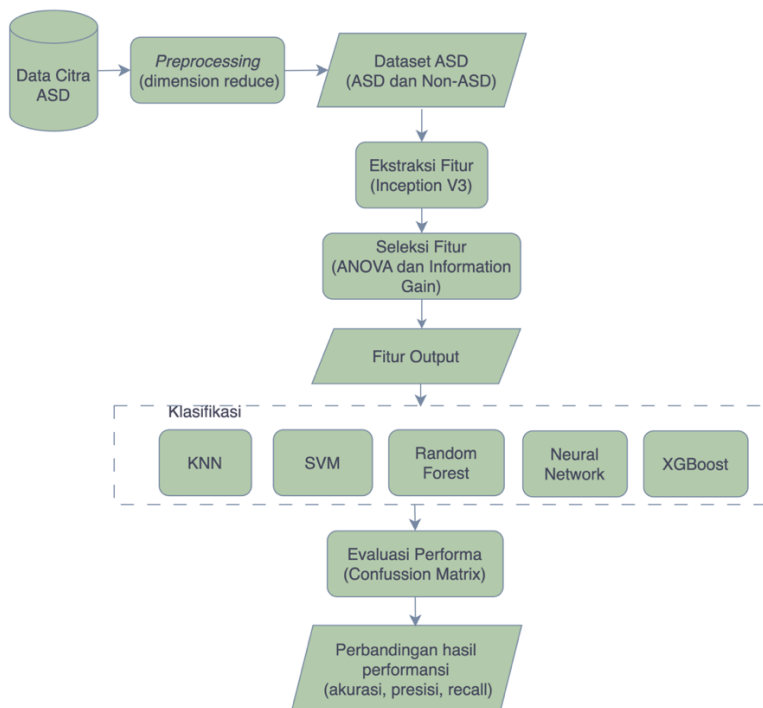
individu ke dalam kategori autistik atau non-autistik, berdasarkan fitur-fitur spesifik yang diekstraksi dari citra wajah[4].

Metode multi-klasifikasi memungkinkan penggunaan beberapa algoritma secara bersamaan untuk meningkatkan akurasi prediksi. Algoritma-algoritma yang digunakan pada penelitian ini adalah *Logistic Regresi*, *Neural Network*, *Random Forest*, *Support Vector Machine (SVM)*. Dengan menggabungkan berbagai algoritma, metode multi-klasifikasi dapat mengatasi keterbatasan yang mungkin dimiliki oleh satu algoritma tunggal.

Selain itu, seleksi fitur adalah langkah penting dalam proses klasifikasi. Seleksi fitur melibatkan pemilihan atribut yang paling relevan dari citra wajah yang dapat memberikan informasi signifikan untuk diagnosis ASD [5]. Fitur-fitur ini bisa berupa karakteristik geometris wajah, tekstur, distribusi warna, dan pola-pola spesifik lainnya yang membedakan individu autistik dari non-autistik [6]. Dengan menggunakan teknik seleksi fitur, model klasifikasi dapat difokuskan pada informasi yang paling penting, sehingga meningkatkan akurasi dan efisiensi deteksi. Beberapa penelitian yang menggabungkan seleksi fitur dan metode multi klasifikasi, antara lain Novianto dkk dengan memanfaatkan seleksi fitur SpSFR dan 6 metode classifier, di mana metode SVM menunjukkan performansi paling baik di antara metode lainnya [7]. Penelitian lainnya dengan menggunakan multi classifier yang dilakukan oleh Shinde dkk, Dimana hasil penelitian menunjukkan akurasi sebesar 97.59% pada metode Linear Regresion, Adaboost dan XGBoost[8]. Beberapa penelitian lainnya deknik menggabungkan teknik transfer learning pada Deep Learning sebagai salah satu metode ekstraksi fitur juga dilakukan oleh [8], [9], [10]. Sehingga penelitian ini bertujuan untuk mengembangkan dan mengevaluasi model klasifikasi berbasis multi-klasifikasi dan metode seleksi serta jumlah fitur pada citra wajah untuk mendeteksi secara dini pada ASD. Dengan memanfaatkan dataset citra wajah yang tersedia secara publik, model ini diharapkan dapat memberikan solusi yang lebih akurat dan efisien untuk mendukung diagnosis klinis ASD pada tahap awal. Melalui pendekatan ini, diharapkan bahwa individu dengan ASD dapat menerima intervensi yang lebih cepat dan tepat, sehingga meningkatkan peluang mereka untuk berkembang dengan lebih baik di masa depan.

2. Metode Penelitian

Penelitian dimulai dari proses akuisisi data, preprocessing data citra, ekstraksi fitur, seleksi fitur, dan proses klasifikasi ASD. Gambar 1 menunjukkan desain alur proses dari penelitian.



Gambar 1. Alur proses klasifikasi ASD berdasarkan pada citra wajah anak-anak.

2.1. Akuisisi dan Preprocessing Citra Wajah Anak-Anak

Penelitian ini menggunakan dataset citra wajah anak-anak yang diambil dari *repository* publik, yaitu pada *Kaggle* (<https://www.kaggle.com/datasets/cihan063/autism-image-data>). Total citra terdiri dari 2940 yang terdiri dari 1470 citra wajah anak dengan label ASD, dan 1470 citra dengan label Non-ASD. Kemudian dataset citra dibagi menjadi data *training* sebanyak 2352, dan data *testing* sebanyak 588 citra. Dalam penelitian ini, langkah *preprocessing* yang dilakukan adalah *reduce dimension*, meliputi proses *resize* dan *rescale* pada citra. Proses *resize* dilakukan untuk mengubah ukuran serta resolusi citra digital, dengan menyesuaikan dimensi citra menjadi 224 x 224 x 3, yang sesuai dengan persyaratan arsitektur Inception-V3. Tujuan dari penggunaan ukuran ini adalah untuk mempercepat proses perhitungan dan menyederhanakan pengolahan data. Sementara itu, *rescale* citra bertujuan untuk menormalkan ukuran piksel. Dalam penelitian ini, *rescale* dilakukan dengan cara membagi nilai piksel terkecil dengan piksel terbesar. Proses *rescale* ini mengubah nilai piksel RGB (dari 0-255) menjadi skala antara 0 dan 1, yang membantu dalam memudahkan proses pelatihan data. Contoh citra wajah anak ditunjukkan pada Gambar 2.



Gambar 2. Contoh citra wajah pada anak ASD dan Non-ASD pada *repository Kaggle*.

2.2. Ekstraksi Fitur

Ekstraksi fitur pada penelitian ini menggunakan konsep atau model *transfer learning* dengan memanfaatkan arsitektur dari *Inception V3*. *Inception V3* adalah arsitektur jaringan saraf konvolusional yang dikembangkan oleh peneliti Google yang secara khusus dirancang untuk meningkatkan efisiensi dan efektivitas komputasi [11]. Arsitektur *Inception V3* terdiri dari beberapa blok *Inception* yang dirancang untuk menangkap fitur dari gambar pada berbagai skala. Struktur *Inception V3* pada penelitian ini ditunjukkan pada Gambar 3.

- a) *Convolutional Layer* : Lapisan konvolusi awal untuk ekstraksi fitur dasar dari input gambar
- b) *Inception Modules* : Kombinasi dari berbagai ukuran filter konvolusi dan pooling untuk menangkap informasi multiskala.
- c) *Reduction Modules* : Blok yang dirancang untuk mengurangi dimensi spasial (*downsampling*) pada fitur.
- d) *Fully Connected Layers* : Lapisan sepenuhnya terhubung di bagian akhir untuk pengklasifikasian.

Model matematis dari *Inception V3* dapat dijelaskan dengan operasi dasar dalam jaringan saraf konvolusional, seperti konvolusi, *pooling*, normalisasi dan aktivasi [11].

1. *Convolution*:

$$\text{output}[i, j, k] = \sum_{m,n} \text{input}[i + m, j + n] \cdot \text{filter}[m, n, k] \quad (1)$$

Operasi konvolusi standar yang menggeser filter di seluruh gambar untuk menghasilkan fitur.

2. *Pooling*:

$$\text{output}[i, j, k] = \max_{m,n} \text{input}[i + m, j + n, k] \quad (2)$$

Pooling mengurangi dimensi spasial dengan mengambil nilai maksimum dalam jendela tertentu.

3. *Batch Normalization*

$$\hat{x}^{(k)} = \frac{x^{(k)} - \mu^{(k)}}{\sqrt{(\sigma^{(k)})^2 + \epsilon}} \quad (3)$$

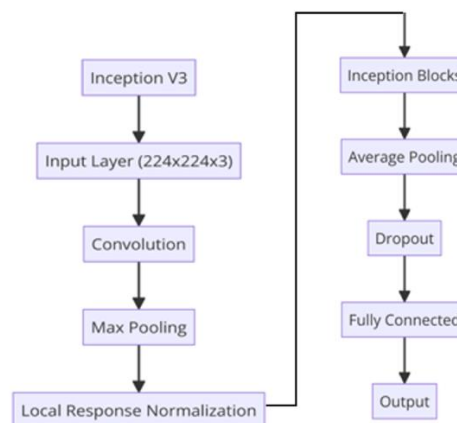
$$y^{(k)} = \gamma^{(k)} \hat{x}^{(k)} + \beta^{(k)}$$

Dimana μ dan σ adalah rata-rata dan variansi *batch*, γ dan β adalah parameter yang dipelajari.

4. *Activation Function* (ReLU, mislanya):

$$\text{output} = \max(0, \text{input}) \quad (4)$$

Fungsi aktivasi non-linear yang menambahkan kemampuan representasi jaringan.



Gambar 3. Model arsitektur *Inception V3*.

2.3. Seleksi Fitur

a. ANOVA

Seleksi fitur menggunakan ANOVA (Analysis of Variance) merupakan teknik statistik yang digunakan untuk menentukan seberapa signifikan perbedaan rata-rata antara kelompok-kelompok yang berbeda. Dalam konteks *machine learning*, khususnya dalam klasifikasi, ANOVA dapat membantu memilih fitur yang paling relevan dengan variabel target kategori.

Prinsip ANOVA:

ANOVA membandingkan variabilitas antara kelompok (variabilitas antar kelompok) dengan variabilitas dalam kelompok (variabilitas intra kelompok). Rasio ini diringkas dalam *F-statistic*. *F-statistic*: Mengukur seberapa banyak variabilitas antar kelompok dibandingkan dengan variabilitas dalam kelompok. Semakin besar *F-statistic*, semakin besar kemungkinan bahwa perbedaan rata-rata antar kelompok bukan karena kebetulan.

b. *Information Gain*

Information Gain merupakan salah satu metode seleksi fitur berbasis filter. Tujuan dari seleksi fitur dengan *Information Gain* adalah untuk mendapatkan fitur yang memiliki informasi yang paling relevan terhadap suatu kelas. *Information Gain* dapat membantu mengurangi *noise* yang disebabkan oleh fitur-fitur yang tidak relevan. *Information Gain* mendeteksi fitur-fitur yang paling banyak memiliki informasi berdasarkan kelas tertentu. Penentuan atribut terbaik dilakukan dengan menghitung nilai *entropy* terlebih dahulu. *Entropy* merupakan ukuran ketidakpastian kelas dengan menggunakan probabilitas kejadian atau atribut tertentu. Untuk mencari nilai dari *Information gain* suatu fitur maka beberapa tahapan persamaan dilakukan, yaitu pada persamaan (5-6).

$$Entropy(S) = \sum_i^c -P_i \log_2 P_i \tag{5}$$

$$Entropy(x) = \sum_i^c -P_i \log_2 P_i \tag{6}$$

Entropy(S) merupakan nilai *entropy* dari *output* kelas, dan *Entropy(X)* merupakan nilai *Entropy* dari suatu fitur, *c* adalah jumlah nilai suatu fitur dan *P_i* merupakan jumlah sampel untuk kelas *i*. Langkah selanjutnya adalah perhitungan nilai *entropy* pada masing-masing fitur terhadap kelas *output* pada persamaan (7) sebagai berikut:

$$E(T, X) = \sum_x P(x) Entropy(x) \tag{7}$$

Dimana *E(T, X)* adalah nilai *entropy* dari suatu fitur *X* terhadap kelas *T*, *P(x)* adalah probabilitas suatu nilai dalam fitur *X* terhadap kelas *T*, dan *Entropy(x)* adalah nilai *entropy* dari suatu nilai di fitur *X* yang didapat dari persamaan (3.32). Setelah nilai *entropy* dari fitur didapat maka untuk mendapatkan nilai *Information Gain* dari fitur tersebut adalah dengan menghitung pengurangan antara *Entropy* kelas target dengan *Entropy* dari fitur seperti pada persamaan (8).

2.4. Multi Classifier

Untuk melakukan analisis pada eksperimen, pada penelitian ini menggunakan beberapa model metode atau algoritma klasifikasi. Dalam pemilihan metode juga telah mempertimbangkan pra-pemrosesan data, over-fitting dan *tuning parameter*. Pra-pemrosesan data mencakup missing value data, transformasi fitur dari kategorikal menjadi data numerik. Model/algoritma klasifikasi yang digunakan adalah sebagai berikut:

a. Random Forest

Dalam menangani klasifikasi biner, metode yang sering digunakan adalah *decision tree*. Namun, *decision tree* memiliki kelemahan utama yaitu *over-fitting*, yang dapat diatasi dengan *Random Forest*. Penggunaan *ensemble of trees* dalam *Random Forest* secara signifikan meningkatkan akurasi klasifikasi dan memungkinkan pemilihan kelas yang paling tepat melalui mekanisme voting. *Random Forest* membentuk sebuah "hutan" dari banyak *decision tree*. *Random Forest* meningkatkan akurasi prediksi dengan melakukan sampling pada dataset. Melalui voting, pohon dengan skor tertinggi dipilih dari hutan tersebut. *Random Forest* juga efektif dalam mengatasi masalah *over-fitting*. Proses kerja *Random Forest* adalah sebagai berikut [12], [13]:

- 1) Pilih secara acak sejumlah *N* sampel data dari dataset pelatihan.
- 2) Bangun *decision tree* menggunakan sampel data yang terpilih.
- 3) Tentukan jumlah *decision tree* yang akan dibentuk.
- 4) Ketika sampel data baru datang, classifier memprediksi keputusan akhir dengan menggunakan suara mayoritas dari *decision tree* yang telah dibentuk.

b. Logistik Regresi

Regresi logistik adalah algoritma prediksi yang digunakan untuk mengelola hubungan antara label kelas biner dan satu atau lebih variabel prediktor. Ini merupakan salah satu metode statistik yang digunakan untuk menyelesaikan masalah prediksi dengan mempertimbangkan probabilitas terjadinya suatu kejadian .

Algoritma ini melakukan klasifikasi label kelas kategorikal menggunakan variabel-variabel independen[14].

c. *Support Vector Machine (SVM)*

SVM umumnya digunakan dalam masalah klasifikasi. Algoritma machine learning ini sangat cepat dan memiliki akurasi yang luar biasa. Dalam SVM, kita mencari hyper-plane di ruang N-dimensi yang mengelompokkan sampel ke dalam kelas biner dengan cara yang optimal. Jarak antara sampel pelatihan terdekat dan hyper-plane disebut margin. SVM menemukan hyper-plane pemisah yang optimal sehingga margin tersebut menjadi maksimum[15]

d. *Artificial Neural Network (ANN)*

ANN merupakan jaringan neural yang terdiri dari banyak neuron yang saling terhubung. Setiap neuron menerima nilai input dan memiliki bobot yang terkait. Jaringan ini terdiri dari tiga lapisan utama, yaitu lapisan input, lapisan tersembunyi, dan lapisan output. Jaringan ini tidak memiliki loop atau siklus [16].

2.5 Evaluasi Performa

Evaluasi performansi dalam klasifikasi machine learning adalah langkah penting untuk menentukan seberapa baik model kita dalam memprediksi kelas yang benar untuk data yang diberikan. Beberapa metrik umum yang digunakan dalam evaluasi performansi model klasifikasi meliputi:

a. Akurasi: Mengukur proporsi prediksi yang benar dari total prediksi

$$Akurasi = \frac{Jumlah\ prediksi\ benar}{Total\ Prediksi} \tag{8}$$

b. Presisi: Mengukur proporsi prediksi positif yang benar dari semua prediksi positif

$$Presisi = \frac{True\ Positive\ (TP)}{True\ Positive\ (TP) + False\ Positive\ (FP)} \tag{9}$$

c. Recall: Mengukur proporsi prediksi positif yang benar dari semua kasus positif sebenarnya.

$$Recall = \frac{True\ Positive\ (TP)}{True\ Positive\ (TP) + False\ Negative\ (FN)} \tag{10}$$

d. AUC (Area Under Curve): mengukur keseluruhan performansi model; semakin dekat nilai AUC ke 1, maka semakin baik model tersebut.

3. Hasil dan Pembahasan

3.1. Penentuan Jumlah Fitur dengan ANOVA dan *Information Gain*

Pada model training, total dari jumlah fitur adalah sebanyak 1000 fitur (n1000). Sehingga dengan memanfaatkan metode seleksi fitur ANOVA dan *Information Gain*, fitur tersebut akan dipilih sesuai dengan skor dan ranking yang diberikan oleh kedua metode seleksi fitur. Tabel 1 merupakan hasil seleksi fitur antara ANOVA dan *Information Gain* dengan menggunakan 25 fitur.

Tabel 1. Contoh Fitur yang telah diseleksi menggunakan ANOVA dan *Information Gain*.

No	ANOVA		<i>Information Gain</i>	
	Fitur	Skor	Fitur	Skor
1	n463	2293.22	n463	0.456
2	n412	2202.63	n412	0.438
3	n899	2107.90	n899	0.424
4	n883	1922.68	n883	0.386
5	n633	1809.44	n633	0.363
6	n898	1806.15	n898	0.358
7	n692	1777.90	n848	0.355

8	n731	1745.41	n692	0.353
9	n427	1731.30	n731	0.352
10	n848	1726.16	n826	0.349
11	n502	1714.68	n427	0.345
12	n826	1710.93	n852	0.341
13	n469	1685.59	n811	0.340
14	n852	1660.88	n469	0.339
15	n473	1660.50	n473	0.338
16	n811	1655.41	n754	0.338
17	n503	1651.60	n503	0.336
18	n615	1632.52	n502	0.332
19	n754	1632.23	n615	0.331
20	n482	1602.69	n643	0.328
21	n747	1577.72	n482	0.325
22	n787	1558.29	n747	0.325
23	n643	1555.85	n855	0.324
24	n805	1554.38	n650	0.321
25	n855	1548.67	n604	0.321

Pada Tabel 1, terlihat perbedaan fitur-fitur yang telah diseleksi oleh metode ANOVA dan *Information Gain*. Metode seleksi fitur pada ANOVA melakukan seleksi fitur berdasarkan nilai Skor yang tertinggi dari keseluruhan fitur, hal ini menunjukkan bahwa fitur tersebut sangat berperan dalam mempelajari pola dari data sehingga dengan memilih/menyeleksi fitur yang relevan dapat meningkatkan kinerja model, sementara fitur yang tidak relevan atau berkualitas rendah dapat menurunkan kinerja model.

3.2. Performansi Hasil Multi Klasifikasi

Performansi hasil klasifikasi terhadap metode fitur seleksi dan jumlah fitur ditunjukkan pada Tabel 2-4.

Tabel 2. Performa dari hasil klasifikasi dengan seleksi fitur ANOVA dan *Information Gain* pada 50 fitur terbaik.

Metode Klasifikasi	ANOVA			<i>Information Gain</i>		
	Akurasi	Presisi	Recall	Akurasi	Presisi	Recall
Logistic Regresi	0.9503	0.9504	0.9505	0.9501	0.9498	0.9502
Neural Network	0.9499	0.9433	0.9400	0.9516	0.9155	0.9600
Random Forest	0.9360	0.9350	0.9381	0.9344	0.9346	0.9335
SVM	0.9378	0.9403	0.9387	0.9323	0.9352	0.9323

Tabel 3. Performa dari hasil klasifikasi dengan seleksi fitur ANOVA dan *Information Gain* pada 100 fitur terbaik.

Metode Klasifikasi	ANOVA			<i>Information Gain</i>		
	Akurasi	Presisi	Recall	Akurasi	Presisi	Recall
Logistic Regresi	0.9617	0.9618	0.9617	0.9593	0.9591	0.9590
Neural Network	0.9627	0.9626	0.9630	0.9624	0.9625	0.9626
Random Forest	0.9438	0.9431	0.9437	0.9434	0.9435	0.9433
SVM	0.9562	0.9570	0.9562	0.9556	0.9562	0.9556

Tabel 4. Performa dari hasil klasifikasi dengan seleksi fitur ANOVA dan *Information Gain* pada 150 fitur terbaik.

Metode Klasifikasi	ANOVA			<i>Information Gain</i>		
	Akurasi	Presisi	Recall	Akurasi	Presisi	Recall
Logistic Regresi	0.9688	0.9687	0.9688	0.9687	0.9686	0.9680
Neural Network	0.9673	0.9676	0.9680	0.9667	0.9668	0.9667
Random Forest	0.9457	0.9458	0.9450	0.9461	0.9462	0.9461

SVM	0.9585	0.9588	0.9580	0.9573	0.9572	0.9570
-----	--------	--------	--------	--------	--------	--------

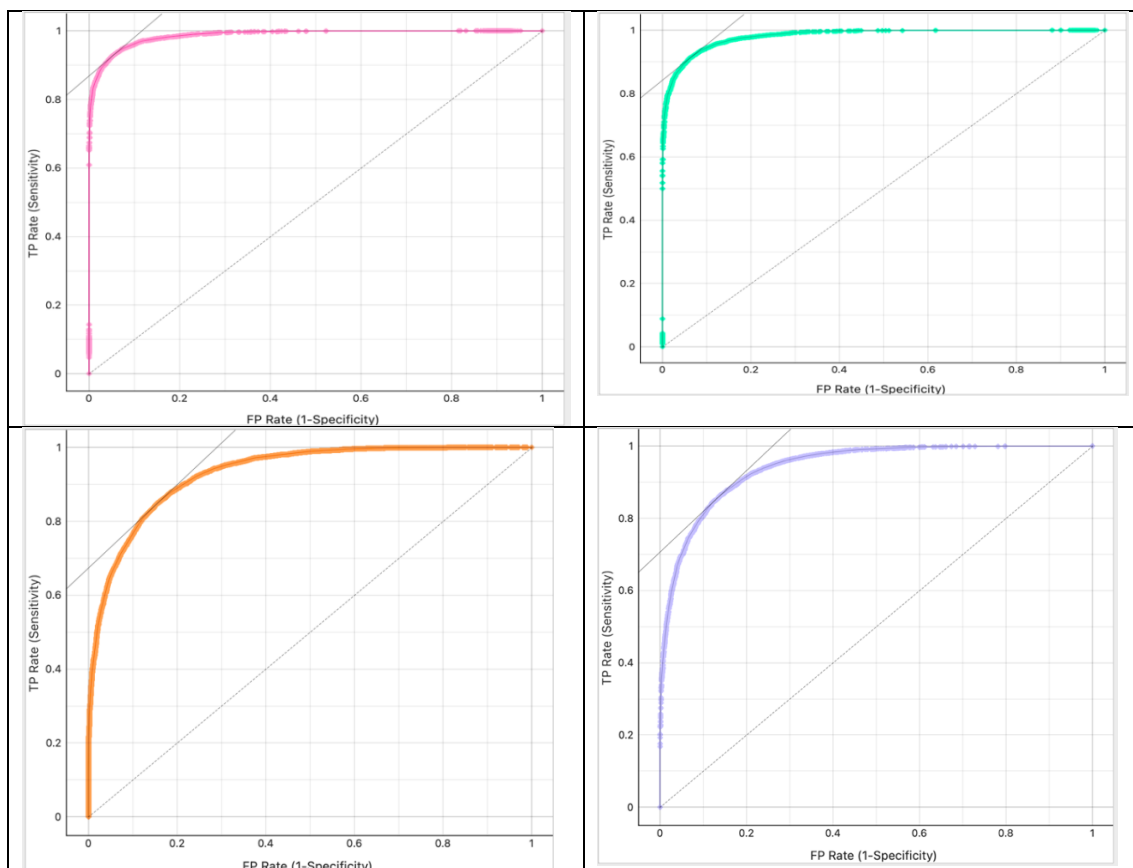
Tabel 2- 4 menunjukkan perpaduan metode klasifikasi *logistic regresi* dengan seleksi fitur ANOVA pada 150 fitur terbaik menghasilkan performansi terbaik dari sisi akurasi sebesar 0.9688, presisi sebesar 0.9687, dan recall sebesar 0.9688, dibandingkan dengan penggunaan metode seleksi fitur *Information Gain* dan jumlah fitur. Hasil pengujian menunjukkan metode *logistic linear* memiliki keunggulan dalam melakukan klasifikasi pada kelas biner dengan fitur yang terbatas. Hasil analisis terhadap metode seleksi fitur menunjukkan ANOVA memiliki performansi yang lebih baik dari *Information Gain* dalam memilih fitur yang paling relevan dengan variabel target kategori.

Pada tabel 5, menunjukkan hasil pengujian terhadap performansi dari sisi Area Under Curve (AUC) pada performa klasifikasi dengan seleksi fitur ANOVA dengan 150 fitur terbaik.

Tabel 5. Performansi dari AUC pada metode klasifikasi dan seleksi fitur Anova pada 150 fitur.

Metode Klasifikasi	AUC
Logistic Regresi	0.9962
Neural Network	0.9962
Random Forest	0.9905
SVM	0.9948

Tabel 5 menunjukkan performansi dari metode klasifikasi *logistic regresi* dan *neural network* menunjukkan hasil AUC sebesar 0.9962, hal ini menunjukkan bahwa model yang dibangun untuk melakukan deteksi dini autisme pada anak berdasarkan citra wajah menghasilkan performansi model yang sangat baik dengan hasil hampir mendekati nilai 1. Gambar 4 menunjukkan kurva dari AUC pada masing-masing metode klasifikasi.



Gambar 4. AUC dari masing-masing metode klasifikasi (pink: *Logistic Regresi*; hijau: *Neural Network*; orange: *Random Forest*; ungu: *SVM*).

4. Kesimpulan

Dengan memanfaatkan dataset citra wajah yang tersedia secara publik, perpaduan metode klasifikasi *logistic regresi* dengan seleksi fitur ANOVA pada 150 fitur terbaik menghasilkan performansi terbaik dari sisi akurasi sebesar 0.9688, presisi sebesar 0.9687, dan recall sebesar 0.9688, dibandingkan dengan penggunaan metode seleksi fitur *Information Gain* dan jumlah fitur. Hasil pengujian menunjukkan metode *logistic linear* memiliki keunggulan dalam melakukan klasifikasi pada kelas biner dengan fitur yang terbatas. Hasil analisis terhadap metode seleksi fitur menunjukkan ANOVA memiliki performansi yang lebih baik dari *Information Gain* dalam memilih fitur yang paling relevan dengan variabel target kategori. Nilai AUC sebesar 0.9962 untuk metode Linear Regression dan Neural Network mengindikasikan bahwa model ini menghasilkan performansi sangat baik dengan hasil hampir mendekati nilai 1. Melalui pendekatan ini, diharapkan bahwa individu dengan ASD dapat menerima intervensi yang lebih cepat dan tepat, sehingga meningkatkan peluang mereka untuk berkembang dengan lebih baik di masa depan. Pada penelitian berikutnya dapat dikembangkan kembali tentang cara ekstraksi fitur baik secara handcrafted ataupun dengan pendekatan convolutional neural network (CNN).

Daftar Pustaka

- [1] P. Mazumdar, G. Arru, and F. Battisti, "Early detection of children with Autism Spectrum Disorder based on visual exploration of images," *Signal Process Image Commun*, vol. 94, no. February, p. 116184, 2021, doi: 10.1016/j.image.2021.116184.
- [2] M. Z. Uddin, M. A. Shahriar, M. N. Mahamood, F. Alnajjar, M. I. Pramanik, and M. A. R. Ahad, "Deep learning with image-based autism spectrum disorder analysis: A systematic review," *Eng Appl Artif Intell*, vol. 127, no. PA, p. 107185, 2024, doi: 10.1016/j.engappai.2023.107185.
- [3] A. V. Shinde and D. D. Patil, "A Multi-Classifer-Based Recommender System for Early Autism Spectrum Disorder Detection using Machine Learning," *Healthcare Analytics*, vol. 4, no. April, p. 100211, 2023, doi: 10.1016/j.health.2023.100211.
- [4] M. S. Farooq, R. Tehseen, M. Sabir, and Z. Atal, "Detection of autism spectrum disorder (ASD) in children and adults using machine learning," *Sci Rep*, vol. 13, no. 1, Dec. 2023, doi: 10.1038/s41598-023-35910-1.
- [5] P. W. Rudnicki W.R., Wrzesień M., *Feature selection for data and pattern classification*. 2013.
- [6] A. A. Z. I. Delowar Hossain Muhammad Ashad Kabir, "Detecting autism spectrum disorder using machine learning techniques: An experimental analysis on toddler, child, adolescent and adult datasets.," vol. 9, no. 1. Springer. doi: 10.1007/S13755-021-00145-9.
- [7] A. Novianto and M. D. Anasanti, "Autism Spectrum Disorder (ASD) Identification Using Feature-Based Machine Learning Classification Model," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 17, no. 3, p. 259, Jul. 2023, doi: 10.22146/ijccs.83585.
- [8] A. V. Shinde and D. D. Patil, "A Multi-Classifer-Based Recommender System for Early Autism Spectrum Disorder Detection using Machine Learning," *Healthcare Analytics*, vol. 4, Dec. 2023, doi: 10.1016/j.health.2023.100211.
- [9] I. Ahmad, J. Rashid, M. Faheem, A. Akram, N. A. Khan, and R. ul Amin, "Autism spectrum disorder detection using facial images: A performance comparison of pretrained convolutional neural networks," *Healthc Technol Lett*, no. January, 2024, doi: 10.1049/htl2.12073.
- [10] T. M. Ghazal, S. Munir, S. Abbas, A. Athar, H. Alrababah, and M. A. Khan, "Early Detection of Autism in Children Using Transfer Learning," *Intelligent Automation and Soft Computing*, vol. 36, no. 1, pp. 11–22, 2023, doi: 10.32604/iasc.2023.030125.
- [11] A. S. Ali Abdullah Yaser Issam Aljanabi, "Developing a convolutional neural network for classifying tumor images using Inception v3," vol. 3, no. 9 (123). doi: 10.15587/1729-4061.2023.281227.
- [12] M. Martin, T. Nguyen, S. Yousefi, and B. Li, "Comprehensive features with randomized decision forests for hand segmentation from color images in uncontrolled indoor scenarios," *Multimed Tools Appl*, vol. 78, no. 15, pp. 20987–21020, 2019, doi: 10.1007/s11042-019-7445-3.
- [13] P. Annangi, S. Frigstad, S. B. Subin, A. Torp, S. Ramasubramaniam, and S. Varna, "An automated bladder volume measurement algorithm by pixel classification using random forests," in *38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2016*, GE Global Research, United States: Institute of Electrical and Electronics Engineers Inc., 2016, pp. 4121–4124. doi: 10.1109/EMBC.2016.7591633.
- [14] K. Vijayalakshmi, M. Vinayakamurthy, and Anuradha, "A Hybrid Recommender System using MultiClassifier Regression Model for Autism Detection," in *2020 International Conference on*

-
- Smart Technologies in Computing, Electrical and Electronics (ICSTCEE)*, 2020, pp. 139–144. doi: 10.1109/ICSTCEE49637.2020.9277034.
- [15] T. Akter, M. I. Khan, M. H. Ali, M. S. Satu, M. J. Uddin, and M. A. Moni, “Improved Machine Learning based Classification Model for Early Autism Detection,” *International Conference on Robotics, Electrical and Signal Processing Techniques*, pp. 742–747, 2021, doi: 10.1109/ICREST51555.2021.9331013.
- [16] H. Ravishankar *et al.*, “Understanding the Mechanisms of Deep Transfer Learning for Medical Images,” pp. 68–76, 2017, doi: 10.1007/978-3-319-46976-8.