JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.9874
Hal 1269-1279
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Klasifikasi Tingkat Kemampuan Bahasa Anak Berdasarkan Sinyal EEG Menggunakan Random Forest dengan Ektraksi Fitur Wavelet

Galuh Saputri[1,*], Anton Yudhana[2], Abdul Fadlil[2]

[1] Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Pamulang, Tangerang Selatan, Indonesia
[2]Fakultas Teknologi Industri, Program Studi Informatika, Universitas Ahmad Dahlan, Yogyakarta, Indonesia
Email: [1,*]dosen02693@unpam.ac.id, [2]eyudhana@ee.uad.ac.id,[3]fadlil@mti.uad.ac.id
Email Penulis Korespondensi: eyudhana@ee.uad.ac.id[*]
Submitted **05-06-2026**; Accepted **02-10-2026**; Published **31-10-2026**

Abstrak

Salah satu indikator penting dalam perkembangan kognitif anak terhadap kemampuan komunikasi, proses pembelajaran, dan literasi adalah kemampuan bahasa, tetapi pengukuran kemampuan bahasa masih sering dilakukan dengan pendekatan konvensional seperti tes akademik dan observasi yang mengarah ke subjektif dan belum bias merepresentasikan aktivitas otak secara langsung. Urgensi penelitian ini mengatasi keterbatasan subjektif konvensional menjadi pendekatan klasifikasi kategori kemampuan bahasa anak menggunakan sinyal *electroencephalography* (EEG). Tujuan penelitian ini adalah mengembangkan pendekatan klasifikasi kategori kemampuan bahasa anak berbasis sinyal EEG sebagai indikator objektif menggunakan algoritma *Random Forest* dan ekstraksi fitur *wavelet*. Data EEG yang digunakan adalah Kumar EEG Imagined Speech Dataset untuk mendukung pengujian karakteristik sinyal EEG yang berkaitan dengan aktivitas bahasa dan pengembangan tahapan pengolahan serta ekstraksi fitur. Sedangkan kategori kemampuan bahasa anak dioperasionalkan berdasarkan nilai akademik bahasa anak yang diperoleh dari tiga sekolah dasar di Yogyakarta. Tahapan penelitian ini, yaitu pra-pemrosesan data, ekstraksi fitur dengan *Discrete Wavelet Transform* (DWT), klasifikasi menggunakan *algoritma Random Forest*, dan evaluasi model. Hasil penelitian ini menunjukkan performa klasifikasi yang memuaskan dengan akurasi 0.86, *precision* 0.89, *recall* 0.83, dan *f1-score* 0.83, Hasil penelitian ini menunjukkan adanya potensi karakteristik fitur EEG yang dapat dimanfaatkan untuk membedakan kategori kemampuan bahasa yang ditetapkan berdasarkan data nilai akademik. Penelitian ini memberikan kontribusi awal dalam menghubungkan representasi neurofisiologis sinyal EEG dengan kategori kemampuan bahasa anak melalui pendekatan machine learning.

Kata Kunci: Klasifikasi; Kemampuan Bahasa; EEG; *Wavelet; Random Forest*

Abstract

Language ability is a key indicator of children's cognitive development influencing communication skills, learning processes, and literacy yet it is frequently measured using conventional approaches such as academic tests and observations, which are prone to subjectivity and fail to directly represent brain activity. This research addresses the limitations of conventional, subjective methods by proposing an approach to classify children's language ability categories using electroencephalography (EEG) signals. The study aims to develop an EEG-based classification approach for children's language ability categories serving as an objective indicator utilizing the Random Forest algorithm and wavelet feature extraction. The Kumar EEG Imagined Speech Dataset was employed to analyze EEG signal characteristics associated with language activity and to develop the signal processing and feature extraction stages. Meanwhile, children's language ability categories were operationalized based on academic language grades obtained from three elementary schools in Yogyakarta. The research stages included data preprocessing, feature extraction using Discrete Wavelet Transform (DWT), classification via the Random Forest algorithm, and model evaluation. The results demonstrate satisfactory classification performance, achieving an accuracy of 0.86, precision of 0.89, recall of 0.83, and an F1-score of 0.83. These findings indicate the potential of specific EEG feature characteristics to distinguish between language ability categories established based on academic grade data. This study makes an initial contribution by linking neurophysiological representations of EEG signals to categories of children's language abilities using a machine learning approach.

Keywords: Classification; Language Ability; EEG; Wavelet; Random Forest

1. PENDAHULUAN

Kecerdasan dapat didefinisikan sebagai kemampuan belajar dari ilmu dan pengalaman untuk beradaptasi dengan lingkungan [1],salah satu bentuk kecerdasan yang dimiliki manusia adalah kecerdasan linguistik yang melandasi kemampuan seseorang dalam berbahasa.. Kemampuan bahasa dipengaruhi oleh perkembangan mental dan kognitif, semakin kompleks pemahaman bahasa yang dikuasai maka semakin matang mental dan kognitif seseorang [2]. Seseorang yang memiliki tingkat kemampuan bahasa yang baik dapat memahami konteks komunikasi, menyampaikan ide, dan cepat beradaptasi di berbagai lingkungan sosial [3].

Kemampuan literasi membaca anak Indoensia sebesar 359, berada di bawah rata-rata negara OECD berdasarkan laporan *Programme for International Student Assesment* (PISA) 2022 [4], [5]. Kondisi tersebut menunjukkan kemampuan literasi dan bahasa anak perlu diperhatikan sejak usia dini, dengan melakukan stimulus pembelajaran yang sesuai potensi kognitif dan karakterstik anak ,bisa menjadi faktor utama untuk meningkatkan kemampuan literasi dan bahasa. Proses perkembangan kemampuan bahasa dapat lebih optimal dengan menegtahui kemampuan anak melalui pendekatan objektif, orangtua atau guru dapat memberikan stimulus dan metode belajar yang sesuai karakteristik kognitif anak [6].

Saat ini, pengukuran kemampuan bahasa anak dilakukan melalui pendekatan yang konvensional, seperti observasi, wawancara, dan tes akademik, tetapi metode tersebut terkadang bersifat subjektif dan bergantung pada interpretasi evaluator [7]. Pendekatan tersebut belum mampu merepresentasikan aktivitas kognitif secara langsung dari otak manusia, sehingga evaluasi objektif dari proses neurokognitif yang menjadi dasar perkembangan bahasa belum tercapai maksimal.

Copyright © 2026 The **Galuh Saputri**, Page 1269
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.9874
Hal 1269-1268
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Dengan memanfaatkan *neuroscience*, perkembangan bahasa secara objektif melalui deteksi *brainwave*[8]. *Electroencephalography* (EEG) merupakan salah satu teknologi yang dapat digunakan untuk mengetahui tingkat perkembangan kemampuan bahasa anak [9]. EEG dapat merekam aktivitas listrik otak anak dengan aman secara langsung dan banyak di pakai dalam berbagai bidang seperti analisis fungsi kognitif, diagnosis gangguan neurologis, dan neurolinguistik [10], [11].

Urgensi penelitian ini ada pada kebutuhan untuk mengembangkan pendekatan berbasis neurofisiologis yang dapat memberikan informasi tambahan mengenai karakteristik aktivitas otak yang berkaitan dengan kemampuan bahasa anak. Penilaian kemampuan bahasa biasanya menggunakan nilai akademik, tes, dan observasi memberikan informasi penting mengenai capaian kemampuan anak, tetapi belum secara langsung merepresentasikan karakteristik aktivitas neurofisiologis selama proses bahasa berlangsung. Pemanfaatan sinyal electroencephalography (EEG) untuk mengeksplorasi karakteristik aktivitas otak yang berkaitan dengan proses bahasa dan menghubungkannya dengan kategori kemampuan bahasa yang ditetapkan dari data nilai akademik bahasa. Skor akademik terbagi menjadi 3 kategori, yaitu Low dengan skor <60, Medium dengan skor 60–79, High dengan skor ≥80. Model machine learning pada penelitian ini mempelajari pola karakteristik EEG yang berkaitan dengan kategori kemampuan bahasa berdasarkan data nilai akademik bahasa. Dataset Kumar digunakan untuk menyediakan dan menguji karakteristik sinyal EEG pada aktivitas imagined speech yang relevan dengan pemrosesan bahasa, khususnya dalam pengembangan dan validasi tahapan pengolahan sinyal serta ekstraksi fitur. Ekstraksi fitur wavelet dipilih karena unggul dalam menganalisis sinyal pada domain waktu dan frekuensi secara bersamaan, menangkap fitur-fitur lokal, dan transien yang krusial dalam sinyal EEG [12]. Tujuan penelitian ini adalah mengembangkan pendekatan klasifikasi kategori kemampuan bahasa anak berbasis sinyal EEG sebagai indikator objektif menggunakan algoritma Random Forest dan ekstraksi fitur wavelet.

Pada penelitian terdahulu, EEG secara efektif mendeteksi respons fisiologis dan berpotensi untuk monitoring kemampuan bahasa melalui pola aktivitas otak [13]. Pada penelitian yang dilakukan oleh Elwasify dkk, melakukan pengembangan metode klasifikasi imagined speech berbasis sinyal EEG menggunakan preprocessing FD-F (ICA + *frequency-domain filtering*) dan *model deep learning* CNN-2-Bi-LSTM pada *dataset* Kumar 30 kelas, model ini berhasil mencapai akurasi di atas 99% [14][15]. Penelitian lainnya yang dilakukan oleh Aldhaheri dkk melakukan penelitian dari respons otak saat mempelajari bahasa kedua menggunakan sinyal EEG dan *machine learning*. Data dikumpulkan dari 8 partisipan menggunkan perangkat Virgo EEG, lalu dianalisis melalui *preprocessing* sinyal dan klasifikasi dengan beberapa algoritma *Random Forest*, SVM, XGBoost, dan CatBoost menggunakan EEGLAB serta MNE-Python. Hasil terbaik diperoleh dengan *model CatBoost* yang mencapai akurasi 74.56%, menunjukkan bahwa EEG dan machine learning efektif untuk menganalisis aktivitas otak dalam proses pembelajaran bahasa kedua [10], fokus penelitian ini untuk respons pembelajaran bahasa kedua menggunakan jumlah partisipan yang terbatas, hal tersebut belum menjawab permasalahan identifikasi tingkat kemampuan bahasa anak berdasarkan representasi fitur EEG.

Research gap pada penelitian terdahulu, yaitu: (1) Fokus penelitian EEG terkait bahasa pada pengenalan imagined speech, analisis respons pembelajaran bahasa atau klasifikasi stimulus bahasa, sedangkan pemanfaatan EEG untuk klasifikasikan tingkat kemampuan bahasa anak terbatas [10], [14]; (2) Penelitian sebelumnya sudah menunjukkan EEG mengandung informasi relevan untuk aktivitas bahasa, namun secara spesifik belum mengeksplorasi fitur wavelet untuk representasi karakteristik temporal frekuensi yang membedakan tingkat kmampuan bahasa [12], [13]; (3) Pemanfaatan fitur EEG perlu pendekatan machine learning dan hasil transformasi wavelet untuk menghasilkan klasifikasi tingkat kemampuan bahasa secara efektif. Penelitian ini menjawab kesenjangan tersebut degan mengusulkan pendekatan klasifikasi tingkat kemampuan bahasa anak berbasis sinyal EEG menggunakan ekstraksi fitur wavelet dan algoritma Random Forest. Penggunaan fitur wavelet untuk mendapat representasi karakteristik sinyal pada domain waktu-frekuensi, penggunaan Random Forest untuk mempelajari pola hubungan antara sinyal EEG dan tingkat kemampuan bahasa. Random Forest dipilih berdasarkan pada kemampuan dalam menangani data berdimensi tinggi, nonlinear,dan kombinasi fitur yang kompleks [16].

Novelty penelitian ini, yaitu penggunaan ekstraksi fitur *wavelet* dan algoritma *Random Forest* untuk mengklasifikasikan tingkat kemampuan bahasa anak yang dapat membantu pengembangan alat pengukuran objektif lebih lanjut. Berbeda dari penelitian terdahulu yang menggunakan EEG untuk mengenali stimulus atau aktivitas bahasa, penelitian ini mengarahkan karakteristik EEG untuk mengklasifikasikan kategori kemampuan bahasa yang ditetapkan berdasarkan data nilai akademik bahasa. Kontribusi penelitian terletak pada representasi neurofisiologis EEG dan indikator kemampuan bahasa yang digunakan sebagai dasar pelabelan.

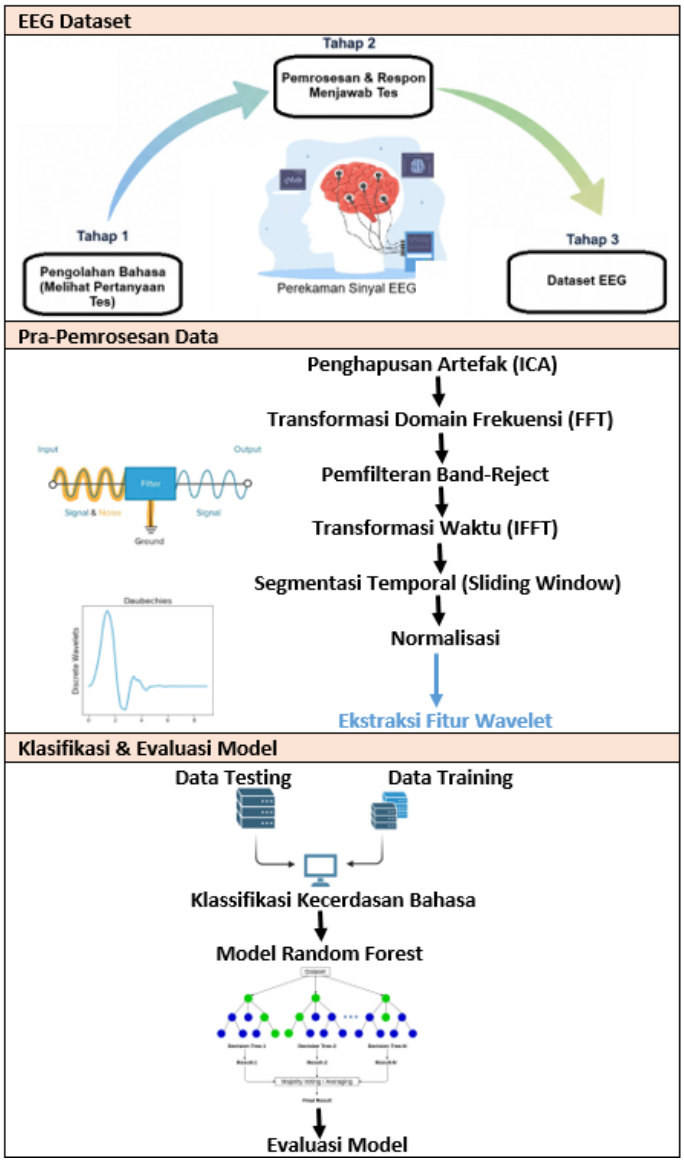
2. METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan eksperimental dengan metode pengolahan sinyal dan machine learning berdasarkan sinyal EEG untuk tingkat kemampuan bahasa anak. Tahapan penelitian dimulai dari pengumpulan data EEG, pra-pemrosesan data, ekstraksi fitur menggunakan *discrete wavelet transform* (DWT), penyeimbang data dengan *Oversampling*, klsifikasi menggunakan algoritma Random Forest, dan evaluasi performa model. Data nilai akademik bahasa digunakan untuk label klasifikasi, yaitu High, Medium, dan Low, kategori High diberikan kepada subjek dengan skor ≥80, Medium diberikan kepada subjek dengan skor 60–79, Low diberikan kepada subjek dengan skor <60. Label tersebut merepresentasikan kategori kemampuan bahasa berdasarkan skor akademik yang digunakan dalam penelitian, bukan pengukuran psikometrik kemampuan secara klinis. Tingkat kemampuaan bahasa dalam penelitian ini digunakan

Copyright © 2026 The **Galuh Saputri**, Page 1270
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.9874
Hal 1269-1268
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

dalam konteks operasional berdasarkan indikator skor akademik bahasa. Untuk mempermudah pemahaman alur penelitian, tahapan Eksperimen dan Analisis Pengujian ditunjukkan pada gambar 1.



Gambar 1. Eksperimen dan Analisis Pengujian

2.1 EEG Dataset

Dataset EEG yang digunakan, yaitu *dataset Kumar EEG Imagined Speech* dari *web Kaggle* yang banyak digunakan dalam penelitian analisis aktivitas kognitif dan *Brain Computer Interface* (BCI) [10], [14]. Dataset terdiri dari 3 jenis stimulus, yaitu *Digit, Char, Image* yang merepresentasikan aktivitas imajinasi *linguistic* dan *visual*. Prosedur perekaman melibatkan 23 peserta, peserta harus membayangkan 10 huruf alfabet Inggris, angka 0-9, dan gambar objek tertentu dalam kondisi fokus. Tujuan dari latihan visualisasi ini adalah untuk memicu aktivitas mental tertentu yang terkait dengan pemrosesan bahasa dan kognitif. Aktivitas EEG yang terkait dengan tugas inimenggunakan perangkat Emotiv EPOC+, yang mampu merekam 14 saluran, masing-masing dengan frekuensi sampling 128 Hz, selama 10 detik, yang setara dengan 1280 sampel untuk setiap saluran. Dataset Kumar dipilih sebagai dataset EEG dalam penelitian ini karena aktivitas imagined speech yang direkam memiliki keterkaitan dengan pemrosesan bahasa di otak dan digunakan untuk memvalidasi pendekatan ekstraksi fitur Wavelet serta klasifikasi pola EEG. Kemudian algoritma Random Forest digunakan untuk mempelajari pola fitur tersebut.

Dataset kedua menggunakan data nilai akademik bahasa dengan subjek penelitian berupa siswa dari SD Muhammadiyah Argosari, SD Muhammadiyah Karangkajen, dan SDN Keputran I Yogyakarta, Prosedur akuisisi data melibatkan pengumpulan data nilai akademik melalui dokumen resmi sekolah pada mata pelajaran bahasa. Proses pengumpulan data dilakukan dengan memperoleh izin dari pihak sekolah dan persetujuan orangtua/wali serta memastikan kerahasiaan data subjek. Data nilai akademik yang terkumpul akan digunakan sebagai label untuk membandingkan respons perkembangan kognitif yang terukur dari sinyal EEG dan menilai seberapa handal ektraksi fitur dan model yang diusulkan.

Copyright © 2026 The **Galuh Saputri**, Page 1271
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.9874
Hal 1269-1268
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

2.2 Pra-Pemrosesan Data

2.2.1 Tahap Pra-Pemrosesan

Tahap pra-pemrosesan diperlukan untuk ekstraksi fitur karena sinyal EEG mentah memiliki karakteristik yang bersifat non-linear dan non-stasioner. Karakteristik tersebut menyulitkan ekstraksi fitur yang berguna dari sinyal mentah, karena menghambat kualitas sinyal dan menyebabkan adanya artefak biologis. Tanpa langkah pra-pemprosesan ini, data menjadi lebih sulit untuk proses klasifikasi *machine learning*.

Pada pra-pemrosesan data terdapat 6 tahap yang dilakukan, yaitu:

a. Tahap 1 penghapusan artefak dengan *Independent Component Analysis* (ICA) yang bertujuan untuk menghilangkan gangguan non-neurologis seperti pergerakan mata (EOG) dan aktivitas otot (EMG) dari sinyal EEG mentah, hal ini meningkatkan kualitas sinyal yang akan digunakan untuk analisis [17].
b. Tahap 2 transformasi domain frekuensi dengan Fast Fourier Transform (FFT) yang bertujuan mengkonversi sinyal dari domain waktu ke domain frekuensi untuk memudahkan proses pemfilteran berikutnya [18], [19].
c. Tahap 3 pemfilteran band-reject untuk menghilangkan frekuensi gangguan spesifik 4-15 Hz yang sering mengandung artefak dan tidak relevan untuk aktivitas kognitif yang akan di analisis. Rumus proses *filtering*:

$$y[n] = \sum_{k=0}^{N-1} h[k] \, . \, x[n-k] \qquad (1)$$

dimana y[n] adaalah sinyal yang telah di *filter*, h[k] adalah respons *impuls filter*, dan x[n-k] adalah sinyal *input*.

d. Tahap 4 transformasi waktu dengan *Inverse Fast Fourier Transform* (IFFT) untuk mengembalikan sinyal yang sudah di filter kembali ke *domain* waktu aslinya [20].
e. Tahap 5 segmentasi *temporal* (*Sliding Window*) bertujuan memecah sinyal EEG yang panjang menjadi potongan *window* yang lebih kecil untuk digunakan sebagai sample pelatihan model [21].
f. Tahap 6 adalah normalisasi adaptif (z-score) melakukan standar data agar model lebih stabil dan menghindari *leakage* data saat pelatihan [22]. Rumus normalisasi (z-score):

$$x_{norm}[n] = \frac{x[n] - \mu}{\sigma} \qquad (2)$$

dimana $x_{norm}[n]$ adalah nilai sinyal hasil normalisasi dalam data ke-n, $x[n]$ adalah nilai sinyal asli dalam data ke-n, μ adalah mean seluruh data sinyal, σ adalah standar deviasi data sinyal.

2.2.2 Ekstraksi Fitur *Wavelet*

Teknik ekstraksi fitur menggunakan *Discrete Wavelet Transform* (DWT) dengan fungsi *wavelet Daubechies-4* (db4). DWT mengubah sinyal EEG dari *domain* waktu ke *domain* waktu frekuensi, memungkinkan analisis pada brbagai skala frekuensi [23]. Tujuan dari langkah ini yaitu mengubah sinyal EEG yang telah dibersihkan dan tersegmentasi menjadi sekumpulan fitur numerik yang dapat dipahami oleh algoritma *Random Forest*. DWT memungkinkan analisis sinyal EEG pada resolusi waktu dan frekuensi yang sangat penting untuk menangkap karakteristik sinyal otak yang kompleks. Dari hasil dekomposisi DWT, didapatkan 2 jenis koefisien utama:

a. Koefisien Aproksimasi (cA) mewakili komponen frekuensi rendah dari sinyal yang memberikan informasi tentang tren umum dalam aktivitas otak [24].

$$cA = \sum_n x[n] . \phi[n] \qquad (3)$$

dimana *cA* adalah koefisien *approximation*, $x[n]$ adalah sinyal EEG diskrit, $\phi[n]$ adalah fungsi skala, dan n adalah *indeks* data sinyal.

b. Koefisien detail (cD) mewakili komponen frekuensi tinggi dari sinyal yang menangkap fluktuasi dan perubahan cepat dalam aktivitas otak [25].

$$cD = \sum_n x[n] . \psi[n] \qquad (4)$$

dimana *cD* adalah koefisien detail, $x[n]$ adalah sinyal EEG diskrit, $\psi[n]$ adalah fungsi *wavelet*, dan n adalah *indeks* data sinyal.

2.2.3 Ketidakseimbangan Kelas (Oversampling)

Oversampling terhadap data pelatihan dilakukan untuk mengurangi dominasi kelas mayoritas. Oversampling tidak dilakukan terhadap keseluruhan dataset sebelum proses pembagian data, karena prosedur tersebut dapat menyebabkan sampel hasil duplikasi atau informasi yang berkaitan dengan data pengujian masuk ke proses pelatihan dan menghasilkan data leakage. Urutan pada setiap iterasi validasi, yaitu pertama menentukan subjek pengujian, kedua membentuk data pelatihan dari subjek lainnya, ketiga melakukan oversampling hanya terhadap data pelatihan, keempat melakukan

Copyright © 2026 The **Galuh Saputri**, Page 1272
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.9874
Hal 1269-1268
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

standardisasi berdasarkan data pelatihan, kelima melatih model Random Forest, keenam mengevaluasi model menggunakan data pengujian asli tanpa oversampling [26].

2.3 Klasifikasi dan Evaluasi Model

2.3.1 Klasifikasi dengan Algoritma *Random Forest*

Random Forest meru[akan salah satu algoritma *machine learning*, menggabungkan banyak pohon keputusan (*decision trees*) untuk menghasilkan prediksi yang lebih akurat dan stabil. Algoritma ini digunakan untuk masalah klasifikasi maupun regresi dengan melatih setiap pohon pada subset data acak [27]. Setiap *node decision tree* saat mengukur tingkat ketidakmurnian data Rumus *Random Forest* dimulai dengan perhitungan *Gini Impurity*, persamaannya:

$$Gini = 1 - \sum_{i=1}^{C} p_i^2 \quad (5)$$

dimana, pi mewakili proporsi sampel kelas I dan C untuk jumlah kelas pada dataset. Untuk menentukan *split* terbaik pada pohon keputusan, digunakan rumus *Gain* atau *Decision Tree Splitting* yang mengukur penurunan impurity, persamaannya:

$$Gain = Gini_{parent} - \sum_{j=1}^{m} \frac{n_j}{n} . Gini_{child_j} \quad (6)$$

dimana $Gini_{parent}$ sebagai nilai *impurity node* induk, $Gini_{child_j}$ sebagai nilai *impurity node* anak ke-j, n_j sebagai jumlah sampel node anak ke-j, n sebagai jumlah total sampel node induk, dan m sebagai jumlah node anak hasil proses split. Persamaan ini untuk mengukur proses pemisahan data setiap node. Semakin besar nilai Gain, semakin baik split tersebut dalam memisahkan kelas-kelas. Setelah semua pohon dalam *forest* dilatih, proses prediksi dilakukan dengan mengambil *Random Forest Prediction*, yaitu modus semua prediksi individual dari setiap pohon [28].

Random Forest mempunyai beberapa *parameter* untuk mengatur proses pembuatan *decision tree* supaya mendapatkan performa klasifikasi optimal [29]. *Parameter* klasifikasi *model Random Forest*, yaitu:

a. *n_estimators = 100*, menyatakan jumlah pohon keputusan yang dipakai untuk proses klasifikasi
b. *criterion = 'gini'*, menentukan *split* terbaik setiap *node* dengan metode pengukuran *impurity* dari persamaan (5) dan (6)
c. *max_depth = None*, menentukan kedalaman maksimal pohon hingga semua node memenuhi kondisi penghentian
d. *min_samples_split = 2*, melakukan pemecahan node dalam menentukan jumlah minimum sampel
e. *min_samples_leaf = 1*, menyatakan jumlah minimum sampel pada node akhiir
f. *max_features = 'auto'*, meningkatkan variasi antar pohon dalam forest untuk memilih *subset* fitur otomatis dalam setiap pembentukan pohon
g. *class_weight = 'balanced'*, menangani imbalanced jumlah data antar kelas supaya model tidak bias terhadap kelas mayoritas
8. *random_state = 40*, mengatur nilai acak supaya hasil pelatihan model bias reproduksi secara konsisten
9. *n_jobs = -1*, mengatur pemakaian jumlah inti prosesor agar pelatihan model berjalan cepat.

2.3.2 Evaluasi Model

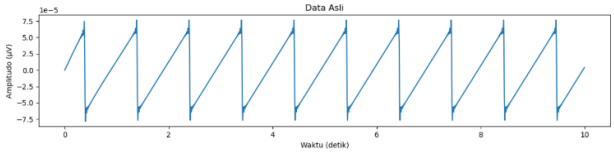
Evaluasi performa model dilakukan menggunakan *cross-validation* dengan *3-fold stratified*, 2 *fold* untuk data training dan satu fold untuk data testing yang berfungsi untuk validitas model. Metrik evaluasi utama meliputi *accuracy*, *precision*, *recall*, *f1-score*, dan *confusion matrix* untuk analisis detail performa per kelas [28][30].

3. HASIL DAN PEMBAHASAN

3.1 Pra-Pemrosesan Data

3.1.1 Tahap Pra-Pemrosesan

a. Tahap 1 penghapusan artefak dengan analisis komponen independen (ICA) menggunakan library MNE untuk memuat data EEG.



Gambar 2. Data Asli

Copyright © 2026 The **Galuh Saputri**, Page 1273
This Journal is licensed under a Creative Commons Attribution 4.0 International License

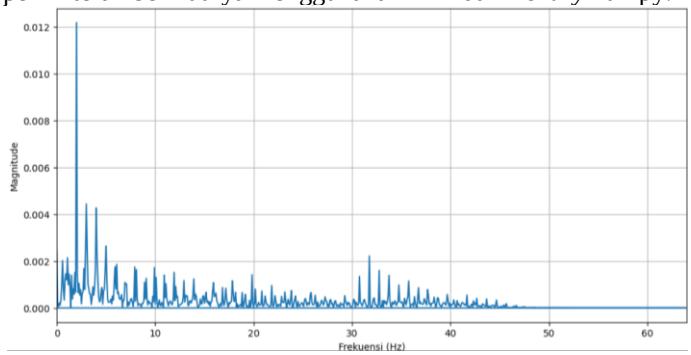
JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.9874
Hal 1269-1268
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom



Gambar 3. Data Setelah ICA

Gambar 2 menampilkan data EEG asli yang mengandung artefak dan noise dan Gambar 3 menampilkan data EEG setelah ICA dilakukan. Hasil *preprocessing* setelah ICA dilakukan menjadi stabil dan bersih untuk proses ekstraksi fitur dan klasifikasi.

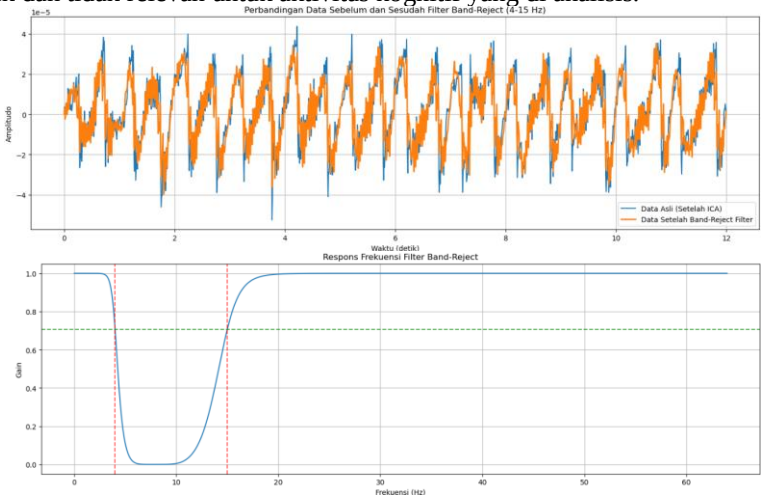
b. Tahap 2 transformasi *domain* frekuensi (FFT) mengkonversi sinyal dari domain waktu ke domain frekuensi untuk mempermudah proses pemfilteran berikutnya menggunakan FFT dari *library* numpy.



Gambar 4. Spektrum Frekuensi

Gambar 4 menampilkan spectrum frekuensi hasil FFT yang menggambarkan distribusi energi sinyal EEG dalam berbagai rentang frekuensi, sumbu vertical menunujukkan kekuatan sinyal dalam setiap frekuensi sedangkan sumbu horizontal menunjukkan nilai frekuensi. Hasil FFT terlihat kekuatan sinyal lebih besar pada frekuensi rendah.

c. Tahap 3 Pemfilteran *Band-Reject* akan menghilangkan frekuensi gangguan spesifik (4-15 Hz) yang sering mengandung artefak dan tidak relevan untuk aktivitas kognitif yang di analisis.



Gambar 5. Perbandingan Data Sebelum dan Sesudah Filter Band Reject (4-15 Hz)

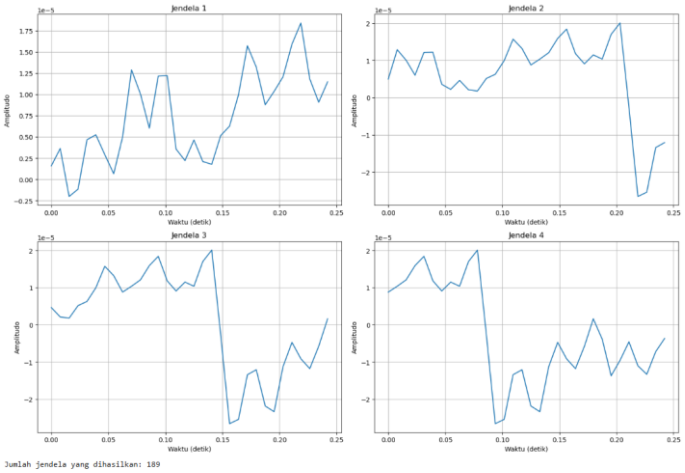
Gambar 5 menampilkan perbandingan sinyal EEG sebelum dan sesudah proses *filter band-reject*. Pada Persamaan (1), setiap sample sinyal EEG akan dikalikan dengan koefisien *filter* (*band-reject filter*) lalu dijumlah agar menghasilkan sinyal keluaran yang sudah di *filter*.Terlihat bentuk sinyal setelah *filtering* terdapat perubahan pada komponen amplitudo yang berhasil direduksi pada rentang 4-15 Hz. Hasil *filtering* ini menghasilkan sinyal EEG yang dapat digunakan untuk tahap ekstraksi fitur *wavelet*.

d. Tahap 4 transformasi waktu dimana mengembalikan sinyal yang sudah di *filter* kembali ke domain aslinya menggunakan *Inverse Fast Fourier Transform* (IFFT).

e. Tahap 5 segmentasi temporal (*sliding window*) akan memecah sinyal EEG yang panjang menjadi potongan-pootongan *windows* yang lebih kecil seperti gambar 6.

Copyright © 2026 The **Galuh Saputri**, Page 1274
This Journal is licensed under a Creative Commons Attribution 4.0 International License

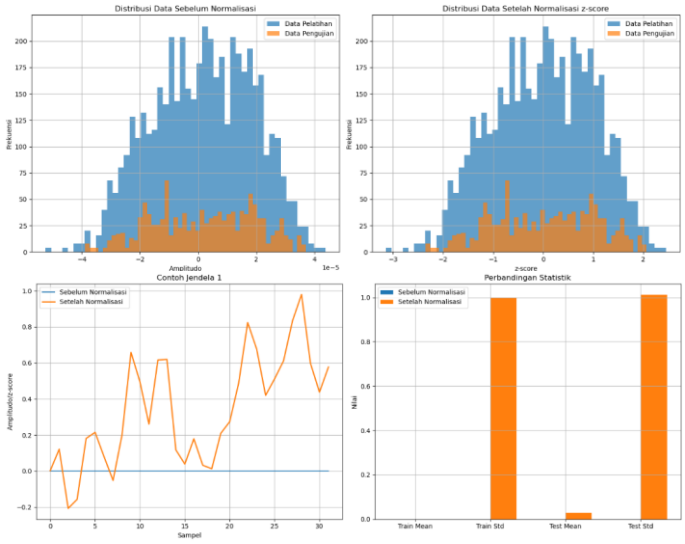
JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.9874
Hal 1269-1268
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom



Gambar 6. Segmentasi Temporal

Gambar 6 menampilkan hasil segmentasi sinyal ke beberapa window waktu yanh merepresentasi aktivitas EEG pada *interval* tertentu, Segmentasi temporal digunakan untuk mempertahankan karakteristik *local* sinyal supaya dapat dianalisis lebih efektif dalam proses klasifikasi, perekaman sinyal EEG dilakukan selama 10 detik dengan frekuensi sampling 128 Hz, menghasilkan 1280 sampel data per saluran untuk setiap subjek (siswa). Kemudian data ini disegmentasi dengan metode sliding window dengan window size sebesar 150 sampel dan langkah pergeseran sebesar 6 sampel dan menghasilkan 189 segmen data per subjek, setelah itu data akan digunakan pada tahap ekstraksi fitur.

f. Tahap 6 normalisasi adaptif akan menghitung rata-rata dan standar deviasi dari data pelatihan, lalu normalisasi setiap *window* data.



Gambar 7. Distribusi Data Sebelum dan Sesudah Normalisasi z-score

Gambar 7 menampilkan distribusi data sebelum dan sesudah normalisasi, normalisasi pada Persamaan (2) menghasilkan pengurangan nilai data terhadap nilai rata-rata lalu dibagi standar deviasi yang dimana setelah normalisasi data mendapat rata-rata yang mendekati nol dan standar deviasi yang lebih seragam.

3.1.2 Ekstraksi Fitur *Wavelet*

Proses ekstraksi fitur *wavelet* berhasil menghasilkan enam fitur representatif dari setiap jendela data EEG. Tabel 1 menunjukkan rincian fitur yang diekstraksi menggunakan transformasi *wavelet Daubechies-4* (db4). Fitur ini meliputi energi, rata-rata, dan standar deviasi dari koefisien aproksimasi (cA) dan detail (cD) yang dapat menangkap karakteristik sinyal EEG pada berbagai skala frekuensi.

Tabel 1. Fitur Importance dari Ekstraksi Wavelet

Fitur	Bobot
Energi_cA	0.287
Rata_cA	0.152
Std_cA	0.118
Energi_cD	0.325
Rata_cD	0.078
Std_cD	0.040

Copyright © 2026 The **Galuh Saputri**, Page 1275
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.9874
Hal 1269-1268
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Sinyal EEG didekomposisi menjadi koefisien aproksimasi (cA) dan koefisien detail (cD) menggunakan *Discrete Wavelet Transform* (DWT) dari persamaan (3) dan (4). Selanjutnya tiga fitur statistik yang ditentukan, yaitu energi, rata (*mean*), dan std (standar deviasi) dari setiap himpunan koefisien dihitung sebagai atribut input pada *Random Forest.* Berdasarkan hasil analisis fitur pada Tabel 1, Energi_cD memberikan kontribusi terbesar dengan bobot 0.325, diikuti oleh Energi_cA dengan bobot 0.287 dan Rata_cA 0.152. Ditunjukkan bahwa fitur energi pada koefisien detail dan aproksimasi berperan penting dalam mengklasifikasikan tingkat kemampuan bahasa berdasarkan sinyal EEG.

3.2 Klasifikasi dan Evaluasi Model

3.2.1 Klasifikasi dengan Random Forest

Tahap klasifikasi ini dilakukan untuk menentukan tingkat kemampuan bahasa menggunakan fitur-fitur yang berasal dari hasil ekstraksi sinyal EEG. Proses klasifikasi dalam penelitian ini menggunakan *model Random Forest* karena dapat menangani data berdimensi tinggi, mengurangi *overfitting*, memberikan informasi mengenai tingkat kepentingan setiap atribut yang digunakan dalam pengambilan keputusan. Setelah proses ekstraksi fitur *wavelet* diperoleh atribut Energi_cA, Rata_cA, Std_cA, Energi_cD, Rata_cD, Std_cD yang dipakai sebagai input features dalam *Random Forest*. Kemudian nilai *Gini Impurity* dalam Persamaan (5) dan (6) akan membangun 100 pohon keputusan dan pada setiap node dilakukan pemilihan atribut terbaik. Atribut dengan hasil nilai *Gain* terbesar dipilih sebagai pemisah data dalam *node* tersebut, proses akan dilakukan berulang sampai seluruh pohon keputusan terbentuk. Hasil perhitungan menunjukan atribut Energi_cD dan Energi_cA paling sering memberikan penurunan *impurity* terbesar sehingga mempunyai peran paling tinggi pada klasifikasi tingkat kemampuan bahasa dengan *Random Forest.* Sebelum melakukan proses klasifikasi, teknik *oversampling* digunakan untuk menyeimbangkan data antar kelas dan mengurangi potensi bias model kepada kelas mayoritas yang ditunjukkan pada Tabel 2.

Tabel 2. Distribusi Kelas Seblum dan Sesudah Oversampling

Kelas	Sebelum Oversampling	Sesudah Oversampling
Low	22	22
Medium	20	22
High	3	22

Distribusi kelas sebelum dan sesudah oversampling ada pada Tabel 2. Data awal menampilkan *imbalance* yang signifikan dengan kelas *Low* memiliki 22 sampel, *Medium* 20 sampel, dan *High* hanya 3 sampel. Melalui teknik *oversampling* dengan metode *resampling*, distribusi kelas berhasil disamakan menjadi 22 sampel untuk setiap kelas.

Tabel 3. Data Training dan Testing

No.	Kelas	Data Training (80%)	Data Testing (20%)	Total Oversampling
1.	*Low*	18	4	22
2.	*Medium*	17	5	22
3.	*High*	17	5	22
Total		52	14	66

Pada tabel 3 sebelum proses pelatihan dilakukan, data yang sudah di *oversampling* dibagi menjadi 2 bagian, yaitu *Data Training* 80% untuk melatih model machine learning dan *Data Testing* 20% untuk menguji performa model yang dapat memprediksi label baru dari pola yang sudah dipelajari.

Tabel 4. Hasil Klasifikasi Random Forest

Kelas	Precision	Recall	F1-Score
Low	1.00	0.50	0.67
Medium	0.83	1.00	0.91
High	0.83	1.00	0.91
Accuracy			**0.86**
Macro Avg	**0.89**	**0.83**	**0.83**

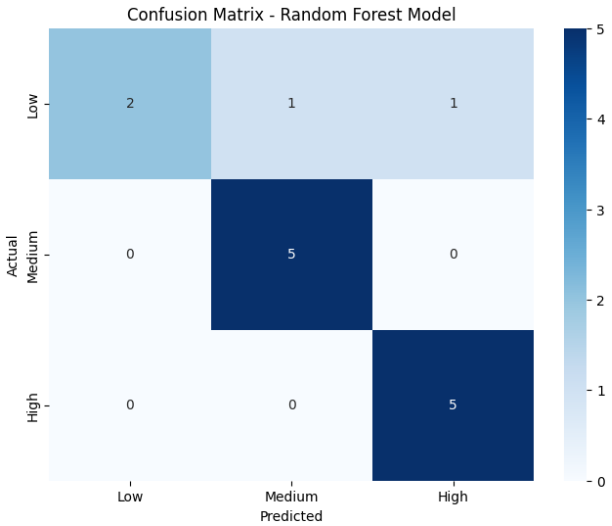
Model Random Forest berhasil mengklasifikasikan tingkat kemampuan bahasa dengan akurasi (*accuracy*) 0.86 pada data *testing*. Tabel 4 menunjukkan hasil evaluasi detail untuk setiap kelas, nilai total (*macro average*) *precision* 0.89, *recall* 0.83, dan *f1-score* 0.83.

3.2.2 Evaluasi Model

Evaluasi model menggunakan *accuracy matrix, precision, recall,* dan *f1-score* menunjukkan hasil yang memuaskan. Akurasi keseluruhan mencapai 0.86 dengan *cross-validation score* sebesar 0.68.

Copyright © 2026 The **Galuh Saputri**, Page 1276
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.9874
Hal 1269-1268
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom



Gambar 8. Confusion Matrix

Confusion matrix pada Gambar 8 menampilkan *model Random Forest* dapat mengklasifikasikan data tingkat kemampuan bahasa dengan baik. Kelas *Medium* dan *High* berhasil diklasifikasikan sebanyak 5 data, lalu kelas *Low* ada 2 data berhasil diklasifikasi dan 2 data gagal klasifikasi. Proses *split* dengan perhitungan *Gini Impurity* dan *Gain* dalam persamaan (5) dan (6) mampu membantu model untuk memisahkan kelas data dari fitur hasil ekstraksi sinyal EEG secara optimal.

4. KESIMPULAN

Kombinasi ekstraksi fitur wavelet dan algoritma Random Forest menghasilkan kemampuan yang baik dalam mengidentifikasi pola sinyal EEG untuk klasifikasi tingkat kemampuan bahasa anak dalam 3 kategori yaitu Low, Medium, High. Penggunaaan sinyal EEG direpresentasikan dengan fitur wavelet sebagai masukan untuk proses pembelajaran Random Forest dan hasil evaluasi memperoleh akurasi 0.86, precision 0.89, recall 0.83, F1-score 0.83. Pada penelitian ini, kategori tersebut digunakan sebagai label operasional yang ditetapkan berdasarkan nilai akademik bahasa, sehingga model mempelajari keterkaitan antara karakteristik neurofisiologis sinyal EEG dan kategori kemampuan bahasa yang sudah ditentukan dari data nilai akademik bahasa, bukan mengukur kemampuan bahasa secara langsung seperti instrumen psikometrik. Ekstraksi fitur wavelet memberikan representasi karakteristik sinyal EEG pada domain waktu frekuensi dan Random Forest digunakan untuk mempelajari hubungan nonlinear antar fitur dalam proses klasifikasi. Kontribusi ilmiah penelitian ini terletak pada penerapan pengolahan sinyal EEG berbasis fitur wavelet dan Random Forest untuk menghubungkan karakteristik aktivitas otak dengan kategori kemampuan bahasa yang ditetapkan melalui indikator akademik, sehingga memperluas pemanfaatan EEG yang sebelumnya banyak digunakan untuk pengenalan stimulus atau aktivitas bahasa menuju kajian klasifikasi kemampuan bahasa. Hasil penelitian ini memberikan bukti awal mengenai potensi karakteristik sinyal EEG sebagai sumber informasi neurofisiologis yang berkaitan dengan kemampuan bahasa anak, tetapi interpretasi hasil perlu mempertimbangkan keterbatasan ukuran sampel, keterkaitan antara rekaman EEG dan nilai akademik pada tingkat subjek, validitas pembentukan label Low, Medium, High, dan konsistensi pembagian data dan prosedur oversampling yang digunakan dalam proses pemodelan. Penelitian berikutnya dapat melibatkan jumlah subjek yang lebih besar, data EEG dan nilai akademik yang terhubung secara konsisten pada subjek yang sama, validasi label menggunakan instrumen kemampuan bahasa yang lebih terstandar, serta validasi eksternal untuk menguji kemampuan generalisasi model. Perbandingan beberapa algoritma klasifikasi dan konfigurasi fitur juga dapat dilakukan untuk memperoleh model yang lebih dipertanggungjawabkan secara ilmiah.

UCAPAN TERIMAKASIH

Terima kasih disampaikan kepada Univeritas Ahmad Dahlan dan Universitas Pamulang atas dukungan dan fasilitas yang diberikan selama pelaksanaan penelitian, terima kasih juga disampaikan kepada SD Muhammadiyah Argosari, SD Muhammadiyah Karangkajen, dan SDN Keputran I Yogyakarta yang telah memberikan izin dalam proses pengumpulan dan analisis data, sehingga selesai penyusunan artikel ini.

REFERENCES

[1] P. Mariani, D. F. Kurnia, and L. Yarni, “Kecerdasan Majemuk (Multiple Intelligence),” *KOLONI*, vol. 2, no. 4, pp. 201–212, Dec. 2023, doi: 10.31004/KOLONI.V2I4.566.

Copyright © 2026 The **Galuh Saputri**, Page 1277
This Journal is licensed under a Creative Commons Attribution 4.0 International License

[2] P. Peng and R. A. Kievit, "The Development of Academic Achievement and Cognitive Abilities: A Bidirectional Perspective," *Child Dev. Perspect.*, vol. 14, no. 1, pp. 15–20, 2020, doi: 10.1111/cdep.12352.
[3] V. Cook, "The language in language and thinking," *Vigo Int. J. Appl. Linguist.*, no. 18, pp. 35–58, Jan. 2021, doi: 10.35869/VIAL.V0I18.3364.
[4] "PISA 2022 Assessment and Analytical Framework," Aug. 2023, doi: 10.1787/DFE0BF9C-EN.
[5] "PISA 2022 Results (Volume III)," Jun. 2024, doi: 10.1787/765EE8C2-EN.
[6] A. P. Fitriani, A. Wijayanti, and D. P. Koesmadi, "Meningkatkan Kecerdasan Linguistik Anak Usia Dini Dengan Menggunakan Buku Language Smart Kids," *J. Pendidik. Anak Usia Dini Undiksha*, vol. 9, no. 2, pp. 270–276, Jul. 2021, doi: 10.23887/PAUD.V9I2.34123.
[7] A. T. Inayah *et al.*, "Objectivity in subjectivity: do students' self and peer assessments correlate with examiners' subjective and objective assessment in clinical skills? A prospective study," *BMJ Open*, vol. 7, no. 5, p. e012289, May 2017, doi: 10.1136/BMJOPEN-2016-012289.
[8] J. Decety, K. L. Meidenbauer, and J. M. Cowell, "The development of cognitive empathy and concern in preschool children: A behavioral neuroscience investigation," *Dev. Sci.*, vol. 21, no. 3, pp. 1–12, 2018, doi: 10.1111/desc.12570.
[9] S. Bagherzadeh, K. Maghooli, A. Shalbaf, and A. Maghsoudi, "A Hybrid EEG-Based Emotion Recognition Approach Using Wavelet Convolutional Neural Networks (WCNN) and Support Vector Machine," *Basic Clin. Neurosci.*, vol. 14, no. 1, pp. 87–102, 2023, doi: 10.32598/bcn.2021.3133.1.
[10] T. A. Aldhaheri, S. B. Kulkarni, P. R. Bhise, and B. Ghaleb, "Investigation of brain response to acquisition and learning the second languages based on EEG signals and machine learning techniques," *Cogent Arts Humanit.*, vol. 11, no. 1, p., 2024, doi: 10.1080/23311983.2024.2416759.
[11] M. Melinda, F. H. Juwono, I. K. A. Enriko, M. Oktiana, S. Mulyani, and K. Saddami, "Application of Continuous Wavelet Transform and Support Vector Machine for Autism Spectrum Disorder Electroencephalography Signal Classification," *Radioelectron. Comput. Syst.*, no. 3(107), pp. 73–90, 2023, doi: 10.32620/reks.2023.3.07.
[12] S. Kumar G S, N. Sampathila, and T. Tanmay, "Wavelet based machine learning models for classification of human emotions using EEG signal," *Meas. Sensors*, vol. 24, no. November, p. 100554, 2022, doi: 10.1016/j.measen.2022.100554.
[13] M. M. Abdulghani, W. L. Walters, and K. H. Abed, "Imagined Speech Classification Using EEG and Deep Learning," *Bioengineering*, vol. 10, no. 6, 2023, doi: 10.3390/bioengineering10060649.
[14] F. Elwasify, E. Shaaban, and R. M. Abdelmoneem, "EEG imagined speech neuro-signal preprocessing and deep learning classification," *Sci. Rep.*, vol. 16, no. 1, pp. 1–19, 2026, doi: 10.1038/s41598-026-39395-6.
[15] P. Kumar, R. Saini, P. P. Roy, P. K. Sahu, and D. P. Dogra, "Envisioned speech recognition using EEG sensors," *Pers. Ubiquitous Comput. 2017 221*, vol. 22, no. 1, pp. 185–199, Sep. 2017, doi: 10.1007/S00779-017-1083-4.
[16] A. Derakhshesh, Z. Dehghanian, R. Ebrahimpour, and H. R. Rabiee, "Cueless EEG Imagined Speech for Subject Identification: Dataset and Benchmarks," *IEEE Trans. Biometrics, Behav. Identity Sci.*, pp. 1–13, 2025, doi: 10.1109/TBIOM.2025.3634273.
[17] A. Delorme, T. Sejnowski, and S. Makeig, "Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis," *Neuroimage*, vol. 34, no. 4, pp. 1443–1449, 2007, doi: 10.1016/j.neuroimage.2006.11.004.
[18] A. Yudhana, A. Muslim, D. E. Wati, I. Puspitasari, A. Azhari, and M. M. Mardhia, "Human emotion recognition based on EEG signal using fast fourier transform and K-Nearest neighbor," *Adv. Sci. Technol. Eng. Syst.*, vol. 5, no. 6, pp. 1082–1088, 2020, doi: 10.25046/aj0506131.
[19] A. Yudhana *et al.*, "Recognizing Human Emotion patterns by applying Fast Fourier Transform based on Brainwave Features," *Proc. - 1st Int. Conf. Informatics, Multimedia, Cyber Inf. Syst. ICIMCIS 2019*, pp. 249–254, Oct. 2019, doi: 10.1109/ICIMCIS48181.2019.8985227.
[20] R. Alfina, I. Arifianto, D. Astharini, and P. Wulandari, "Mendisain GUI Untuk Menampilkan Nilai FFT dan IFFT Menggunakan LabVIEW," *TESLA J. Tek. Elektro*, vol. 21, no. 1, pp. 50–56, Mar. 2019, doi: 10.24912/TESLA.V21I1.3250.
[21] A. Solichin, "Analisis Komparatif Akurasi Deteksi Teks Dokumen Keuangan Menggunakan CTPN dan EAST," vol. 18, no. 2, pp. 274–286, 2026.
[22] D. Haryanto, M. E. Agustian, K. Muhammad, and W. Hidayat, "AUTENTIKASI BERKELANJUTAN BERBASIS POLA KEYSTROKE MENGGUNAKAN METODE DEEP LEARNING," *KHARISMA Tech*, vol. 21, no. 1, pp. 179–187, Mar. 2026, doi: 10.55645/KHARISMATECH.V21I1.692.
[23] Y. Lu, W. Wang, B. Lian, and C. He, "Feature Extraction and Classification of Motor Imagery EEG Signals in Motor Imagery for Sustainable Brain–Computer Interfaces," *Sustain.*, vol. 16, no. 15, 2024, doi: 10.3390/su16156627.
[24] M. Muthulakshmi, R. Munirathinam, M. Gayathri, and K. L. Nayana Sree, "EEG-Based Identification of Huntington's Disease using Daubechies Wavelet features with Decision Tree Classifier," *2025 Int. Conf. Sensors Relat. Networks, SENNET 2025 - Spec. Focus Digit. Healthc.*, 2025, doi: 10.1109/SENNET64220.2025.11135963.
[25] T. B. Putri, S. Saidah, B. Hidayat, F. Qothrunnada, T. Telekomunikasi, and U. Telkom, "Deteksi Emosi Berdasarkan Sinyal Suara Manusia Menggunakan Discrete Wavelet Transform (DWT) Dengan Klasifikasi Support Vector Machine (SVM)," vol. 3, no. 1, pp. 1–10, 2023.

Copyright © 2026 The **Galuh Saputri**, Page 1278
This Journal is licensed under a Creative Commons Attribution 4.0 International License

[26] M. A. Rasheed *et al.*, "Use of artificial intelligence on Electroencephalogram (EEG) waveforms to predict failure in early school grades in children from a rural cohort in Pakistan," *PLoS One*, vol. 16, no. 2 February, pp. 1–12, 2021, doi: 10.1371/journal.pone.0246236.
[27] A. Fadlil, I. Riadi, I. Julda, and D. E. Purwadi Putra, "Comparison of Machine Learning Performance Using Naive Bayes and Random Forest Methods to Classify Batik Fabric Patterns", doi: 10.18280/ria.370214.
[28] M. A. M. Mustofa, Sucipto, and A. Nugroho, "PENERAPAN RANDOM FOREST UNTUK DETEKSI DINI PENYAKIT PARKINSON'S DENGAN DATA FREKUENSI SUARA," *J. Qua Tek.*, vol. 15, no. 02, pp. 25–37, Sep. 2025, doi: 10.35457/QUATEKNIKA.V15I02.4563.
[29] H. Sumarti, "EEG Classification while Listening to Murottal Al-Quran and Classical Music using Random Forest Method EEG Classification while Listening to Murottal Al-Quran and Classical Music using Random Forest Method," vol. 6, no. 2, 2023.
[30] J. Mega Reski, A. Yudhana, U. Ahmad Dahlan Jl Ringroad Selatan, K. Banguntapan, and D. Istimewa Yogyakarta, "Classification of Crystallization Images of Pharmaceutical Raw Materials Using Convolutional Neural Network Algorithm," *Int. J. Adv. Data Inf. Syst.*, vol. 6, no. 3, pp. 705–715, Dec. 2025, doi: 10.59395/IJADIS.V6I3.1440.

Copyright © 2026 The **Galuh Saputri**, Page 1279
This Journal is licensed under a Creative Commons Attribution 4.0 International License