

Implementasi Model MobileNetV2 pada ESP-32-CAM untuk Klasifikasi Botol Plastik

^{1*}Edo Ranov Anjasmara, ²Farniwati Fattah, ³Andi Widya Mufila Gaffar

^{1,2,3}Fakultas Ilmu Komputer, Universitas Muslim Indonesia, Makassar, Indonesia,

*Korespondensi: 13020220010@student.umi.ac.id

Submit : 09 Mei 2026 | Diterima : 04 Jun 2026 | Terbit : 07 Jun 2026

ABSTRACT

Waste management, particularly in sorting plastic bottle and non-bottle waste, remains a challenge in supporting effective recycling systems. This study aims to implement the MobileNetV2 model on the ESP32-CAM device using the Edge Impulse platform and to evaluate the model's performance in real-time object classification. The dataset consists of 200 images containing bottle and non-bottle objects with variations in lighting conditions, shooting angles, and backgrounds. The model was trained using Edge Impulse and then converted into TensorFlow Lite format for deployment on the ESP32-CAM device. The training results show that the model achieved high performance with an accuracy of 92.50%, supported by an AUC of 0.97, precision of 0.98, recall of 0.97, and F1-score of 0.97. Based on the simplified confusion matrix with visual verification, the model achieved 100% accuracy in detecting bottle objects and 95% accuracy for non-bottle objects, with a 5% misclassification rate. However, during real-world implementation on the ESP32-CAM device, the model's performance decreased to 65.7% accuracy due to differences between training and real-world conditions as well as hardware limitations. Despite this, the system successfully performed real-time image classification on an embedded device. This study demonstrates that the edge artificial intelligence approach using MobileNetV2 and Edge Impulse can be effectively applied to resource-constrained devices, although improvements are still needed in terms of model generalization and dataset diversity.

Keywords: *edge artificial intelligence, Edge Impulse, ESP32-CAM, image classification, MobileNetV2.*

ABSTRAK

Permasalahan dalam pengelolaan sampah, khususnya pada proses pemilahan antara sampah botol plastik dan non-botol, masih menjadi tantangan dalam mendukung sistem daur ulang yang efektif. Penelitian ini bertujuan untuk mengimplementasikan model MobileNetV2 pada perangkat ESP32-CAM dengan memanfaatkan platform Edge Impulse, serta mengevaluasi performa model dalam melakukan klasifikasi objek secara real-time. Dataset yang digunakan terdiri dari 200 citra yang mencakup objek botol dan non-botol dengan variasi kondisi pencahayaan, sudut pengambilan gambar, dan latar belakang. Proses pelatihan model dilakukan menggunakan platform Edge Impulse, kemudian model dikonversi ke format TensorFlow Lite untuk diimplementasikan pada perangkat ESP32-CAM. Hasil pelatihan menunjukkan bahwa model memiliki performa yang sangat baik dengan akurasi sebesar 92,50%, serta didukung oleh nilai AUC sebesar 0,97, precision sebesar 0,98, recall sebesar 0,97, dan F1-score sebesar 0,97. Berdasarkan hasil confusion matrix yang telah disederhanakan melalui verifikasi visual, model mampu mengenali objek botol dengan akurasi 100% dan objek non-botol sebesar 95%, dengan kesalahan klasifikasi sebesar 5%. Namun, pada tahap implementasi di perangkat ESP32-CAM, performa model mengalami penurunan dengan akurasi sebesar 65,7% akibat perbedaan kondisi antara data pelatihan dan kondisi nyata serta keterbatasan perangkat. Meskipun demikian, sistem yang dikembangkan berhasil melakukan klasifikasi citra secara real-time pada perangkat embedded. Penelitian ini menunjukkan bahwa pendekatan edge artificial intelligence menggunakan MobileNetV2 dan Edge Impulse dapat diterapkan pada perangkat dengan sumber daya terbatas, meskipun masih memerlukan peningkatan dalam hal generalisasi model dan variasi dataset.

Kata Kunci: edge artificial intelligence, Edge Impulse, ESP32-CAM, klasifikasi citra, MobileNetV2.

PENDAHULUAN

Permasalahan dalam pengelolaan sampah, khususnya pada proses pemilahan antara sampah botol plastik dan non-botol, masih menjadi tantangan dalam mendukung sistem daur ulang yang efektif. Pemilahan secara manual cenderung kurang efisien karena memerlukan waktu yang lama serta berpotensi menimbulkan kesalahan akibat keterbatasan manusia dalam mengenali jenis sampah secara konsisten. Oleh karena itu, diperlukan sistem otomatis yang mampu melakukan klasifikasi objek secara cepat dan akurat. Perkembangan teknologi *computer vision* dan *deep learning*, khususnya metode *convolutional neural network* (CNN), telah memberikan solusi dalam klasifikasi citra dengan kemampuan mengekstraksi fitur secara otomatis serta menghasilkan performa tinggi (Chen et al., 2021; Wasim Khan, 2024). Penerapan CNN memungkinkan proses identifikasi objek botol dan non-botol dilakukan lebih efisien dan dapat berjalan secara real-time. Seiring berkembangnya berbagai arsitektur CNN, diperlukan pemilihan model yang tepat agar sistem dapat bekerja optimal, terutama pada perangkat dengan keterbatasan sumber daya.

Dalam pengembangan model pada perangkat terbatas, digunakan arsitektur CNN ringan dengan efisiensi komputasi tinggi. Penelitian ini membandingkan tiga model, yaitu MobileNetV2, ShuffleNet, dan SqueezeNet, yang dirancang untuk perangkat embedded dengan ukuran model kecil namun tetap memiliki performa yang baik. Hasil pengujian menunjukkan bahwa MobileNetV2 memiliki performa terbaik dengan akurasi 86% dan F1-score 79%, dibandingkan ShuffleNet (84%) dan SqueezeNet (82%) (Jurnal Andika). Hal ini menunjukkan bahwa MobileNetV2 lebih unggul dalam mengekstraksi fitur citra serta menghasilkan klasifikasi yang lebih seimbang. Keunggulan ini didukung oleh arsitektur *depthwise separable convolution* dan *inverted residual* yang mampu mengurangi kompleksitas komputasi tanpa menurunkan kualitas pembelajaran. Oleh karena itu, MobileNetV2 dipilih sebagai model yang paling sesuai untuk diimplementasikan pada perangkat seperti ESP32-CAM. Pemilihan model yang efisien sangat penting dalam sistem berbasis Internet of Things (IoT), karena perangkat dituntut mampu melakukan pemrosesan secara mandiri dengan sumber daya terbatas.

Seiring perkembangan IoT, konsep *edge computing* semakin banyak digunakan untuk memungkinkan pemrosesan data dilakukan langsung pada perangkat tanpa bergantung pada server eksternal. Pendekatan ini dapat mengurangi latensi, meningkatkan efisiensi, serta mendukung sistem real-time pada perangkat dengan keterbatasan sumber daya (Zhou et al., 2019). Dalam bidang *computer vision*, perangkat seperti ESP32-CAM mulai dimanfaatkan untuk menjalankan model *deep learning* secara langsung di sisi edge, sehingga proses klasifikasi dapat dilakukan lebih cepat tanpa memerlukan komputasi berbasis cloud (Sarker, 2021). Untuk mendukung hal tersebut, dibutuhkan platform yang mampu mengintegrasikan proses pengembangan model, mulai dari pengumpulan data, preprocessing, pelatihan, hingga deployment ke perangkat edge. Salah satu platform yang banyak digunakan adalah Edge Impulse, yang menyediakan lingkungan pengembangan terpadu untuk implementasi *embedded machine learning* pada perangkat IoT dengan sumber daya terbatas.

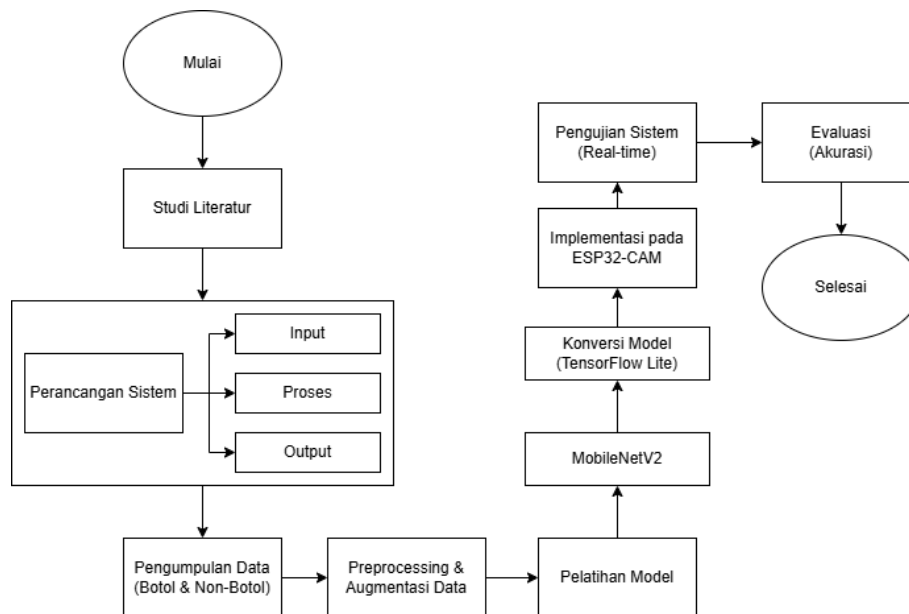
ESP32-CAM merupakan perangkat embedded berbasis mikrokontroler yang dilengkapi kamera dan konektivitas Wi-Fi, sehingga banyak digunakan dalam aplikasi IoT. Perangkat ini memiliki keunggulan dari segi biaya yang terjangkau, konsumsi daya rendah, dan ukuran yang ringkas, sehingga cocok untuk sistem berbasis *edge computing*. Namun, perangkat ini memiliki keterbatasan pada memori dan kemampuan komputasi, sehingga tidak semua model *deep learning* dapat dijalankan secara optimal (Banbury et al., 2021). Oleh karena itu, diperlukan model yang ringan seperti MobileNetV2 agar tetap dapat diimplementasikan tanpa mengurangi performa secara signifikan. Dengan dukungan Edge Impulse, proses optimasi dan deployment model menjadi lebih mudah, sehingga model dapat langsung dijalankan pada perangkat untuk melakukan klasifikasi citra secara real-time (Sarker, 2021). Kombinasi ESP32-CAM, MobileNetV2, dan Edge Impulse menjadi solusi potensial dalam pengembangan sistem klasifikasi citra berbasis *edge artificial intelligence* (Deng, Li, Han, Shi, & Xie, 2020).

Penelitian ini bertujuan untuk mengimplementasikan model MobileNetV2 pada ESP32-CAM dengan memanfaatkan Edge Impulse sebagai platform pengembangan dan deployment. Selain itu, penelitian ini juga mengevaluasi performa model dalam mengklasifikasikan objek botol dan non-botol secara real-time. Metode penelitian meliputi perancangan sistem, pengumpulan data, preprocessing, serta pelatihan model menggunakan MobileNetV2. Model

kemudian diimplementasikan pada ESP32-CAM dan diuji untuk mengetahui kinerja sistem. Dataset yang digunakan terdiri dari citra botol dan non-botol dengan variasi kondisi lingkungan untuk meningkatkan generalisasi model. Hasil penelitian menunjukkan bahwa model dapat diimplementasikan dengan baik dan mampu melakukan klasifikasi secara real-time, meskipun terjadi penurunan performa pada kondisi nyata. Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan sistem berbasis *edge artificial intelligence* serta menjadi referensi untuk penelitian selanjutnya.

METODE PENELITIAN

Penelitian ini menggunakan metode eksperimen dengan pendekatan rekayasa sistem untuk merancang, mengimplementasikan, dan mengevaluasi sistem klasifikasi objek berbasis ESP32-CAM. Metode ini dipilih karena memungkinkan pengujian langsung terhadap performa model pada perangkat dengan keterbatasan sumber daya komputasi, khususnya dalam klasifikasi objek secara real-time. Alur metodologi penelitian dapat di lihat pada Gambar 1.



Gambar 1 Research Design Flowchart

Gambar tersebut merepresentasikan alur *research design* yang digunakan dalam penelitian ini, yang mencakup tahapan-tahapan utama dalam pengembangan sistem mulai dari studi literatur hingga analisis data. Setiap tahapan memiliki peran penting dalam memastikan proses penelitian berjalan secara sistematis dan terarah. Oleh karena itu, metode penelitian pada studi ini dijelaskan dalam beberapa tahap sebagai berikut.

Studi Literatur

Tahap pertama adalah studi literatur. Pada tahap ini dilakukan kajian terhadap perkembangan *computer vision*, *deep learning*, serta implementasi *convolutional neural network* (CNN) pada perangkat *embedded*. Studi literatur juga mencakup penggunaan model *lightweight* seperti MobileNetV2 yang dirancang untuk perangkat *edge* karena memiliki efisiensi komputasi tinggi dan ukuran model yang kecil (Raza, Abidi, Masduzzaman, & Shin, 2025).

Perancangan Sistem

Tahap kedua adalah perancangan sistem, di mana sistem yang dikembangkan dibagi menjadi tiga bagian utama, yaitu input, proses, dan output. Bagian input berupa data citra objek yang diperoleh dari kamera ESP32-CAM, kamera handphone, serta dataset tambahan dari Kaggle yang terdiri dari kategori botol dan non-botol. Pada bagian proses dilakukan tahapan pengolahan data yang meliputi preprocessing, augmentasi, pelatihan model menggunakan MobileNetV2, konversi ke TensorFlow Lite, serta implementasi pada ESP32-CAM menggunakan konsep edge computing. Penerapan edge computing memungkinkan proses inferensi dilakukan

langsung pada perangkat ESP32 sehingga mengurangi latensi dan ketergantungan terhadap server eksternal (Chang, Wu, & Lin, 2025). Selain itu, penggunaan Convolutional Neural Network (CNN) pada perangkat embedded memerlukan model yang ringan agar dapat berjalan optimal pada keterbatasan sumber daya (Lim et al., 2022). Implementasi CNN berbasis ESP32-CAM juga telah menunjukkan kemampuan klasifikasi citra secara real-time dengan efisiensi komputasi yang tinggi (Ben Aicha, 2025). Sistem kemudian diuji secara real-time untuk memperoleh hasil klasifikasi. Adapun bagian output berupa hasil klasifikasi objek dalam bentuk label “botol” atau “non-botol” yang ditampilkan secara real-time dan digunakan sebagai dasar dalam analisis kinerja sistem. Penggunaan model machine learning pada perangkat embedded seperti microcontroller juga memungkinkan evaluasi performa dilakukan secara langsung pada perangkat dengan tetap mempertahankan akurasi yang baik (Diab & Rodriguez-Villegas, 2022).

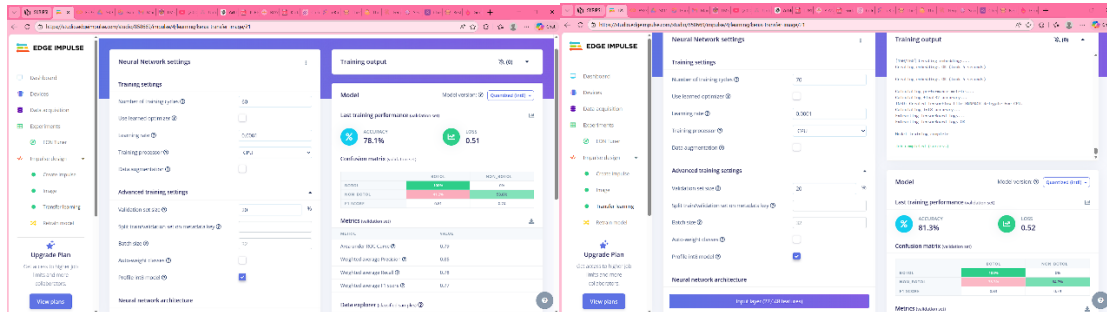
Pengumpulan Data dan Preprocessing

Tahap ketiga adalah pengumpulan data. Dataset citra diperoleh melalui pengambilan gambar objek botol dan non-botol menggunakan ESP32-CAM, kamera handphone dengan ukuran kamera 64 Megapixel, serta dataset publik dari Kaggle. Pada penelitian ini digunakan total 200 data citra yang terdiri dari 100 citra botol plastik dan 100 citra non-botol plastik. Untuk meningkatkan performa model, digunakan pendekatan variasi data melalui perubahan kondisi citra yang meliputi pencahayaan, sudut pengambilan gambar, dan latar belakang. Variasi pencahayaan pada dataset terdiri dari pencahayaan terang sebanyak 50 data, pencahayaan normal sebanyak 98 data, dan pencahayaan gelap sebanyak 52 data. Pencahayaan terang didefinisikan sebagai kondisi pengambilan gambar di luar ruangan pada siang hari atau di dalam ruangan dengan bantuan pencahayaan lampu yang cukup intens. Pencahayaan normal merupakan kondisi pengambilan gambar di dalam ruangan pada siang hari tanpa paparan langsung cahaya matahari dan tanpa bantuan lampu tambahan, sehingga menghasilkan intensitas cahaya yang relatif stabil dan tidak terlalu kontras. Sementara itu, pencahayaan gelap merupakan kondisi pengambilan gambar di dalam ruangan pada siang hari dengan lampu dimatikan serta meminimalkan masuknya cahaya dari luar ruangan, sehingga menghasilkan citra dengan intensitas cahaya rendah. Selain itu, variasi sudut pengambilan gambar dilakukan dari beberapa perspektif, yaitu sudut depan (0°) sebanyak 194 data, sudut miring ($\pm 30^\circ$ – 45°) sebanyak 4 data, dan sudut samping ($\pm 60^\circ$ – 90°) sebanyak 2 data. Variasi latar belakang juga digunakan untuk meningkatkan kompleksitas data, yang terdiri dari latar terang sebanyak 76 data, latar netral/ sederhana sebanyak 82 data, dan latar gelap sebanyak 42 data. Variasi ini mencakup kondisi latar polos, latar dengan sedikit objek, hingga latar kompleks dengan banyak elemen visual. Pendekatan variasi data tersebut bertujuan untuk meningkatkan kemampuan generalisasi model dalam mengenali objek pada berbagai kondisi lingkungan dan terbukti efektif dalam klasifikasi citra berbasis deep learning (Yong, Ma, Sun, & Du, 2023). Akuisisi citra menggunakan perangkat ESP32-CAM dengan berbagai skenario objek serta variasi jumlah data juga menunjukkan bahwa keberagaman data berperan penting dalam meningkatkan performa model klasifikasi berbasis CNN (Gaffar, Widyawati, Fahmi, & Ulfiani, 2025).

Sebelum digunakan dalam proses pelatihan, data citra terlebih dahulu melalui tahap preprocessing untuk memastikan keseragaman dan kualitas input model. Proses resizing dilakukan dengan mengubah seluruh citra menjadi ukuran 96×96 piksel agar sesuai dengan spesifikasi input model MobileNetV2 yang digunakan pada perangkat embedded. Selain itu, dilakukan normalisasi nilai piksel dengan mengubah rentang nilai dari 0–255 menjadi 0–1 melalui pembagian setiap nilai piksel dengan 255, sehingga proses pelatihan menjadi lebih stabil dan mempercepat konvergensi model. Untuk meningkatkan variasi data, diterapkan teknik data augmentation yang meliputi rotasi dengan rentang sudut $\pm 15^\circ$ hingga $\pm 30^\circ$, horizontal flipping untuk mensimulasikan perubahan orientasi objek, serta penyesuaian pencahayaan melalui peningkatan dan penurunan intensitas cahaya (brightness adjustment) pada kisaran tertentu. Selain itu, dilakukan juga variasi kontras ringan untuk mensimulasikan kondisi lingkungan yang berbeda. Teknik augmentasi ini bertujuan untuk meningkatkan kemampuan model dalam mengenali objek pada berbagai kondisi nyata, terutama yang berkaitan dengan perubahan sudut pandang dan pencahayaan yang telah terdapat pada dataset. Setelah tahap *preprocessing*, dataset kemudian dibagi menjadi dua bagian, yaitu data pelatihan sebanyak 160 sampel dan data pengujian sebanyak 40 sampel. Pembagian ini dilakukan untuk memastikan bahwa model dapat dilatih secara optimal sekaligus diuji pada data yang tidak pernah dilihat sebelumnya, sehingga mampu mengukur kemampuan generalisasi model secara objektif (Zahirah, Purnawansyah, Kurniati, & Darwis, 2024).

Pelatihan Model

Tahap keempat adalah pelatihan model menggunakan platform Edge Impulse dengan arsitektur MobileNetV2. Model ini dipilih karena efisiensi komputasi tinggi dan sesuai untuk perangkat *embedded*. Sebelum menentukan parameter terbaik, dilakukan beberapa percobaan pelatihan dengan variasi jumlah epoch untuk melihat pengaruhnya terhadap akurasi dan loss model. Hasil pengujian tersebut ditunjukkan pada Gambar 2.



Gambar 2 Pengujian Epoch

Berdasarkan hasil eksperimen yang dilakukan, jumlah epoch yang digunakan dalam pelatihan ditetapkan sebanyak 70. Pemilihan nilai ini didasarkan pada perbandingan dengan 60 epoch, di mana 70 epoch menghasilkan akurasi sebesar 81,3% dengan loss 0,52, sedangkan 60 epoch menghasilkan akurasi 78,1% dengan loss 0,51. Meskipun loss sedikit lebih tinggi, peningkatan akurasi menunjukkan pembelajaran model yang lebih optimal. Pelatihan menggunakan learning rate sebesar 0,0001 untuk menjaga stabilitas konvergensi. Proses pelatihan meliputi *feature extraction* dan *classification*, serta validasi model untuk mencegah *overfitting*. Model kemudian dikonversi ke format TensorFlow Lite untuk implementasi pada ESP32-CAM. Model MobileNetV2 digunakan karena memiliki arsitektur efisien berbasis *depthwise separable convolution* yang mampu mengurangi kompleksitas komputasi tanpa mengorbankan akurasi secara signifikan. Penelitian terbaru menunjukkan bahwa MobileNetV2 efektif digunakan dalam klasifikasi sampah berbasis citra dengan performa yang baik (Yong et al., 2023). Model dilatih menggunakan framework TensorFlow dan kemudian dikonversi ke format TensorFlow Lite agar dapat dijalankan pada perangkat *embedded*.

Implementasi Sistem

Tahap kelima adalah implementasi sistem. Model yang telah dilatih diintegrasikan ke dalam sistem ESP32-CAM. Proses inferensi dilakukan secara *real-time* menggunakan pendekatan *edge computing* sehingga tidak bergantung pada server eksternal. Pendekatan ini meningkatkan efisiensi sistem serta mengurangi latensi dalam proses klasifikasi (Kong et al., 2023).

Pengujian Sistem

Tahap keenam adalah pengujian sistem dengan memberikan berbagai objek ke kamera ESP32-CAM. Proses dimulai dari pengambilan citra, preprocessing (resizing dan normalisasi), hingga inferensi menggunakan model MobileNetV2 dalam format TensorFlow Lite. Model kemudian mengekstraksi fitur citra dan menghasilkan output klasifikasi berupa label "botol" atau "non-botol" secara *real-time*. Seluruh proses dilakukan langsung pada perangkat (*on-device inference*) tanpa koneksi ke server.

Evaluasi Sistem

Tahap ketujuh adalah evaluasi sistem dengan membandingkan hasil klasifikasi terhadap kondisi sebenarnya. Evaluasi dilakukan menggunakan metrik akurasi serta confusion matrix yang meliputi true positive, false positive, true negative, dan false negative (Zheng, Wang, Duan, Pang, & Tan, 2023). Tahap ini bertujuan untuk mengukur tingkat keandalan sistem sebagai alat berbasis IoT dalam melakukan klasifikasi objek secara *real-time*.

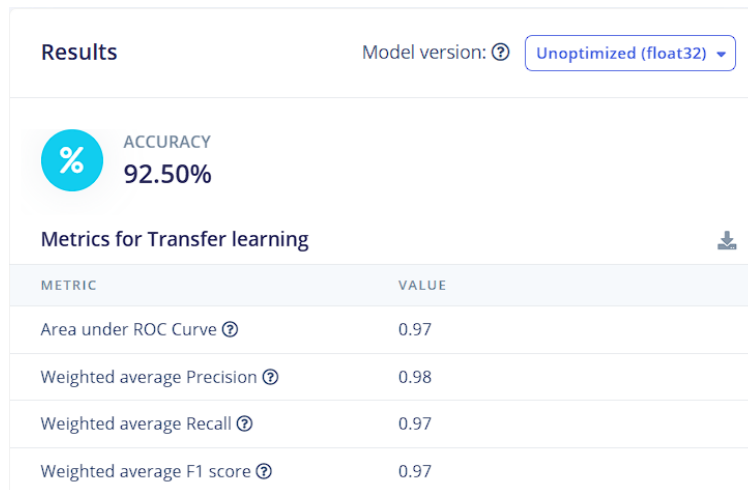
HASIL DAN PEMBAHASAN

Hasil penelitian ini diperoleh melalui tahapan implementasi, pengujian, evaluasi, dan analisis terhadap sistem klasifikasi objek botol dan non-botol berbasis ESP32-CAM. Sistem

yang dikembangkan mengintegrasikan model MobileNetV2 ke dalam perangkat *embedded* sehingga mampu melakukan klasifikasi citra secara *real-time*. Pada tahap implementasi, model yang telah dikonversi ke dalam format TensorFlow Lite berhasil dijalankan pada perangkat ESP32-CAM. Sistem mampu menangkap citra objek secara langsung melalui kamera dan memprosesnya tanpa memerlukan koneksi ke server eksternal. Hal ini menunjukkan bahwa sistem telah berhasil menerapkan konsep *edge computing*, di mana proses inferensi dilakukan langsung pada perangkat

Hasil Pelatihan Model pada Edge Impulse

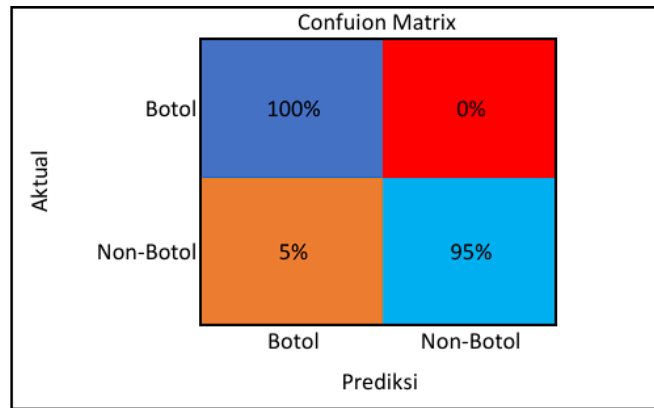
Hasil pelatihan model pada platform Edge Impulse ditunjukkan pada Gambar 1, yang menampilkan nilai akurasi, *loss*, serta confusion matrix pada data validasi.



Gambar 3 Hasil performa model pada tahap pelatihan di Edge Impulse

Berdasarkan Gambar 3, model memperoleh akurasi sebesar 92,50% pada tahap pelatihan menggunakan transfer learning. Selain itu, nilai Area Under ROC Curve (AUC) sebesar 0,97 menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam membedakan antara kelas botol dan non-botol. Nilai weighted average precision sebesar 0,98 serta weighted average recall sebesar 0,97 menunjukkan bahwa model mampu melakukan klasifikasi dengan tingkat ketepatan dan sensitivitas yang tinggi. Selain itu, nilai weighted average F1-score sebesar 0,97 menunjukkan bahwa model memiliki keseimbangan yang sangat baik antara precision dan recall, sehingga dapat disimpulkan bahwa model yang dihasilkan memiliki performa yang optimal dalam proses klasifikasi citra.

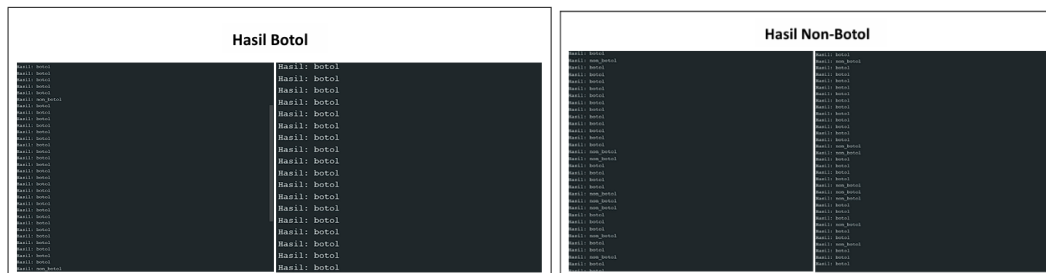
Dilakukan penyederhanaan menjadi dua kelas dengan mengeliminasi kelas *uncertain*. Pada penelitian ini, prediksi *uncertain* tidak langsung dianggap sebagai kesalahan klasifikasi, melainkan dilakukan verifikasi ulang secara visual terhadap data. Berdasarkan hasil pengamatan tersebut, sebagian besar data yang termasuk dalam kategori *uncertain* merupakan objek non-botol, sehingga dalam analisis ini dimasukkan ke dalam kelas non-botol. Hasil evaluasi menunjukkan bahwa model mampu mengenali objek botol dengan sangat baik dengan tingkat akurasi mencapai 100%. Selain itu, setelah dilakukan penyesuaian berdasarkan verifikasi visual, kelas non-botol menunjukkan tingkat klasifikasi sebesar 95%, di mana sebagian besar data non-botol berhasil dikenali dengan benar. Namun, masih terdapat kesalahan klasifikasi sebesar 5%, yaitu satu data non-botol yang diklasifikasikan sebagai botol. Munculnya prediksi *uncertain* pada tahap awal menunjukkan bahwa model masih memiliki tingkat kepercayaan yang rendah terhadap beberapa objek non-botol, terutama yang memiliki kemiripan visual dengan botol. Dengan demikian, meskipun performa model tergolong tinggi, masih diperlukan peningkatan pada aspek *confidence* agar sistem dapat memberikan hasil klasifikasi yang lebih konsisten secara otomatis tanpa memerlukan verifikasi tambahan. Gambar penyederhanaan *confusion matrix* menjadi dua kelas dapat dilihat pada Gambar 4.



Gambar 4 Confusion Matrix Hasil Klasifikasi Setelah Penyederhanaan Kelas

Hasil Implementasi dan Pengujian Sistem

Pada tahap pengujian, sistem diuji menggunakan 102 data, yang terdiri dari 52 data botol plastik dan 50 data non-botol plastik, dengan variasi kondisi lingkungan seperti pencahayaan, sudut pengambilan gambar, dan latar belakang. Pengujian dilakukan secara langsung untuk melihat kemampuan sistem dalam mendeteksi objek pada kondisi nyata. Proses deteksi dimulai ketika kamera ESP32-CAM menangkap citra objek. Citra tersebut kemudian melalui tahap *preprocessing* berupa *resizing* menjadi 96x96 piksel dan normalisasi nilai piksel. Selanjutnya, citra diproses oleh model MobileNetV2 untuk melakukan ekstraksi fitur dan klasifikasi objek. Hasil dari proses ini ditampilkan dalam bentuk label "botol" atau "non-botol" secara *real-time*. Untuk memperjelas proses kerja sistem, ditampilkan beberapa hasil pengujian berupa hasil klasifikasi yang dihasilkan secara *real-time*, seperti ditunjukkan pada Gambar 5.



Gambar 5 Tampilan hasil klasifikasi objek pada sistem ESP32-CAM

Performa Model

Evaluasi performa model dilakukan pada dua tahap, yaitu tahap pelatihan menggunakan Edge Impulse dan tahap implementasi pada perangkat ESP32-CAM. Hasil dari masing-masing tahap disajikan secara terpisah dalam tabel. Hasil pelatihan di Edge Impulse disajikan pada Tabel 2, sedangkan hasil pengujian pada perangkat ESP32-CAM disajikan pada Tabel 3.

Tabel 2 Performa Model pada Edge Impulse

No	Kategori	Terdeteksi	Presentase
1	True Positive (TP)	20	(100%)
2	False Negative (FN)	0	(0%)
3	False Positive (FP)	1	(5%)
4	True Negative (TN)	19	(90%)
5	Total Data	40	

Berdasarkan Tabel 2, nilai *true positive* (TP) sebesar 20 (100%) menunjukkan bahwa seluruh data botol berhasil diklasifikasikan dengan benar oleh model. Nilai *false negative* (FN) sebesar 0 (0%) menunjukkan bahwa tidak terdapat data botol yang salah diklasifikasikan sebagai non-botol. Pada kelas non-botol, nilai *true negative* (TN) sebesar 19 (95%) menunjukkan bahwa hampir seluruh data non-botol berhasil dikenali dengan benar. Namun, masih terdapat nilai *false positive* (FP) sebesar 1 (5%), yang menunjukkan bahwa sebagian kecil data non-botol diklasifikasikan sebagai botol.

Tabel 3 Performa Model pada ESP32-CAM

No	Kategori	Terdeteksi	Presentase Benar	Presentase Salah
1	Botol	52 Botol	50 (96,2%)	2(3,8%)
2	Non-Botol	50 Non-Botol	33 (66,0%)	17 (34,0%)
3	Total Data		102	

Berdasarkan Tabel 2, dari total 52 data botol, sebanyak 50 data berhasil diklasifikasikan dengan benar sebagai botol (96,2%), sedangkan 2 data lainnya diklasifikasikan sebagai non-botol (3,8%). Hal ini menunjukkan bahwa sistem memiliki kemampuan yang sangat baik dalam mengenali objek botol. Namun, pada kelas non-botol, dari total 50 data, hanya 17 data yang berhasil diklasifikasikan dengan benar sebagai non-botol (34,0%), sementara 33 data lainnya justru diklasifikasikan sebagai botol (66,0%). Hal ini menunjukkan bahwa sistem masih mengalami kesulitan dalam membedakan objek non-botol, terutama yang memiliki kemiripan visual dengan botol. Dengan demikian, dapat disimpulkan bahwa sistem cenderung mengklasifikasikan objek sebagai botol, yang menunjukkan adanya bias terhadap kelas botol. Kondisi ini menjadi penyebab utama tingginya kesalahan klasifikasi pada objek non-botol.

Pembahasan

Berdasarkan hasil pengujian dan evaluasi yang telah dilakukan, terlihat adanya perbedaan performa model antara tahap pelatihan menggunakan Edge Impulse dan tahap implementasi pada perangkat ESP32-CAM. Pada tahap pelatihan, model MobileNetV2 yang diimplementasikan melalui platform Edge Impulse menunjukkan performa yang sangat baik dengan tingkat akurasi sebesar 92,50%. Nilai AUC sebesar 0,97 serta nilai *precision*, *recall*, dan *F1-score* yang masing-masing mencapai 0,98, 0,97, dan 0,97 menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam membedakan antara kelas botol dan non-botol. Berdasarkan hasil *confusion matrix* yang telah disederhanakan melalui verifikasi visual, model mampu mengenali objek botol dengan tingkat akurasi mencapai 100%, sedangkan pada kelas non-botol menunjukkan tingkat akurasi sebesar 95%, dengan kesalahan klasifikasi sebesar 5% berupa satu data non-botol yang diklasifikasikan sebagai botol. Hal ini menunjukkan bahwa secara umum model telah mampu mempelajari karakteristik objek dengan baik pada kondisi pelatihan, meskipun masih terdapat sedikit kesalahan pada objek non-botol yang memiliki kemiripan visual dengan botol.

Namun, setelah model diimplementasikan pada perangkat ESP32-CAM dan diuji pada kondisi nyata, terjadi penurunan performa yang cukup signifikan dengan tingkat akurasi sebesar 65,7%. Meskipun sistem masih mampu mengenali objek botol dengan baik, kesalahan klasifikasi pada objek non-botol meningkat secara signifikan, di mana sebagian besar objek non-botol diklasifikasikan sebagai botol. Hal ini menunjukkan bahwa sistem memiliki kecenderungan untuk mengklasifikasikan objek sebagai botol (*bias terhadap kelas boto*), yang menjadi penyebab utama penurunan akurasi pada tahap implementasi.

Perbedaan performa ini menunjukkan adanya kesenjangan antara kondisi pelatihan dan kondisi implementasi di dunia nyata (*real-world environment gap*). Pada tahap pelatihan, data yang digunakan relatif lebih terkontrol dari segi pencahayaan, sudut pengambilan gambar, dan latar belakang. Sementara itu, pada tahap implementasi, sistem harus menghadapi kondisi lingkungan yang lebih dinamis dan tidak terkontrol. Selain itu, distribusi dataset yang belum sepenuhnya seimbang, terutama pada variasi sudut pengambilan gambar yang didominasi oleh sudut depan, serta keterbatasan variasi objek non-botol, turut mempengaruhi kemampuan generalisasi model.

Selain itu, sistem yang dikembangkan hanya dilatih menggunakan citra dengan satu objek utama dalam satu gambar. Oleh karena itu, ketika terdapat lebih dari satu objek dalam satu citra, misalnya botol dan non-botol secara bersamaan, sistem cenderung langsung mengklasifikasikan citra tersebut sebagai non-botol. Hal ini terjadi karena model tidak memiliki referensi data untuk kondisi multi-objek, sehingga belum mampu melakukan deteksi dan klasifikasi secara simultan.

Meskipun demikian, sistem yang dikembangkan telah berhasil mengimplementasikan model MobileNetV2 pada perangkat ESP32-CAM dan mampu melakukan klasifikasi citra secara real-time. Hal ini menunjukkan bahwa pendekatan *edge artificial intelligence* menggunakan

platform Edge Impulse dapat diterapkan dengan baik pada perangkat *embedded* dengan keterbatasan sumber daya.

Keterbatasan perangkat ESP32-CAM juga menjadi faktor yang mempengaruhi performa sistem. Kualitas citra yang dihasilkan, resolusi yang terbatas, serta penggunaan model yang telah dikompresi dalam format TensorFlow Lite (*quantized model*) menyebabkan penurunan kemampuan model dalam mengekstraksi fitur secara optimal. Hal ini berdampak pada meningkatnya kesalahan klasifikasi, khususnya pada objek non-botol yang memiliki kemiripan visual dengan botol.

Selain itu, terdapat beberapa keterbatasan dalam penelitian ini. Objek botol yang digunakan dalam dataset terbatas pada kemasan botol air mineral, sehingga model belum mampu mengenali jenis botol lain secara umum. Akibatnya, objek botol dengan bentuk atau karakteristik yang berbeda dari botol air mineral berpotensi diklasifikasikan sebagai non-botol.

KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa model MobileNetV2 berhasil diimplementasikan pada perangkat ESP32-CAM dengan memanfaatkan platform Edge Impulse sebagai sarana pengembangan dan deployment model. Pada tahap pelatihan, model menunjukkan performa yang sangat baik dengan akurasi sebesar 92,50% serta didukung oleh nilai AUC, precision, recall, dan F1-score yang tinggi. Hasil confusion matrix yang telah disederhanakan melalui verifikasi visual menunjukkan bahwa model mampu mengenali objek botol dengan akurasi 100% dan objek non-botol sebesar 95%, dengan kesalahan klasifikasi sebesar 5%. Hal ini menunjukkan bahwa model memiliki kemampuan yang baik dalam membedakan objek botol dan non-botol pada kondisi terkontrol. Namun, pada tahap implementasi di perangkat ESP32-CAM, performa model mengalami penurunan dengan akurasi sebesar 65,7%. Meskipun sistem tetap menunjukkan kemampuan yang baik dalam mengenali objek botol, kesalahan klasifikasi pada objek non-botol masih cukup tinggi. Hal ini disebabkan oleh perbedaan kondisi antara data pelatihan dan kondisi nyata (*real-world environment gap*), keterbatasan perangkat *embedded*, serta kurangnya variasi dataset yang mempengaruhi kemampuan generalisasi model. Selain itu, penelitian ini memiliki beberapa keterbatasan, di antaranya objek botol yang digunakan hanya terbatas pada botol air mineral sehingga model belum mampu mengenali jenis botol lain secara umum. Selain itu, sistem yang dikembangkan belum mampu menangani citra dengan lebih dari satu objek dalam satu waktu, sehingga ketika terdapat kombinasi objek dalam satu gambar, hasil klasifikasi cenderung tidak optimal. Meskipun demikian, sistem yang dikembangkan telah berhasil menunjukkan bahwa implementasi model *deep learning* berbasis *edge artificial intelligence* pada perangkat dengan keterbatasan sumber daya seperti ESP32-CAM dapat dilakukan dan mampu berjalan secara *real-time*. Hal ini menunjukkan bahwa pendekatan yang digunakan memiliki potensi untuk dikembangkan lebih lanjut dalam aplikasi nyata. Untuk pengembangan selanjutnya, diperlukan peningkatan pada variasi dan keseimbangan dataset, khususnya pada objek non-botol dan jenis botol yang lebih beragam, penerapan teknik data augmentation yang lebih kompleks, serta proses *fine-tuning* menggunakan data dari kondisi nyata. Selain itu, pengembangan sistem ke arah deteksi multi-objek serta peningkatan kualitas input citra juga dapat menjadi solusi untuk meningkatkan akurasi dan keandalan sistem secara keseluruhan.

REFERENSI

- Banbury, C. R., Reddi, V. J., Lam, M., Fu, W., Fazel, A., Holleman, J., ... Yadav, P. (2021). Benchmarking TinyML Systems: Challenges and Direction. Retrieved from <http://arxiv.org/abs/2003.04821>
- Ben Aicha, F. (2025). Optimized embedded AI: efficient implementation of CNNs on ESP32-CAM for real-time image classification. *Computing*, 107(11), 206. <https://doi.org/10.1007/s00607-025-01559-z>
- Chang, Y. H., Wu, F. C., & Lin, H. W. (2025). Design and Implementation of ESP32-Based Edge Computing for Object Detection. *Sensors*, 25(6). <https://doi.org/10.3390/s25061656>
- Chen, L., Li, S., Bai, Q., Yang, J., Jiang, S., & Miao, Y. (2021). Review of Image Classification Algorithms Based on Convolutional Neural Networks. *Remote Sensing*, 13(22), 4712. <https://doi.org/10.3390/rs13224712>
- Deng, B. L., Li, G., Han, S., Shi, L., & Xie, Y. (2020). Model Compression and Hardware Acceleration for Neural Networks: A Comprehensive Survey. *Proceedings of the IEEE*, 108(4), 485–532. <https://doi.org/10.1109/JPROC.2020.2976475>

- Diab, M. S., & Rodriguez-Villegas, E. (2022). Performance Evaluation of Embedded Image Classification Models Using Edge Impulse for Application on Medical Images. In *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* (pp. 2639–2642). IEEE. <https://doi.org/10.1109/EMBC48229.2022.9871108>
- Gaffar, A. W. M., Widyawati, D., Fahmi, F., & Ulfiani, S. (2025). Deteksi dan Klasifikasi Jenis Kacang-Kacangan Menggunakan CNN Berbasis ESP32-Cam. *Jurnal Minfo Polgan*, *14*(2), 2234–2241. <https://doi.org/10.33395/jmp.v14i2.15359>
- Kong, L., Tan, J., Huang, J., Chen, G., Wang, S., Jin, X., ... Das, S. K. (2023). Edge-computing-driven Internet of Things: A Survey. *ACM Computing Surveys*, *55*(8), 1–41. <https://doi.org/10.1145/3555308>
- Lim, S.-H., Kang, S.-H., Ko, B.-H., Roh, J., Lim, C., & Cho, S.-Y. (2022). An Integrated Analysis Framework of Convolutional Neural Network for Embedded Edge Devices. *Electronics*, *11*(7), 1041. <https://doi.org/10.3390/electronics11071041>
- Raza, S. M., Abidi, S. M. H., Masuduzzaman, M., & Shin, S. Y. (2025). Lightweight deep learning for visual perception: A survey of models, compression strategies, and edge deployment challenges. *Neurocomputing*, *656*, 131357. <https://doi.org/10.1016/j.neucom.2025.131357>
- Sarker, I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science*, *2*(3), 160. <https://doi.org/10.1007/s42979-021-00592-x>
- Wasim Khan. (2024). Image Classification using modified Convolutional Neural Network. *Journal of Electrical Systems*, *20*(3), 3465–3472. <https://doi.org/10.52783/jes.4982>
- Yong, L., Ma, L., Sun, D., & Du, L. (2023). Application of MobileNetV2 to waste classification. *PLOS ONE*, *18*(3), e0282336. <https://doi.org/10.1371/journal.pone.0282336>
- Zahirah, D., Purnawansyah, P., Kurniati, N., & Darwis, H. (2024). DIGITAL IMAGE CLASSIFICATION OF HERBAL LEAVES USING KNN AND CNN WITH GLCM FEATURES. *Jurnal Teknik Informatika (Jutif)*, *5*(1), 61–67. <https://doi.org/10.52436/1.jutif.2024.5.1.1162>
- Zheng, D., Wang, R., Duan, Y., Pang, P. C.-I., & Tan, T. (2023). Focus-RCNet: a lightweight recyclable waste classification algorithm based on focus and knowledge distillation. *Visual Computing for Industry, Biomedicine, and Art*, *6*(1), 19. <https://doi.org/10.1186/s42492-023-00146-3>
- Zhou, Z., Chen, X., Li, E., Zeng, L., Luo, K., & Zhang, J. (2019). Edge Intelligence: Paving the Last Mile of Artificial Intelligence With Edge Computing. *Proceedings of the IEEE*, *107*(8), 1738–1762. <https://doi.org/10.1109/JPROC.2019.2918951>