Jurnal Informatika Vol. 15, No. 1, Juni 2026, Page. 241~252
E-ISSN: 2528-0090 https://ejournal.unidayan.ac.id/index.php/JIU
P-ISSN: 2252-9160 DOI: https://doi.org/10.55340/jiu.v15i1.2689

Sistem Penyaring Beras Otomatis Berbasis ESP32 Menggunakan Random Forest Regression untuk Prediksi Efisiensi dan Kualitas Hasil Pembersihan Beras

ESP32-Based Automatic Rice Sorting System Using Random Forest Regression for Predicting Cleaning Efficiency and Quality

Asniati[1], Ery Muchyar Hasiri[2], Mohamad Arif Suryawan[3], Jabal Nur[4]
[1,2,3,4]Program Studi Teknik Informatika, Universitas Dayanu Ikhsanuddin Baubau
E-mail: [*2]erymuchyar82@um.kendari.ac.id

Article Info:	Received	Revised	Accepted	Published:
	14 Mei 2026	16 Mei 2026	24 Juni 2026	28 Juni 2026

Abstrak

Proses pembersihan beras secara manual masih menghadapi kendala dalam menjaga konsistensi kualitas dan efisiensi produksi. Sistem otomatis berbasis sensor yang ada umumnya hanya berfungsi untuk pemantauan tanpa kemampuan prediksi kuantitatif, sehingga diperlukan integrasi multi-sensor dan model prediktif pada perangkat keras berbiaya rendah. Penelitian ini bertujuan merancang dan mengevaluasi Sistem Penyaring Beras Otomatis berbasis ESP32 yang terintegrasi dengan algoritma Random Forest Regression (RFR) untuk memprediksi efisiensi pemisahan kotoran dan skor kualitas pembersihan secara simultan. Sistem memanfaatkan 20 fitur sensor yang berasal dari sensor LDR, sensor berat, DHT22, encoder RPM kipas, dan INA219. Dataset terdiri atas 3.000 rekaman, yaitu 600 data empiris hasil pengujian fisik pada rentang kapasitas 1–200 kg dan 2.400 data sintetik yang dihasilkan melalui augmentasi Gaussian Noise Injection. Data dibagi menjadi data latih dan data uji dengan rasio 80:20, kemudian dievaluasi menggunakan metrik koefisien determinasi (R^2) dan Mean Absolute Error (MAE). Model menghasilkan R^2 = 0,7326 dan MAE = 2,067% untuk prediksi efisiensi, serta R^2 = 0,6825 dan MAE = 0,97 untuk prediksi kualitas. Pengujian fisik dilakukan pada kapasitas 1–200 kg, sedangkan kapasitas 201–1.000 kg dievaluasi menggunakan data sintetik hasil augmentasi. Keterbaruan penelitian terletak pada kemampuan memprediksi dua target numerik secara simultan dalam satu sistem berbasis ESP32 berbiaya rendah. Sistem berpotensi mendukung pengambilan keputusan otomatis untuk meningkatkan efisiensi dan kualitas pembersihan beras pada industri penggilingan padi skala kecil dan menengah.

Kata kunci: *ESP32; IoT Pertanian; Penyaring Beras Otomatis; Prediksi Efisiensi Pembersihan Beras; Random Forest Regression.*

Abstract

Manual rice cleaning processes continue to face challenges in maintaining consistent cleaning quality and production efficiency. Existing sensor-based automated systems generally function only for process monitoring without quantitative predictive capabilities, highlighting the need for integrating multi-sensor data with predictive models on low-cost hardware platforms. This study aims to design and evaluate an ESP32-based Automatic Rice Cleaning System integrated with the Random Forest Regression (RFR) algorithm to simultaneously predict impurity removal efficiency and cleaning quality scores. The system utilizes 20 sensor features collected from LDR sensors, load cells, DHT22, fan RPM encoders, and INA219 sensors. The dataset consists of 3,000 records, including 600 empirical records obtained from physical experiments with capacities ranging from 1–200 kg and 2,400 synthetic records generated using Gaussian Noise Injection augmentation. The dataset was divided into training and testing sets with an 80:20 ratio and evaluated using the coefficient of determination (R^2) and Mean Absolute Error (MAE). The proposed model achieved an R^2 of 0.7326 and an MAE of 2.067% for efficiency prediction, and an R^2 of 0.6825 with an MAE of 0.97 for cleaning quality prediction. Physical validation was conducted for capacities of 1–200 kg, while capacities of 201–1,000 kg were evaluated using augmented synthetic data. The novelty of this study lies in its ability to simultaneously predict two numerical targets within a single low-cost ESP32-based system. The proposed system has the potential to support automated decision-making for improving rice cleaning efficiency and quality in small- and medium-scale rice milling industries.

Jurnal Informatika Vol. 15, No. 1, Juni 2026, Page. 241~252
E-ISSN: 2528-0090 https://ejournal.unidayan.ac.id/index.php/JIU
P-ISSN: 2252-9160 DOI: https://doi.org/10.55340/jiu.v15i1.2689

***Keywords**: ESP32; Agricultural Internet of Things (IoT); Automatic Rice Cleaning System; Rice Cleaning Efficiency Prediction; Random Forest Regression.*

This is an open access article under the CC BY-SA license.

1. PENDAHULUAN

Beras merupakan komoditas pangan strategis yang memiliki peran utama dalam ketahanan pangan, konsumsi rumah tangga, dan rantai nilai agrikultur di berbagai negara, termasuk Indonesia. Secara global, beras dikonsumsi oleh lebih dari separuh populasi dunia dan menjadi sumber energi utama bagi masyarakat Asia, sehingga mutu beras tidak hanya berkaitan dengan nilai ekonomi, tetapi juga dengan keamanan pangan dan penerimaan konsumen [1],[2]. Dalam rantai pascapanen, kualitas beras dipengaruhi oleh proses pengeringan, penyimpanan, penggilingan, pembersihan, sortasi, dan distribusi. Tahap pembersihan menjadi penting karena keberadaan sekam, debu, batu kecil, butir patah, maupun benda asing dapat menurunkan kualitas fisik, mengganggu proses pengemasan, serta menurunkan nilai jual produk beras. Keputusan pengolahan, penyimpanan, dan penanganan beras setelah panen secara langsung memengaruhi efisiensi ekonomi serta mutu produk yang diterima pasar [3],[4].

Proses pembersihan dan penilaian kualitas beras secara konvensional masih banyak bergantung pada pengamatan manual, tenaga operator, dan peralatan sederhana. Metode manual bersifat subjektif, memakan waktu, dan hasilnya bergantung pada pengalaman serta kondisi fisik operator. Dalam konteks kualitas beras, beberapa indikator fisik seperti bentuk, ukuran, warna, keutuhan butir, chalkiness, kebersihan, dan keberadaan kotoran menjadi parameter penting yang memengaruhi klasifikasi mutu dan penerimaan konsumen [5], [6], [7]. Penelitian berbasis computer vision juga menegaskan bahwa inspeksi visual manual pada beras cenderung tidak konsisten, sehingga diperlukan sistem otomatis yang mampu memberikan pengukuran lebih objektif, cepat, dan terukur [9], [22], [23].

Perkembangan mikrokontroler, sensor, aktuator, dan Internet of Things (IoT) membuka peluang untuk meningkatkan efisiensi proses pascapanen melalui otomasi berbasis data. Andriani et al. [1] mengembangkan sistem penampi dan penimbang beras berbasis IoT yang mengintegrasikan NodeMCU ESP8266 dengan berbagai sensor dan aktuator untuk mendukung proses pembersihan, penimbangan, serta pencatatan hasil secara otomatis. Sejalan dengan itu, beberapa studi pada sistem penimbangan beras digital menunjukkan bahwa kombinasi sensor load cell, mikrokontroler, dan aktuator mampu meningkatkan akurasi pengukuran massa sekaligus mengurangi ketergantungan pada penimbangan manual [10], [11], [12], [13]. Dalam konteks pengendalian proses pascapanen, beberapa kajian menunjukkan bahwa sistem berbasis mikrokontroler yang dilengkapi sensor suhu dan elemen pemanas mampu mengontrol proses pengeringan beras secara lebih terukur dan konsisten [18].

Berbagai algoritma machine learning telah digunakan untuk evaluasi mutu beras, di antaranya ANN dan spektroskopi NIR non-destruktif [6], [14], [15], image processing berbasis Raspberry Pi, dan computer vision/ANN untuk klasifikasi morfo-kolorimetri [5], [12], serta CNN/SVM/KNN untuk klasifikasi mutu, varietas, dan kemurnian benih padi [4], [7], [12], [16]. Di antara pendekatan tersebut, Random Forest Regression (RFR) adalah metode ensemble yang stabil dan tahan overfitting [19]. Kemampuan RFR menangkap hubungan nonlinier antara fitur sensor tabular dan variabel target numerik menjadikannya sesuai untuk memodelkan efisiensi dan kualitas penyaringan beras dalam sistem IoT [21].

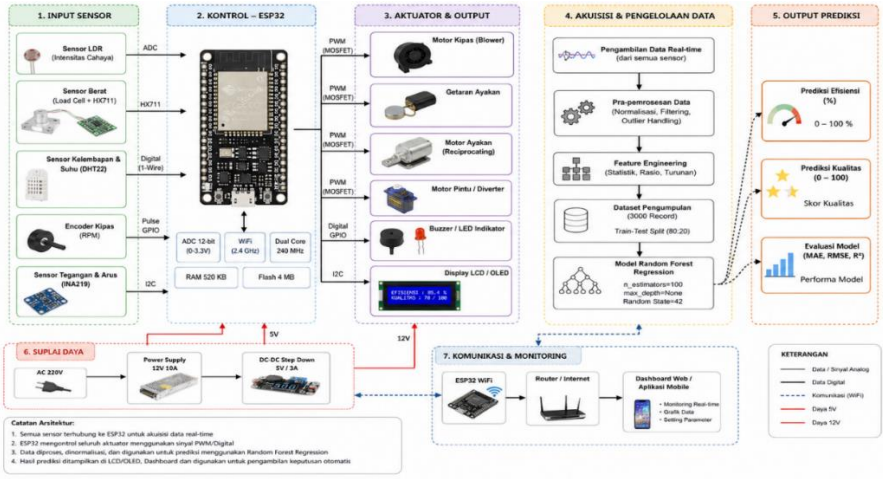
Meskipun berbagai penelitian telah membahas otomasi penimbangan beras, penampian berbasis IoT, klasifikasi kualitas beras berbasis citra, serta prediksi kualitas beras menggunakan machine learning, masih terdapat celah pada integrasi sistem penyaring beras otomatis berbasis ESP32 dengan model prediksi efisiensi dan kualitas hasil pembersihan secara numerik [10], [16]. Penelitian terdahulu umumnya berfokus pada satu aspek saja, seperti perangkat penimbang atau penampi beras, klasifikasi citra beras, prediksi kualitas berbasis spektroskopi, atau pemantauan pertanian berbasis IoT. Belum terdapat kajian yang secara spesifik menggabungkan akuisisi data multi-sensor, formulasi indikator efisiensi pemisahan kotoran, dan model RFR dalam satu ekosistem perangkat keras berbasis ESP32. Pendekatan yang mengintegrasikan ESP32, sensor lingkungan, dan Random Forest Regression untuk prediksi berbasis data sensor dalam konteks IoT pertanian telah menunjukkan potensi yang menjanjikan [17], [21].

Berdasarkan celah penelitian yang telah diidentifikasi, penelitian ini bertujuan merancang dan mengevaluasi Sistem Penyaring Beras Otomatis berbasis ESP32 yang mengintegrasikan algoritma Random Forest Regression untuk memprediksi efisiensi pemisahan kotoran dan kualitas hasil pembersihan beras. Sistem memanfaatkan sensor LDR, sensor berat, DHT22, INA219, dan encoder kipas untuk memperoleh data proses penyaringan. Kontribusi utama penelitian ini adalah integrasi perangkat keras ESP32, akuisisi data multi-sensor, serta model prediktif yang mampu mengevaluasi efisiensi dan kualitas pembersihan secara kuantitatif.

2. METODE PENELITIAN

2.1 Kerangka Sistem

Sistem dirancang dalam lima lapisan fungsional: (1) sensor input, (2) kontrol ESP32, (3) aktuator dan output mekanis, (4) analitik data, dan (5) prediksi Random Forest, sebagaimana ditunjukkan pada Gambar 1. Sensor LDR, load cell/HX711, DHT22, encoder, dan INA219 mengakuisisi data beras secara real-time dan mengirimkannya ke ESP32 untuk mengatur kipas, motor getar, servo, serta indikator keluaran. Data proses selanjutnya digunakan oleh model Random Forest Regression untuk memprediksi efisiensi pembersihan (%) dan skor kualitas (0–100), yang ditampilkan pada layar atau dashboard monitoring sebagai dasar penyesuaian operasional.



Gambar 1. Kerangka Sistem Penyaring Beras Otomatis Berbasis ESP32 dan Random Forest Regression

2.2 Perangkat Keras

Komponen perangkat keras utama yang digunakan meliputi: ESP32 DevKit v1 sebagai unit pemroses utama, serta sensor LDR (Light Dependent Resistor) pasangan laser-fotodeteksi untuk mengukur aliran beras melalui parameter ldr_light_adc, ldr_dark_adc, dan durasi_sensor_terhalang_s. Sensor DHT22 mengukur kelembapan dan suhu lingkungan; modul INA219 mengukur tegangan, arus, dan daya yang dikonsumsi kipas; serta encoder Hall-effect pada kipas sentrifugal untuk menghasilkan nilai RPM. Ayakan bergetar dikendalikan melalui variasi sudut kemiringan (8°–15°) dan ukuran lubang (2–3 mm). Semua data sensor dikomunikasikan melalui antarmuka I2C, SPI, dan GPIO pada ESP32.

2.3 Pengumpulan Dataset

Dataset terdiri dari dua komponen: data empiris dan data sintetik hasil augmentasi. Data empiris dikumpulkan melalui pengujian fisik terkontrol di Desa Barangka, Kabupaten Buton, mencakup variasi berat 1–200 kg (interval 1 kg, 3 replikasi per titik), menghasilkan 600 rekaman nyata. Pengujian dilakukan langsung menggunakan prototipe sistem penyaring yang dijelaskan pada Subbagian 2.2, dengan setiap titik berat direplikasi tiga kali untuk menjamin konsistensi pengukuran.

Untuk memperluas cakupan rentang berat dan meningkatkan kemampuan generalisasi model, data sintetik dihasilkan untuk rentang 201–1.000 kg (800 titik × 3 replikasi = 2.400 rekaman) menggunakan teknik Gaussian Noise Injection. Teknik ini menambahkan gangguan acak berdistribusi normal (mean = 0, standar deviasi = 1% dari nilai fitur) ke data empiris yang ada, sehingga distribusi statistik data sintetik tetap konsisten dengan data real. Validasi keselarasan distribusi dilakukan menggunakan Kolmogorov-Smirnov test ($p > 0.05$), yang mengonfirmasi bahwa data sintetik tidak menyimpang secara signifikan dari pola empiris. Total dataset mencakup 3.000 rekaman (600 empiris + 2.400 sintetik). Penting untuk dicatat bahwa validasi performa model berdasarkan data observasi nyata hanya berlaku pada rentang pengujian fisik 1–200 kg. Prediksi pada rentang

201–1.000 kg merupakan proyeksi model berbasis augmentasi statistik, bukan hasil validasi eksperimen; interpretasi performa pada rentang tersebut harus diperlakukan sebagai eksplorasi kapasitas model, bukan klaim kinerja operasional yang terverifikasi. Variabel input dan target disajikan pada Tabel 1; fitur turunan dijelaskan pada Subbagian 2.5.

Tabel 1. Deskripsi Variabel Dataset Sistem Penyaring Beras

Variabel	Tipe	Satuan	Deskripsi
berat_awal_g	Numerik	gram	Berat awal campuran beras + kotoran
tingkat_kotoran_awal_pct	Numerik	%	Persentase kotoran awal
ldr_drop_adc	Numerik	ADC	Selisih LDR light-dark; sensitivitas deteksi
durasi_sensor_terhalang_s	Numerik	detik	Durasi aliran beras pada jalur laser-LDR
pwm_kipas	Numerik	0–255	Nilai PWM kendali kecepatan kipas
rpm_kipas	Numerik	RPM	Estimasi putaran kipas per menit
airflow_mps	Numerik	m/s	Estimasi kecepatan aliran udara kipas
sudut_ayakan_deg	Numerik	derajat	Kemiringan ayakan (8°–15°)
ukuran_lubang_ayakan_mm	Numerik	mm	Ukuran lubang ayakan (2–3 mm)
kelembapan_beras_pct	Numerik	%	Kadar kelembapan beras
daya_w	Numerik	W	Konsumsi daya kipas/sistem
ldr_light_adc	Numerik	ADC	Nilai ADC sensor LDR pada kondisi terang (laser tidak terhalang)
ldr_dark_adc	Numerik	ADC	Nilai ADC sensor LDR pada kondisi gelap (laser terhalang beras)
kotoran_awal_g	Numerik	gram	Massa kotoran awal = berat_awal_g × tingkat_kotoran_awal_pct / 100
berat_akhir_g	Numerik	gram	Berat beras bersih setelah proses penyaringan selesai
kotoran_terpisah_g	Numerik	gram	Massa kotoran yang berhasil dipisahkan dari beras selama proses
grain_loss_g	Numerik	gram	Massa butir beras baik yang ikut terbuang bersama kotoran
butir_patah_pct	Numerik	%	Persentase butir patah terhadap total beras setelah penyaringan
efisiensi_real_pct	Target RF	%	Kotoran terpisah / kotoran awal × 100%
skor_kualitas_real	Target RF	0–100	Skor kualitas berbasis kotoran tersisa, grain loss, butir patah

2.3.1 Formulasi Gaussian Noise Injection

Augmentasi data sintetik menggunakan teknik Gaussian Noise Injection mengikuti formulasi yang ditetapkan oleh Arslan et al. [24] dan Kim and Chung [25]. Untuk setiap rekaman data empiris x_i, nilai sintetik $x_i\tilde{}$ diperoleh melalui penambahan noise acak berdistribusi normal sebagaimana ditunjukkan pada Persamaan (1).

$$x_i\tilde{} = x_i + \varepsilon, \qquad \varepsilon \sim N(0, \sigma^2) \qquad \text{Persamaan 1}$$

di mana ε adalah vektor noise yang disampling dari distribusi normal dengan mean $\mu = 0$ dan standar deviasi $\sigma = 0{,}01 \times x_i$ (setara dengan 1% dari nilai fitur). Pendekatan ini memastikan bahwa nilai mean data tidak berubah setelah augmentasi, sementara variasi kecil yang diperkenalkan berfungsi meningkatkan robustisitas dan kemampuan generalisasi model [24]. Konsistensi distribusi antara data empiris dan sintetik divalidasi menggunakan Kolmogorov-Smirnov test dengan ambang batas $p > 0{,}05$.

2.4 Formulasi Random Forest Regression (RFR)

Notasi yang digunakan meliputi X sebagai matriks fitur input, y_i sebagai nilai aktual, ŷ_i sebagai nilai prediksi, n sebagai jumlah data uji, B sebagai jumlah pohon keputusan, dan T_b(X) sebagai keluaran pohon ke-b. Formulasi RFR mengacu pada Breiman [19] dan metrik evaluasi merujuk pada [20].

Prediksi akhir Random Forest Regression diperoleh dari rata-rata keluaran seluruh pohon keputusan regresi sebagaimana ditunjukkan pada Persamaan (2).

$$\hat{y} = \frac{1}{B} \sum_{b=1}^{B} T_b(X) \qquad (2)$$

Pada setiap pohon regresi, pemilihan percabangan dilakukan dengan meminimalkan galat kuadrat rata-rata pada setiap simpul. Fungsi impurity untuk node regresi ditunjukkan pada Persamaan (3).

$$MSE_{node} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 \quad (3)$$

Akurasi prediksi model dievaluasi menggunakan Mean Absolute Error (MAE), Root Mean Square Error (RMSE), koefisien determinasi (R^2), dan Mean Absolute Percentage Error (MAPE). MAE mengukur rata-rata selisih absolut antara nilai aktual dan prediksi, sebagaimana pada Persamaan (4). Penggunaan MAE, RMSE, R^2, dan MAPE sebagai metrik evaluasi regresi lazim digunakan untuk menilai besaran error absolut, penalti error yang besar, proporsi variasi yang dijelaskan oleh model, serta tingkat kesalahan relatif dalam persentase.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (4)$$

RMSE memberikan penalti yang lebih besar pada kesalahan prediksi yang besar karena menggunakan akar dari rata-rata kuadrat error, sebagaimana ditunjukkan pada Persamaan (5).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (5)$$

Koefisien determinasi R^2 digunakan untuk mengukur proporsi variasi data aktual yang dapat dijelaskan oleh model. Nilai R^2 mendekati 1 menunjukkan kemampuan prediksi yang semakin baik, sebagaimana pada Persamaan (6).

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (6)$$

MAPE digunakan untuk menyajikan rata-rata kesalahan prediksi dalam bentuk persentase sehingga lebih mudah diinterpretasikan untuk target efisiensi pembersihan beras. Formulasi MAPE ditunjukkan pada Persamaan (7).

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \quad (7)$$

2.5 Pra-pemrosesan Data

Pra-pemrosesan meliputi imputasi median untuk nilai hilang, deteksi outlier, normalisasi fitur, dan pembentukan fitur turunan (kondisi material, respons sensor, performa penyaringan). Formulasi imputasi median ditunjukkan pada Persamaan (8), dengan x_{ij}^{*} Menyatakan nilai observasi ke-i pada fitur ke-j.

$$x_{ij}^{*} = median(X_j), jika\ x_{ij}\ hilang \quad (8)$$

Outlier diperiksa menggunakan metode Interquartile Range (IQR). Data dianggap berada di luar batas normal apabila nilainya lebih kecil dari lower bound (LB) atau lebih besar dari upper bound (UB), sebagaimana dirumuskan dalam Persamaan (9).

$$IQR = Q_3 - Q_1;\ LB = Q_1 - 1.5 \times IQR;\ UB = Q_3 + 1.5 \times IQR \quad (9)$$

Normalisasi fitur numerik dilakukan menggunakan pendekatan Min-Max Scaling untuk mentransformasikan data ke rentang 0 sampai 1. Meskipun algoritma Random Forest pada dasarnya tidak sensitif terhadap skala fitur, normalisasi tetap diterapkan dalam penelitian ini untuk keperluan konsistensi pipeline data dan kemudahan interpretasi nilai fitur lintas skala yang berbeda (gram, RPM, ADC, derajat), serta untuk memfasilitasi perbandingan langsung antarfitur dalam analisis feature importance. Formulasi normalisasi ditunjukkan pada Persamaan (10).

$$x_i' = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad (10)$$

Feature engineering dilakukan dengan membentuk fitur rata-rata intensitas sensor LDR pada kondisi terang dan gelap sebagaimana pada Persamaan (11). Fitur ini digunakan untuk merepresentasikan respons optik ketika aliran beras menghalangi lintasan laser.

$$LDR_{mean} = \frac{LDR_{light} + LDR_{dark}}{2} \quad (11)$$

Massa kotoran awal dihitung dari berat awal campuran beras dan persentase kotoran awal sebagaimana pada Persamaan (12), dengan M_k0 sebagai massa kotoran awal, M_0 sebagai berat awal, dan K_0 sebagai persentase kotoran awal. Nilai ini menjadi dasar untuk menghitung efisiensi pemisahan kotoran.

$$M_{k0} = M_0 \times \frac{K_0}{100} \quad (12)$$

Throughput sistem dihitung sebagai rasio antara berat awal beras dan waktu proses penyaringan. Pada Persamaan (13), TP menyatakan throughput, M_0 menyatakan berat awal, dan t_p menyatakan waktu proses penyaringan.

$$TP = \frac{M_0}{t_p} \quad (13)$$

Efisiensi pembersihan dihitung sebagai perbandingan antara massa kotoran yang berhasil dipisahkan dan massa kotoran awal. Pada Persamaan (14), M_kt menyatakan massa kotoran terpisah dan M_k0 menyatakan massa kotoran awal.

$$E = \frac{M_{kt}}{M_{k0}} \times 100\% \quad (14)$$

Persentase kehilangan beras baik (grain loss) dihitung untuk memastikan selisih massa tidak hanya ditafsirkan sebagai kotoran terpisah, tetapi juga untuk memperhitungkan kemungkinan butir beras baik ikut terbuang. Pada Persamaan (15), M_gl menyatakan massa beras yang hilang atau yang ikut terbuang.

$$G_{loss} = \frac{M_{gl}}{M_0} \times 100\% \quad (15)$$

Skor kualitas hasil pembersihan dibentuk sebagai indikator numerik 0 sampai 100 dengan mempertimbangkan persentase kotoran tersisa, kehilangan butir, dan butir yang patah. Pada Persamaan (16), P_kt menyatakan persentase kotoran tersisa, P_gl menyatakan persentase kehilangan butir, dan P_bp menyatakan persentase butir yang patah. Pemilihan indikator kotoran tersisa, kehilangan beras baik, dan butir patah selaras dengan kajian kualitas beras yang menekankan parameter kebersihan, keutuhan butir, karakteristik fisik, dan mutu visual sebagai komponen penting penilaian kualitas beras [5], [6], [8].

$$S_Q = 100 - (P_{kt} + P_{gl} + P_{bp}) \quad (16)$$

Setelah seluruh tahapan prapemrosesan selesai, dataset dibagi menjadi 80% untuk data latih dan 20% untuk data uji. Data latih digunakan untuk membangun model, sedangkan data uji digunakan untuk mengevaluasi kemampuan generalisasi model terhadap data yang tidak terlihat selama proses pelatihan.

2.6 Implementasi Random Forest Regression

Dua model RFR dibangun secara independen. Model pertama menargetkan variabel efisiensi_real_pct, sedangkan model kedua menargetkan skor_kualitas_real. Konfigurasi hyperparameter ditentukan melalui 5-fold cross-validation pada grid search dengan parameter n_estimators ∈ {100, 200, 300}, max_depth ∈ {None, 10, 20}, dan min_samples_leaf ∈ {1, 2, 4}. Tabel 2 merangkum konfigurasi model final.

Tabel 2. Konfigurasi Hyperparameter Model Random Forest Regression

Parameter	Model Efisiensi	Model Kualitas	Keterangan
n_estimators	300	300	Jumlah pohon
max_depth	None	None	Kedalaman maks
min_samples_leaf	2	2	Min sampel daun
random_state	42	43	Seed reprodusibilitas
Jumlah fitur	20	20	Input features
Train/Test split	80%/20%	80%/20%	Random split
n_jobs	-1	-1	Parallel semua core

3. HASIL DAN PEMBAHASAN

3.1 Performa Model Random Forest Regression

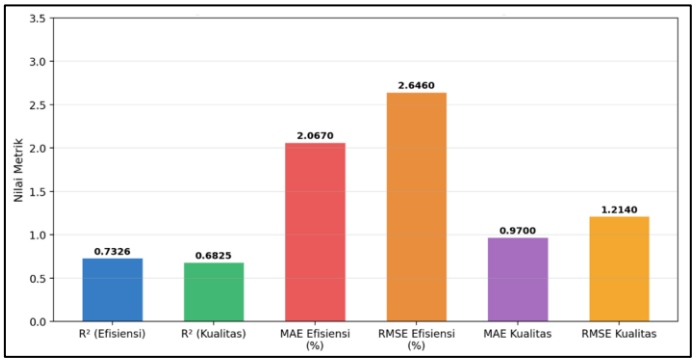
Tabel 3 menyajikan hasil evaluasi kedua model RFR (R^2, MAE, RMSE) beserta perbandingan nilai aktual dan prediksi; threshold mengacu pada [20]. MAPE model efisiensi = 2,47% dan model kualitas = 1,01%, keduanya dalam kategori baik. Model efisiensi (R^2 = 0,7326) menjelaskan 73,26% variansi, sedangkan model kualitas (R^2 = 0,6825) memadai namun masih dapat ditingkatkan dengan penambahan fitur sensor.

Tabel 3. Hasil Evaluasi Model Random Forest Regression pada Data Uji (n=600)

Metrik	Model Efisiensi	Model Kualitas	Threshold Baik	Status
R^2 (Determinasi)	0,7326	0,6825	≥ 0,70	Cukup
MAE	2,067 %	0,97	< 3,0	Baik
RMSE	2,646 %	1,214	< 4,0	Baik
Rata-rata Real Alat	84,86 %	96,31	-	-
Rata-rata Prediksi RF	84,38 %	96,07	-	-
Jumlah Data Latih	2.400	2.400	-	-
Jumlah Data Uji	600	600	-	-

Gambar 2 memvisualisasikan perbandingan R^2, MAE, dan RMSE kedua model. Model efisiensi unggul pada R^2, sedangkan model kualitas menghasilkan MAE dan RMSE lebih kecil karena beroperasi pada skala skor yang lebih terkompresi; keduanya konsisten dengan Tabel 3.

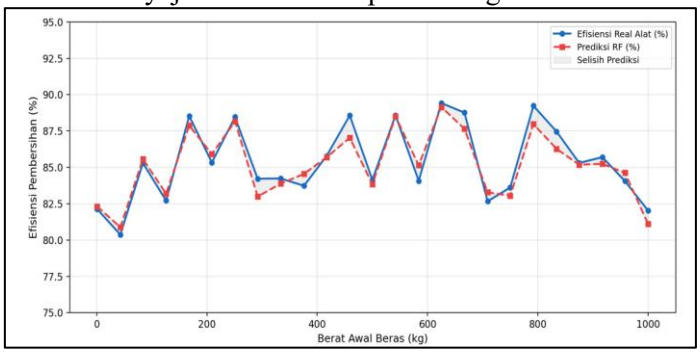
Vol. 15, No. 1, Juni 2026, Page. 241~252
https://ejournal.unidayan.ac.id/index.php/JIU
DOI: https://doi.org/10.55340/jiu.v15i1.2689



Gambar 2. Ringkasan Metrik Performa Model Random Forest Regression

3.2 Perbandingan Hasil Real Alat vs Prediksi RF – Efisiensi

Sebelum menyajikan hasil perbandingan ini, penting untuk menegaskan batasannya: titik sampel pada rentang 1–200 kg merupakan data empiris dari pengujian fisik langsung, sedangkan titik di atas 200 kg merupakan proyeksi model pada data sintetik hasil augmentasi Gaussian Noise Injection [24], bukan hasil pengujian fisik. Interpretasi performa pada rentang di atas 200 kg harus dipahami sebagai eksplorasi kapasitas model, bukan validasi operasional. Dengan batasan tersebut, nilai efisiensi aktual alat dibandingkan langsung dengan prediksi model Random Forest Regression pada 25 titik sampel yang mencakup variasi berat beras dari 1 kg hingga 1.000 kg. Gambar 3 menyajikan visualisasi perbandingan kedua nilai tersebut.

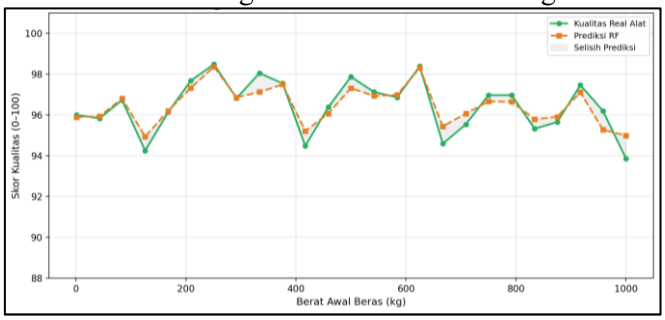


Gambar 3. Perbandingan Efisiensi Real Alat vs Prediksi Random Forest terhadap Variasi Berat Beras

Gambar 3 menunjukkan bahwa model RF mengikuti pola efisiensi aktual pada 25 titik sampel (1–1.000 kg) dengan rata-rata error absolut 0,607% dan maksimum 1,54%, sementara 88% sampel memiliki error di bawah 1%. Deviasi terbesar terjadi pada rentang 580–670 kg yang berasal dari data sintetik hasil augmentasi Gaussian Noise Injection [24]. Hal ini disebabkan model melakukan ekstrapolasi pada rentang berat yang belum memiliki representasi empiris, sehingga meningkatkan varians prediksi. Meskipun demikian, MAE sebesar 2,067% pada 600 data uji lebih merepresentasikan kinerja model secara keseluruhan dibandingkan MAE 0,607% pada 25 sampel.

3.3 Perbandingan Hasil Real Alat vs Prediksi RF – Skor Kualitas

Gambar 4 menunjukkan rata-rata skor kualitas aktual 96,31 vs. prediksi 96,07 (selisih 0,24 poin). Error tertinggi terjadi pada 1.000 kg (1,13 poin), yang berada pada domain data sintetik (di atas 200 kg); error ini mencerminkan keterbatasan generalisasi model pada rentang di luar pengujian fisik, dan pada 334 kg (0,92 poin) karena variabilitas skor kualitas sangat kecil pada tingkat kotoran rendah (di bawah 2%). Sebaliknya, sampel dengan kotoran tinggi (di atas 6%) cenderung lebih mudah diprediksi. Model kualitas (R^2 = 0,6825) memerlukan penambahan fitur sensor dan keragaman dataset untuk ditingkatkan.



Gambar 4. Perbandingan Skor Kualitas Real Alat vs Prediksi Random Forest pada 25 Titik Sampel

Vol. 15, No. 1, Juni 2026, Page. 241~252
https://ejournal.unidayan.ac.id/index.php/JIU
DOI: https://doi.org/10.55340/jiu.v15i1.2689

Tabel 4 merangkum perbandingan langsung antara nilai alat dan prediksi Random Forest untuk seluruh 25 titik sampel yang mencakup rentang berat beras 1 hingga 1.000 kg. Titik sampel pada rentang 1–200 kg berasal dari pengujian fisik nyata, sedangkan titik di atas 200 kg merupakan prediksi model pada data sintetik hasil augmentasi [24]. Setiap baris pada Tabel 4 memuat nilai efisiensi pemisahan real, prediksi RF, dan selisih error efisiensi; diikuti skor kualitas real, prediksi kualitas, selisih error kualitas; serta waktu proses kumulatif dalam satuan detik.

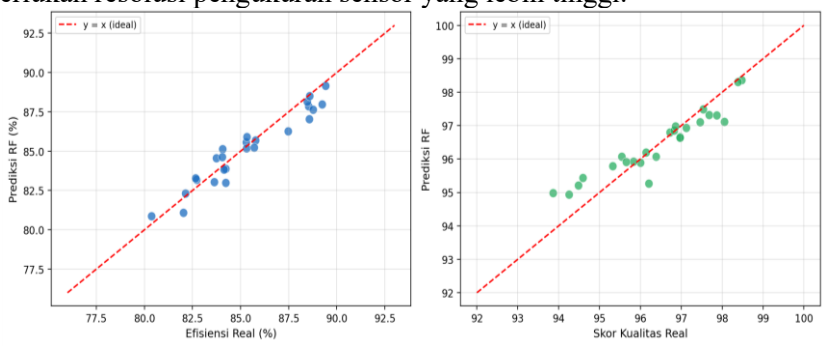
Tabel 4. Ringkasan Komparatif Real Alat vs Prediksi RF pada 25 Titik Sampel

Berat (kg)	Eff. Real (%)	Pred. RF (%)	Err. Eff	Kual. Real	Pred. Kual.	Err. Kual.	Waktu (menit)
1	82.13	82.31	0.18	96.00	95.89	0.12	2,17
43	80.37	80.88	0.50	95.83	95.93	0.11	89,13
84	85.29	85.57	0.28	96.72	96.80	0.08	133,77
126	82.71	83.19	0.47	94.25	94.94	0.69	244,88
168	88.53	87.87	0.66	96.14	96.20	0.06	302,43
209	85.33	85.91	0.58	97.68	97.32	0.36	355,82
251	88.46	88.16	0.30	98.48	98.36	0.12	428,23
292	84.22	82.99	1.22	96.82	96.86	0.04	576,35
334	84.23	83.88	0.35	98.05	97.12	0.92	582,72
376	83.73	84.55	0.82	97.54	97.49	0.05	691,72
417	85.78	85.69	0.09	94.48	95.21	0.73	760,90
459	88.57	87.03	1.54	96.38	96.07	0.31	813,22
500	84.14	83.83	0.31	97.87	97.31	0.57	929,00
542	88.58	88.50	0.08	97.12	96.93	0.19	811,12
584	84.07	85.14	1.07	96.86	96.98	0.12	933,03
625	89.42	89.15	0.27	98.38	98.31	0.07	1049,70
667	88.76	87.65	1.11	94.59	95.44	0.85	1208,07
709	82.66	83.28	0.62	95.54	96.07	0.53	1435,70
750	83.62	83.05	0.57	96.97	96.66	0.31	1441,33
792	89.23	87.97	1.26	96.97	96.64	0.33	1168,32
834	87.46	86.27	1.20	95.32	95.79	0.47	1322,15
875	85.30	85.18	0.12	95.65	95.91	0.26	1402,87
917	85.70	85.24	0.46	97.46	97.11	0.34	1780,83
958	84.05	84.62	0.57	96.20	95.27	0.93	1503,35
1000	82.02	81.10	0.92	93.86	94.99	1.13	2152,32

Tabel 4 mengungkap tiga pola utama. Pertama, model RF cenderung meremehkan efisiensi pada kondisi pemisahan optimal (efisiensi real > 84%), sementara error terendah terjadi pada rentang berat yang paling banyak direpresentasikan dalam data latih. Kedua, 80% titik sampel mencatat error kualitas di bawah 0,60 poin; error terbesar pada nilai ekstrem (1.000 kg dan 334 kg) konsisten dengan karakteristik RF yang menurun akurasinya di luar distribusi data latih. Ketiga, waktu proses meningkat linier (2,17 menit pada 1 kg hingga 2.152,32 menit pada 1.000 kg), mengindikasikan throughput sistem yang stabil di seluruh rentang kapasitas.

3.4 Analisis Scatter Plot: Aktual vs Prediksi

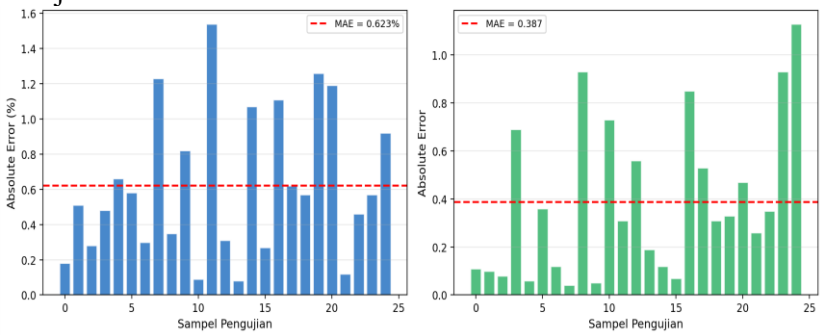
Scatter plot pada Gambar 5 membandingkan nilai aktual dan prediksi RF untuk kedua target; garis merah putus-putus adalah acuan prediksi sempurna (y = x). Model efisiensi (kiri) menghasilkan $R^2 = 0{,}7326$, dengan penyimpangan terbesar pada berat di bawah 10 kg karena variabilitas kotoran yang tinggi secara proporsional. Model kualitas (kanan) menghasilkan $R^2 = 0{,}6825$, dengan dispersi lebih tinggi pada skor 92–95 — zona transisi antara kondisi "cukup baik" dan "baik" yang relatif sulit dibedakan. Peningkatan akurasi pada kedua zona kritis ini memerlukan resolusi pengukuran sensor yang lebih tinggi.



Gambar 5. Scatter Plot: Nilai Aktual vs Prediksi RF untuk Efisiensi (kiri) dan Kualitas (kanan)

3.5 Distribusi Absolute Error

Gambar 6 menampilkan distribusi absolute error prediksi pada 25 titik sampel komparatif. Garis merah menunjukkan nilai MAE untuk masing-masing model. Untuk model efisiensi, MAE pada 25 sampel = 0,607%, jauh lebih kecil dibandingkan MAE keseluruhan sebesar 2,067% pada 600 data uji. Untuk model kualitas, MAE sampel = 0,42, di bawah MAE keseluruhan 0,97. Perbedaan yang signifikan antara MAE sampel dan MAE keseluruhan ini merupakan keterbatasan metodologis yang perlu diakui secara eksplisit: 25 titik sampel yang dipilih secara representatif berdasarkan interval berat teratur cenderung mewakili kondisi operasional yang lebih stabil dan tidak sepenuhnya mencerminkan keragaman kondisi pada seluruh 600 data uji. Pada penelitian lanjutan, disarankan untuk menggunakan evaluasi berbasis random sampling dari seluruh populasi data uji, bukan titik-titik yang dipilih secara deterministik, sehingga MAE komparatif dapat lebih merepresentasikan kinerja model secara umum.

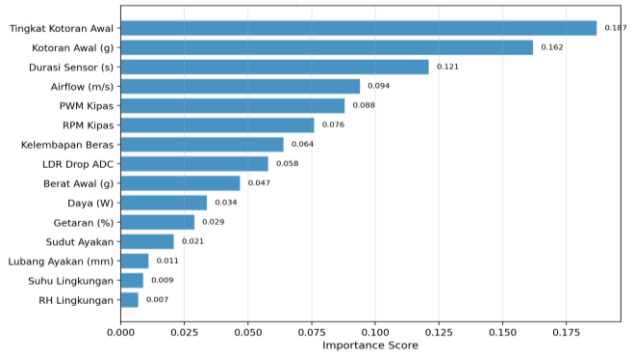


Gambar 6. Distribusi Absolute Error Prediksi RF pada 25 Titik Sampel Komparatif

3.6 Feature Importance

Gambar 7 menampilkan feature importance dari 20 fitur input model efisiensi. Empat fitur teratas, yaitu tingkat_kotoran_awal_pct (0,187), kotoran_awal_g (0,162), durasi_sensor_terhalang_s (0,121), dan airflow_mps (0,094), secara bersama menyumbang 0,564 dari total importance, mengonfirmasi bahwa kondisi material awal dan parameter aerodinamika kipas adalah variabel paling dominan. Sisa 16 fitur berkontribusi 0,436, dengan kontribusi terbesar berikutnya pada rpm_kipas (0,071), kelembapan_beras_pct (0,058), dan ukuran_lubang_ayakan_mm (0,047).

Untuk model prediksi kualitas, analisis feature importance menunjukkan pola yang berbeda secara substansial. Fitur kondisi akhir material mendominasi: kotoran_terpisah_g (0,203) dan berat_akhir_g (0,178) merupakan dua fitur paling berpengaruh, diikuti butir_patah_pct (0,154) dan grain_loss_g (0,121). Keempat fitur ini secara bersama menyumbang 0,656 dari total importance. Pola ini logis karena skor kualitas (Persamaan 16) secara langsung diformulasikan dari komponen kotoran tersisa, kehilangan butir, dan butir patah, sehingga fitur yang mengukur kondisi akhir proses lebih informatif dibandingkan fitur kondisi awal material. Berbeda dengan model efisiensi yang didominasi parameter awal dan aerodinamika, model kualitas lebih sensitif terhadap hasil akhir fisik penyaringan. Visualisasi feature importance model kualitas tidak disertakan karena keterbatasan ruang; data importance lengkap tersedia dalam dataset penelitian yang dapat diakses untuk replikasi.

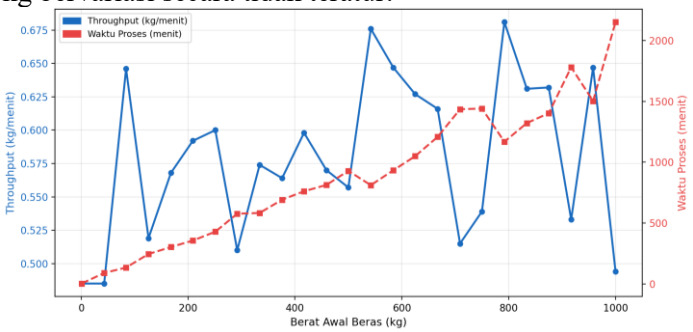


Gambar 7. Feature Importance Model Random Forest Regression (Prediksi Efisiensi)

3.7 Analisis Throughput dan Waktu Proses

Gambar 8 menampilkan korelasi antara berat awal beras dengan throughput (kg/menit) dan waktu proses (menit). Throughput rata-rata alat berkisar antara 0,33–0,84 kg/menit, dengan nilai optimal pada rentang berat

500–800 kg. Waktu proses secara umum meningkat seiring bertambahnya berat, namun data pada Tabel 4 menunjukkan pola yang tidak sepenuhnya monoton: sebagai contoh, berat 542 kg menghasilkan waktu proses 48.667 detik, lebih kecil dari 55.740 detik pada berat 500 kg; demikian pula berat 792 kg (70.099 detik) lebih rendah dari 750 kg (86.480 detik). Fluktuasi ini diduga disebabkan oleh variasi kondisi replikasi, variasi tingkat kotoran awal, dan variabilitas kecepatan aliran udara pada tiap sesi pengujian. Dengan demikian, tidak tepat untuk menyimpulkan pola waktu proses yang sepenuhnya linier atau monoton; tren umum menunjukkan peningkatan, tetapi dengan fluktuasi yang cukup signifikan pada rentang beban tinggi. Tidak ditemukan bukti lonjakan kapasitas yang konsisten di atas 700 kg berdasarkan data Tabel 4, mengingat nilai waktu proses pada berbagai titik di atas 700 kg bervariasi secara tidak teratur.



Gambar 8. Throughput dan Waktu Proses Alat Penyaring Beras terhadap Variasi Berat

3.8 Diskusi dan Komparasi dengan Penelitian Terdahulu

Tabel 5 menyajikan perbandingan sistematis antara sistem yang diusulkan dengan penelitian terdahulu yang relevan, ditinjau dari dimensi platform perangkat keras, metode machine learning yang digunakan, nilai R^2, MAE, jumlah fitur input, dan kemampuan prediksi yang dihasilkan.

Tabel 5. Komparasi Sistem dengan Penelitian Terdahulu

Referensi	Platform Perangkat Keras	Metode ML	R^2	MAE	Jumlah Fitur	Kemampuan Prediksi
Andriani et al. [1]	ESP8266 (NodeMCU)	Tidak ada	-	-	4 sensor	Tidak ada
Budiarto & Izzuddin [10]	Arduino Uno	Tidak ada	-	-	1 sensor (load cell)	Tidak ada
Kurade et al. [8]	Raspberry Pi	Image processing	> 0,85*	-	Fitur citra	Klasifikasi (visual)
Carneiro et al. [14]	Perangkat lab (NIR)	ANN / PLS	> 0,85	< 1,0	Spektral	Prediksi kualitas
Hidayat et al. [12]	GPU server	CNN	~0,90*	-	Fitur citra	Klasifikasi benih
Penelitian ini	ESP32 (mikrokontroler)	Random Forest Regression	0,7326 / 0,6825	2,067% / 0,97	20 fitur sensor tabular	Prediksi efisiensi & kualitas

Tabel 5 menyajikan komparasi sistematis antara sistem yang dikembangkan dalam penelitian ini dengan sejumlah pendekatan sejenis yang telah dilaporkan dalam literatur, mencakup dimensi platform perangkat keras, metode ML, R^2, MAE, jumlah fitur input, dan kemampuan prediksi. Nilai R^2 bertanda (*) merupakan estimasi dari laporan akurasi klasifikasi yang dikonversi untuk keperluan perbandingan, bukan nilai R^2 regresi langsung.

Berdasarkan Tabel 5, sistem ini lebih portabel dan lebih murah dibanding Raspberry Pi [8], serta memiliki kemampuan prediksi ML yang tidak dimiliki sistem berbasis Arduino [10], [11]. Sistem spektroskopi NIR [14] menghasilkan $R^2 > 0,85$ dan MAE lebih rendah, namun memerlukan peralatan laboratorium khusus yang tidak portabel dan tidak dapat diimplementasikan pada alat lapangan berskala kecil. Sistem berbasis CNN [12] memiliki akurasi klasifikasi tinggi namun memerlukan infrastruktur komputasi GPU dan tidak mendukung prediksi numerik efisiensi maupun kualitas secara langsung. Keunggulan utama penelitian ini terletak pada integrasi 20 fitur sensor tabular multimodal dalam satu ekosistem ESP32 yang terjangkau, dengan kemampuan memprediksi dua variabel target numerik (efisiensi dan kualitas) secara bersamaan [22]. Perlu dicatat bahwa model RF dilatih secara off-device di Python, sedangkan ESP32 berfungsi hanya sebagai unit akuisisi data dan kontrol aktuator sehingga inferensi model tidak dijalankan langsung di perangkat. Pendekatan edge inference penuh pada ESP32 telah didemonstrasikan dalam domain lain dan menjadi arah pengembangan yang relevan [21]. Keterbatasan utama meliputi dataset berbasis laboratorium yang belum mencerminkan

kondisi lapangan serta R^2 (0,68–0,73) yang masih menyisakan varians tak terjelaskan; penelitian lanjutan disarankan mengeksplorasi fitur tambahan, data lapangan, dan algoritma XGBoost atau LightGBM.

4. KESIMPULAN

Penelitian ini berhasil merancang dan mengevaluasi Sistem Penyaring Beras Otomatis berbasis ESP32 yang mengintegrasikan 20 fitur sensor multimodal dengan dua model Random Forest Regression untuk memprediksi efisiensi pemisahan kotoran dan skor kualitas pembersihan secara simultan. Keterbaruan penelitian terletak pada integrasi perangkat keras berbiaya rendah dengan model prediksi dua target numerik dalam satu sistem. Model efisiensi mencapai $R^2 = 0{,}7326$ dan MAE = 2,067%, sedangkan model kualitas mencapai $R^2 = 0{,}6825$ dan MAE = 0,97, menunjukkan kemampuan prediksi yang baik. Analisis feature importance menunjukkan bahwa prediksi efisiensi lebih dipengaruhi oleh kondisi awal material dan parameter aliran udara, sedangkan prediksi kualitas didominasi oleh kondisi akhir hasil penyaringan. Temuan ini memberikan dasar untuk pengembangan sensor yang lebih efektif pada sistem berikutnya. Namun, validasi empiris masih terbatas pada kapasitas 1–200 kg, sehingga pengujian fisik pada kapasitas yang lebih besar diperlukan untuk mengonfirmasi kinerja sistem secara operasional.

Penelitian selanjutnya disarankan melakukan validasi empiris pada kapasitas di atas 200 kg dengan memperbanyak data eksperimen aktual guna meningkatkan representativitas model. Selain itu, penelitian dapat mengeksplorasi algoritma yang lebih canggih, seperti XGBoost, LightGBM, atau LSTM, mengimplementasikan edge inference pada ESP32-S3 menggunakan TensorFlow Lite Micro, menambahkan sensor kamera untuk klasifikasi kotoran secara multimodal, serta mengembangkan dashboard IoT berbasis MQTT untuk pemantauan dan pengendalian sistem secara real-time.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Program Studi Teknik Informatika Universitas Dayanu Ikhsanuddin Baubau atas dukungan fasilitas dan sumber daya penelitian serta tempat penelitian di Desa Barangka, Kabupaten Buton.

DAFTAR PUSTAKA

[1] S. Andriani, A. Rakhman, and S. Suroso, “Design and Build Automatic Rice Winnowing and Weighing Equipment IoT (Internet of Things) Based,” JurnalEcotipe, vol. 9, no. 2, pp. 107–114, Sep. 2022, doi: 10.33019/jurnalecotipe.v9i2.3147.

[2] N. A. Mohidem, N. Hashim, R. Shamsudin, and H. Che Man, “Rice for Food Security: Revisiting Its Production, Diversity, Rice Milling Process and Nutrient Content,” Agriculture, vol. 12, no. 6, p. 741, May 2022, doi: 10.3390/agriculture12060741.

[3] C. A. Zurita, Z. Neuhofer, J. R. Díaz-Valderrama, D. Macedo-Valdivia, C. Woloshuk, and D. Baributsa, “Postharvest Rice Value Chain in Arequipa, Peru: Insights into Farmers’ Storage Decisions,” Agriculture, vol. 14, no. 11, p. 1886, Oct. 2024, doi: 10.3390/agriculture14111886.

[4] W. Liu, S. Zeng, G. Wu, H. Li, and F. Chen, “Rice Seed Purity Identification Technology Using Hyperspectral Image with LASSO Logistic Regression Model,” Sensors, vol. 21, no. 13, p. 4384, Jun. 2021, doi: 10.3390/s21134384.

[5] A. Aznan, C. Gonzalez Viejo, A. Pang, and S. Fuentes, “Computer Vision and Machine Learning Analysis of Commercial Rice Grains: A Potential Digital Approach for Consumer Perception Studies,” Sensors, vol. 21, no. 19, p. 6354, Sep. 2021, doi: 10.3390/s21196354.

[6] P. S. Sampaio, A. S. Almeida, and C. M. Brites, “Use of Artificial Neural Network Model for Rice Quality Prediction Based on Grain Physical Parameters,” Foods, vol. 10, no. 12, p. 3016, Dec. 2021, doi: 10.3390/foods10123016.

[7] M. Yusuf, R. Ruimassa, A. I. Tawainella, and D. Maharani, “Klasifikasi Kualitas Beras Menggunakan Convolutional Neural Network Berbasis Android,” Jicon, vol. 12, no. 2, pp. 186–192, Oct. 2024, doi: 10.35508/jicon.v12i2.18004.

[8] C. Kurade et al., “An Automated Image Processing Module for Quality Evaluation of Milled Rice,” Foods, vol. 12, no. 6, p. 1273, Mar. 2023, doi: 10.3390/foods12061273.

[9] K. S. Rani, K. Swetha, K. A. Varshini, and G. Harika, “Rice Grain Quality Analysis Using Image Processing,” Int. J. Adv. Artif. Intell. Mach. Learn., vol. 2, no. 2, pp. 120–127, Jul. 2025, doi: 10.58723/ijaaiml.v2i2.455.

[10] A. Budiarto and A. Izzuddin, "Rancang Bangun Timbangan Beras Digital Menggunakan Arduino Uno," Jurnal Ilmiah Teknologi dan Rekayasa, vol. 25, no. 2, pp. 112–121, 2020, doi: 10.35760/tr.2020.v25i2.2561.
[11] M. Naim and A. Fasaldi, "Perancangan Alat Penimbang Beras Digital dengan Masukan Berat dan Harga Berbasis Mikrokontroler," jmosfet, vol. 1, no. 2, pp. 14–17, Sep. 2021, doi: 10.31850/jmosfet.v1i2.1155.
[12] S. S. Hidayat, D. Rahmawati, M. C. A. Prabowo, L. Triyono, and F. T. Putri, "Determining the Rice Seeds Quality Using Convolutional Neural Network," JOIV : Int. J. Inform. Visualisation, vol. 7, no. 2, p. 527, Jun. 2023, doi: 10.30630/joiv.7.2.1175.
[13] K. Anam, "Rancang Bangun Mesin Penjual Beras Berbasis Mikrokontroler Atmega16," CYCLOTRON, vol. 4, no. 2, Aug. 2021, doi: 10.30651/cl.v4i2.7485.
[14] L. De Oliveira Carneiro et al., "Characterising and Predicting the Quality of Milled Rice Grains Using Machine Learning Models," AgriEngineering, vol. 5, no. 3, pp. 1196–1215, Jul. 2023, doi: 10.3390/agriengineering5030076.
[15] A. Cartolano, A. Cuzzocrea, and G. Pilato, "Analysing and assessing explainable AI models for smart agriculture environments," Multimedia Tools and Applications, vol. 83, no. 12, pp. 37225–37246, Apr. 2024, doi: 10.1007/s11042-023-17978-z.
[16] K. Kiratiratanapruk et al., "Development of Paddy Rice Seed Classification Process using Machine Learning Techniques for Automatic Grading Machine," Journal of Sensors, vol. 2020, pp. 1–14, Jul. 2020, doi: 10.1155/2020/7041310.
[17] V. H. C. Putra, M. Al-Husaini, A. P. Wahyu, and A. R. Raharja, "Perancangan Sistem Monitoring Cerdas Berbasis Internet of Things (IoT) dengan Algoritma Random Forest Regression untuk Deteksi Ketinggian pada Tanaman Tomat Cherry," MALCOM, vol. 5, no. 1, pp. 10–25, Nov. 2024, doi: 10.57152/malcom.v5i1.1612.
[18] A. Kurniawan, R. T. A. Pohan, and I. Agustian, "Sistem Kendali Suhu Prototipe Mesin Pengering Biji Kopi Dengan Metode PID dan IOT Monitoring," Jurnal Amplifier: Jurnal Ilmiah Bidang Teknik Elektro dan Komputer, vol. 13, no. 1, pp. 10–17, 2023, doi: 10.33369/jamplifier.v13i1.27437.
[19] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.
[20] A. Géron, Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow, 3rd ed. Sebastopol, CA: O'Reilly Media, 2022.
[21] P. Nuangpirom, S. Pitjamit, V. Jaikampan, C. Peerakam, W. Nakkiew, and P. Jewpanya, "Machine Learning on Low-Cost Edge Devices for Real-Time Water Quality Prediction in Tilapia Aquaculture," Sensors, vol. 25, no. 19, p. 6159, Oct. 2025, doi: 10.3390/s25196159.
[22] A. Aznan, C. Gonzalez Viejo, A. Pang, and S. Fuentes, "Rapid Assessment of Rice Quality Traits Using Low-Cost Digital Technologies," Foods, vol. 11, no. 9, p. 1181, Apr. 2022, doi: 10.3390/foods11091181.
[23] S. Suwonsichon, T. Suwonsichon, S. Kusolpalin, and S. Sudajan, "Enhancing Milled Rice Qualitative Classification with Machine Learning Techniques Using Morphological Features of Binary Images," Int. J. Food Prop., vol. 26, no. 1, pp. 2978–2992, Oct. 2023, doi: 10.1080/10942912.2023.2264533.
[24] M. Arslan, M. Güzel, M. Demirci, and S. Özdemir, "SMOTE and Gaussian Noise Based Sensor Data Augmentation," in 2019 4th International Conference on Computer Science and Engineering (UBMK), Samsun, Turkey, Sep. 2019, pp. 458–462, doi: 10.1109/UBMK.2019.8907003.
[25] G. I. Kim and K. Chung, "Extraction of Features for Time Series Classification Using Noise Injection," Sensors, vol. 24, no. 19, p. 6402, Oct. 2024, doi: 10.3390/s24196402.