

BAB II

KAJIAN PUSTAKA

2.1 Dasar Teori

Bagian ini memaparkan landasan konseptual yang mendasari pengembangan sistem, meliputi teori citra digital dan pengolahan citra, pengenalan citra bahan pangan, arsitektur Vision Transformer sebagai ekstraktor fitur, algoritma XGBoost untuk klasifikasi multiclass, deteksi objek dengan YOLOv8n sebagai modul lokalisasi bahan pangan, serta konsep kalori dan pemanfaatan Tabel Komposisi Pangan Indonesia sebagai referensi deterministik perhitungan kalori.

a. Citra Digital dan Pengolahan Citra

Citra digital merupakan representasi visual yang tersusun atas piksel sebagai unit dasar penyimpan informasi intensitas dan warna. Setiap piksel pada citra berwarna umumnya terdiri atas tiga kanal utama, yaitu merah, hijau, dan biru, yang membentuk representasi RGB (Archana et al., 2024). Dalam sistem, citra digital diperlakukan sebagai matriks numerik yang dapat dianalisis secara komputasional (Kumar et al., 2022).

Pengolahan citra berfungsi sebagai tahap prapemrosesan sebelum proses ekstraksi fitur dilakukan. Tahapan yang umum diterapkan meliputi penyesuaian ukuran citra agar seragam, normalisasi distribusi nilai piksel, serta transformasi tambahan untuk meningkatkan kualitas representasi data. Penyeragaman resolusi penting untuk memastikan konsistensi masukan model, terutama ketika menggunakan arsitektur pra latih seperti Vision Transformer yang memiliki spesifikasi dimensi tetap (Dosovitskiy, 2020).

Selain itu, teknik augmentasi data digunakan untuk memperkaya variasi visual pada data pelatihan. Transformasi yang diterapkan meliputi rotasi terbatas, pembalikan horizontal, penyesuaian kecerahan, zoom in, zoom out, dan Gaussian blur, yang secara bersama-sama merepresentasikan variasi kondisi pengambilan gambar seperti perbedaan sudut, orientasi, pencahayaan, jarak, dan ketajaman objek. Keseluruhan transformasi tersebut tidak mengubah identitas semantik objek sehingga label kelas tetap konsisten. Strategi ini terbukti efektif dalam mengurangi risiko overfitting pada dataset berukuran terbatas (Shorten, 2019). Dalam penelitian ini, augmentasi hanya diterapkan pada subset data pelatihan untuk menjaga integritas evaluasi.

b. Pengenalan Citra Bahan Pangan

Pengenalan citra bahan pangan termasuk dalam kategori *fine-grained image classification*, yaitu klasifikasi objek dengan tingkat kemiripan visual yang tinggi antar kelas (Wei et al., 2022). Beberapa bahan pangan, seperti tomat dan cabai merah atau bawang merah dan bawang putih, memiliki warna dan tekstur yang hampir serupa. Kondisi ini menuntut model untuk mampu menangkap perbedaan visual yang relatif halus agar proses identifikasi dapat dilakukan secara tepat.

Berbeda dengan penelitian yang memperkirakan kalori secara langsung dari citra melalui pendekatan regresi (Thames et al., 2021), penelitian ini memformulasikan permasalahan sebagai klasifikasi multiclass. Model berperan untuk mengenali jenis bahan pangan berdasarkan citra, sedangkan nilai kalori dihitung setelah kelas bahan diketahui dengan mengacu pada Tabel Komposisi Pangan Indonesia. Pemisahan antara tahap identifikasi visual dan tahap perhitungan kalori dilakukan agar alur sistem lebih jelas (Jiang et al., 2020). Dengan cara ini, proses klasifikasi dan perhitungan dapat dianalisis secara terpisah sesuai fungsinya masing-masing.

c. Vision Transformer (ViT)

Vision Transformer merupakan arsitektur *deep learning* yang mengadaptasi mekanisme Transformer ke ranah visi komputer (Dosovitskiy, 2020). Berbeda dari Convolutional Neural Network yang mengandalkan operasi konvolusi (Lecun et al., 2015), Vision Transformer membagi citra menjadi *patch* berukuran tetap. Setiap *patch* diproyeksikan ke ruang *embedding* dan diproses menggunakan mekanisme *self attention*. Melalui *self attention*, model dapat menangkap hubungan antar bagian citra secara menyeluruh, tidak hanya berfokus pada pola lokal tetapi juga mempertimbangkan konteks global (S. Khan et al., 2022). Kemampuan ini telah dibuktikan efektif pada berbagai tugas pengenalan visual berbasis citra, salah satunya pada sistem pengenalan gestur tangan yang berhasil mencapai kinerja tinggi dengan memanfaatkan mekanisme self-attention untuk menangkap hubungan spasial antarbagian citra (Haq et al., 2024).

Pendekatan ini relevan untuk membedakan bahan pangan yang sering memiliki perbedaan tekstur dan bentuk yang relatif halus (X. Gao et al., 2024). Dalam penelitian ini, Vision Transformer digunakan sebagai ekstraktor fitur visual. Model pra latih berbasis ImageNet dimanfaatkan untuk menghasilkan representasi *embedding* berdimensi tetap. Selama pelatihan, bobot Vision Transformer tidak diperbarui sehingga model berfungsi sebagai penyedia fitur, sedangkan proses penyesuaian parameter dilakukan pada tahap klasifikasi.

d. XGBoost untuk Klasifikasi

XGBoost merupakan algoritma *gradient boosting* berbasis pohon keputusan yang dirancang untuk membangun model secara bertahap melalui kombinasi sejumlah pohon dengan fungsi objektif terregularisasi (T. Chen & Guestrin, 2016). Pendekatan ini memungkinkan model memperbaiki kesalahan prediksi pada setiap iterasi sehingga terbentuk model akhir yang lebih stabil. Selain digunakan pada regresi, XGBoost juga mendukung klasifikasi melalui fungsi objektif *softmax* (Li et al., 2024).

Dalam penelitian ini, XGBoost berperan untuk memetakan vektor *embedding* hasil ekstraksi Vision Transformer ke dalam salah satu dari sepuluh kelas bahan pangan. Kemampuan XGBoost dalam memodelkan hubungan nonlinier dan menerapkan regularisasi menjadikannya sesuai untuk mengolah fitur berdimensi tinggi (Dong et al., 2024). Integrasi Vision Transformer sebagai ekstraktor fitur dan XGBoost sebagai *classifier* membentuk arsitektur yang memisahkan tahap representasi visual dan tahap pengambilan keputusan klasifikasi.

e. Deteksi Objek dengan YOLO

Deteksi objek berbeda dari klasifikasi citra karena tidak sekadar mengenali kelas, tetapi juga menentukan posisi setiap objek lewat bounding box (Zou et al., 2023). Kemampuan ini dibutuhkan saat satu gambar memuat beberapa bahan pangan yang perlu dikenali satu per satu. YOLO termasuk detektor satu tahap dimana lokalisasi dan klasifikasi dilakukan sekaligus dalam satu forward pass, tanpa tahap pengajuan region terpisah seperti pada Faster R-CNN, sehingga inferensinya lebih cepat (Redmon et al., 2016). YOLOv8, yang dirilis oleh Ultralytics, memakai backbone CSPDarknet, neck PANet untuk fusi fitur multiskala, serta decoupled head yang memisahkan prediksi kelas dari regresi bounding box.

Arsitektur ini juga bersifat anchor-free sehingga tidak bergantung pada anchor box yang didefinisikan manual, dan lebih fleksibel terhadap objek dengan rasio aspek beragam. YOLOv8 memiliki lima varian ukuran: nano, small, medium, large, dan extra-large. Penelitian ini menggunakan varian nano (YOLOv8n) karena dua alasan praktis yaitu kapasitas VRAM GPU yang terbatas, dan kompleksitas tugas yang relatif rendah hanya sepuluh kelas bahan pangan dengan latar belakang cukup seragam. Meski berukuran kecil, YOLOv8n tetap memanfaatkan bobot pra-latih COCO (80 kelas objek umum) sebagai titik awal fine-tuning pada domain bahan pangan (Farooq et al., 2023).

f. Kalori dan Energi Pangan

Kalori merupakan satuan energi yang diperoleh tubuh dari konsumsi bahan pangan dan dinyatakan dalam kilokalori. Energi pangan terutama berasal dari karbohidrat, protein, dan lemak. Secara umum, faktor konversi yang digunakan adalah 4 kilokalori per gram untuk karbohidrat dan protein, serta 9 kilokalori per gram untuk lemak, sebagaimana dijelaskan dalam sistem faktor (Ensminger et al., 2023). Dalam penelitian ini, kalori dihitung setelah jenis bahan pangan teridentifikasi. Nilai tersebut tidak dipelajari oleh model klasifikasi, melainkan diperoleh melalui perhitungan proporsional berdasarkan berat aktual dan nilai kalori per 100 gram yang tercantum dalam referensi TKPI.

g. Tabel Komposisi Pangan Indonesia (TKPI)

Tabel Komposisi Pangan Indonesia merupakan basis data nasional yang diterbitkan oleh Kementerian Kesehatan Republik Indonesia dan memuat informasi kandungan gizi bahan pangan yang umum dikonsumsi masyarakat (Kementerian Kesehatan RI, 2020). Data dalam TKPI diperoleh melalui analisis laboratorium dan kompilasi penelitian yang disesuaikan dengan karakteristik pangan lokal. Dalam penelitian ini, TKPI digunakan sebagai referensi nilai kalori per 100 gram bahan pangan. Setelah sistem mengidentifikasi jenis bahan, nilai kalori diambil dari basis data TKPI dan dihitung secara proporsional sesuai berat yang dimasukkan pengguna. Nilai dalam TKPI merupakan nilai rata-rata yang dapat bervariasi tergantung varietas dan kondisi bahan. Oleh karena itu, hasil yang diperoleh dipahami sebagai estimasi berbasis referensi standar, bukan representasi seluruh variasi kondisi nyata.

2.1.1 Pustaka

Berbagai penelitian mengenai estimasi kalori makanan berbasis citra menunjukkan bahwa penggabungan analisis visual dan pemodelan numerik merupakan pendekatan yang menjanjikan untuk memperkirakan nilai kalori makanan secara otomatis. Kajian sistematis oleh Konstantakopoulos et al. (2024) mengungkapkan bahwa sebagian besar sistem estimasi kalori berbasis citra menggunakan alur kerja yang diawali dengan analisis visual makanan, seperti deteksi atau segmentasi objek, kemudian dilanjutkan dengan pemodelan kuantitatif untuk memperkirakan atribut nutrisi, termasuk volume, berat, dan nilai kalori. Pendekatan tersebut terbukti efektif dalam lingkungan terkontrol dengan anotasi yang lengkap, namun kompleksitas sistem meningkat seiring bertambahnya jumlah parameter yang harus diestimasi.

Sejumlah penelitian menekankan pentingnya efisiensi komputasi dalam pengembangan sistem estimasi makanan berbasis citra, khususnya untuk aplikasi yang memerlukan respons

cepat. Oleh karena itu, model berbasis Convolutional Neural Network (CNN) masih banyak digunakan karena mampu memberikan kinerja yang baik dengan kebutuhan sumber daya komputasi yang relatif rendah. Namun demikian, sebagian besar pendekatan berbasis CNN masih berfokus pada klasifikasi jenis makanan, sehingga informasi kuantitatif berupa nilai kalori belum diprediksi secara langsung sebagai variabel kontinu.

Di Indonesia, penelitian terkait citra makanan umumnya masih terbatas pada tugas pengenalan atau klasifikasi jenis makanan. Studi oleh Wibisono, (2020) yang mengembangkan dataset *Traditional Food Knowledge Indonesia* menunjukkan bahwa pendekatan *transfer learning* dari ImageNet mampu menghasilkan akurasi klasifikasi yang tinggi pada makanan tradisional. Meskipun demikian, pendekatan tersebut belum memungkinkan estimasi nilai kalori makanan secara otomatis karena kalori tidak dimodelkan sebagai target regresi.

Perkembangan terkini dalam bidang *computer vision* menunjukkan bahwa arsitektur berbasis Transformer, khususnya Vision Transformer (ViT), menjadi alternatif yang kompetitif terhadap CNN dalam pengolahan citra. Vision Transformer memanfaatkan mekanisme *self-attention* untuk menangkap hubungan global antarbagian citra, sehingga mampu merepresentasikan konteks visual secara lebih menyeluruh dibandingkan pendekatan berbasis konvolusi (Dosovitskiy et al., 2020). Beberapa penelitian melaporkan bahwa model berbasis Transformer lebih efektif dalam menangani citra makanan dengan variasi visual yang tinggi dan komposisi yang kompleks, karena mampu memodelkan dependensi spasial jarak jauh yang sulit ditangkap oleh CNN konvensional (Konstantakopoulos et al., 2024).

Dalam aspek pelabelan data, sebagian besar penelitian estimasi kalori berbasis citra masih mengandalkan basis data nutrisi internasional, seperti USDA FoodData Central atau FatSecret. Penggunaan basis data tersebut berpotensi menghasilkan estimasi yang kurang relevan ketika diterapkan pada makanan lokal Indonesia, mengingat adanya perbedaan bahan baku, resep, serta ukuran porsi konsumsi. Tabel Komposisi Pangan Indonesia (TKPI) merupakan rujukan nasional yang lebih sesuai untuk konteks pangan Indonesia, namun pemanfaatannya sebagai *ground truth* dalam penelitian estimasi kalori berbasis citra masih relatif terbatas.

Penelitian oleh Adjuik et al. (2024) menunjukkan performa estimasi kalori yang tinggi, namun menggunakan data komposisi nutrisi terstruktur sebagai masukan utama model. Pendekatan tersebut belum merepresentasikan tantangan nyata penggunaan citra sebagai sumber data, karena informasi nutrisi telah tersedia secara eksplisit sebelum proses pemodelan. Dengan demikian, integrasi langsung antara ekstraksi fitur visual dari citra makanan dan

pemodelan regresi nilai kalori yang mengacu pada standar gizi nasional masih belum banyak dikaji.

Berdasarkan tinjauan pustaka tersebut, sebagian besar penelitian estimasi makanan berbasis citra masih berfokus pada klasifikasi jenis makanan atau menggunakan basis data nutrisi internasional sebagai acuan. Penelitian yang secara khusus memodelkan estimasi nilai kalori makanan Indonesia secara langsung dari citra dengan mengacu pada standar gizi nasional masih terbatas. Untuk menjawab keterbatasan tersebut, penelitian ini mengusulkan pengembangan sistem estimasi kalori makanan Indonesia berbasis citra menggunakan pendekatan hibrida Vision Transformer dan XGBoost dengan acuan nilai kalori dari Tabel Komposisi Pangan Indonesia.

2.1.2 Obyek Penelitian

Objek penelitian dalam studi ini mencakup data visual bahan pangan, referensi nilai kalori berbasis standar nasional TKPI, serta komponen model komputasional yang digunakan untuk melakukan klasifikasi jenis bahan pangan. Secara konseptual, objek penelitian tidak difokuskan pada prediksi langsung nilai kalori dari citra, melainkan pada pengembangan sistem yang mengintegrasikan identifikasi visual bahan pangan dan perhitungan kalori berbasis referensi.

Komponen pertama berupa citra bahan pangan sebagai data masukan sistem. Citra dikumpulkan dari sumber publik dan sebagian melalui pengambilan gambar langsung dalam kondisi terkontrol. Objek yang direpresentasikan terbatas pada sepuluh jenis bahan pangan yang umum dikonsumsi, seperti beras putih, telur ayam ras segar, kentang segar, dan sayuran tertentu. Setiap citra merepresentasikan satu jenis bahan tunggal untuk menjaga konsistensi pelabelan dan meminimalkan ambiguitas komposisi.

Citra yang digunakan tidak mencakup makanan siap saji atau hidangan multi-komponen. Pembatasan ini dilakukan untuk memastikan bahwa setiap kelas memiliki identitas visual yang jelas dan dapat dipetakan secara langsung ke nilai kalori referensial berdasarkan Tabel Komposisi Pangan Indonesia. Komponen kedua berupa nilai kalori referensi yang diperoleh dari Tabel Komposisi Pangan Indonesia (TKPI) yang diterbitkan oleh Kementerian Kesehatan Republik Indonesia (Kementerian Kesehatan RI, 2020).

Nilai energi dalam TKPI dinyatakan dalam satuan kilokalori per 100 gram bahan pangan. Dalam penelitian ini, nilai tersebut tidak digunakan sebagai target model, melainkan sebagai dasar perhitungan deterministik setelah proses klasifikasi selesai dilakukan. Dengan demikian, ground truth pada model ini adalah label kelas bahan pangan, bukan nilai kalori

numerik. Perhitungan energi dilakukan secara terpisah menggunakan rumus proporsional berbasis berat aktual yang dimasukkan pengguna dan nilai referensi dari TKPI. Komponen ketiga adalah model komputasional yang terdiri atas Vision Transformer sebagai ekstraktor fitur visual dan XGBoost sebagai *classifier*.

Vision Transformer digunakan untuk menghasilkan representasi *embedding* dari citra bahan pangan (Dosovitskiy, 2020), sedangkan XGBoost memetakan representasi tersebut ke dalam salah satu kelas bahan pangan. Model tidak dirancang untuk memprediksi nilai kalori secara langsung, melainkan untuk melakukan identifikasi jenis bahan sebagai dasar perhitungan kalori berbasis referensi. Dengan struktur ini, objek penelitian mencakup sistem klasifikasi visual bahan pangan yang terintegrasi dengan mekanisme perhitungan kalori berbasis standar nasional, sehingga membentuk pipeline dua tahap yang terpisah secara fungsional namun terhubung secara sistematis.

2.1.3 Metode, Teori, dan Algoritma

Penelitian analisis pangan berbasis citra umumnya memanfaatkan pendekatan *machine learning* dan *deep learning* seperti Convolutional Neural Network, Multilayer Perceptron, serta berbagai metode klasifikasi dan regresi. CNN banyak digunakan karena mampu mengekstraksi fitur visual secara otomatis melalui operasi konvolusi. Namun, performanya sangat dipengaruhi oleh ketersediaan data latih dalam jumlah besar dan variasi visual yang memadai (A. Khan et al., 2023). Evaluasi komparatif antara beberapa metode klasifikasi merupakan praktik yang umum dilakukan untuk menentukan pendekatan yang paling sesuai dengan karakteristik data dan tugas yang dikerjakan (Andana et al., 2019). Pada bahan pangan yang memiliki perbedaan tekstur, warna, dan bentuk yang relatif halus, representasi visual yang mampu menangkap konteks global menjadi penting.

Pendekatan konvolusional cenderung berfokus pada pola lokal, sehingga dalam beberapa kasus memerlukan arsitektur yang lebih dalam untuk menangkap hubungan visual secara menyeluruh. Perkembangan arsitektur berbasis Transformer menghadirkan alternatif melalui Vision Transformer (ViT). Model ini membagi citra menjadi sejumlah *patch* yang diproses menggunakan mekanisme *self attention* untuk memodelkan hubungan antar bagian citra secara global (Dosovitskiy, 2020). Dalam penelitian ini, ViT digunakan sebagai ekstraktor fitur visual dengan memanfaatkan model pra latih berbasis ImageNet untuk menghasilkan vektor *embedding* berdimensi tetap. Bobot ViT tidak diperbarui selama pelatihan sehingga proses penyesuaian parameter difokuskan pada tahap klasifikasi.

Untuk pengambilan keputusan, digunakan XGBoost sebagai *classifier*. XGBoost merupakan algoritma *gradient boosting* berbasis pohon keputusan yang membangun model secara bertahap melalui optimasi fungsi objektif terregularisasi (Chen & Guestrin, 2016). Algoritma ini mampu memodelkan hubungan nonlinier pada fitur berdimensi tinggi dan digunakan untuk memetakan *embedding* hasil ekstraksi fitur ke dalam kelas bahan pangan. Berbeda dari penelitian yang memperkirakan kalori melalui regresi langsung dari citra, penelitian ini memformulasikan permasalahan sebagai klasifikasi bahan pangan.

Nilai kalori tidak dipelajari oleh model, melainkan dihitung secara deterministik berdasarkan referensi Tabel Komposisi Pangan Indonesia dan berat yang dimasukkan pengguna. Integrasi Vision Transformer dan XGBoost membentuk arsitektur dua tahap yang memisahkan proses representasi visual dan proses klasifikasi. Dengan struktur ini, tahap identifikasi bahan dan tahap perhitungan kalori dapat dianalisis secara terpisah sesuai fungsinya masing-masing.

2.2 Penelitian Terdahulu

Penelitian mengenai analisis pangan berbasis citra berkembang signifikan dalam dua dekade terakhir, khususnya pada pengenalan jenis makanan dan estimasi atribut kuantitatif seperti volume, berat, dan kalori. Secara umum, sistem berbasis citra mencakup beberapa tahapan, yaitu deteksi atau segmentasi objek, ekstraksi fitur visual, klasifikasi, dan pemetaan ke basis data nutrisi (Lo et al., 2020). Studi tersebut menunjukkan bahwa ketepatan pada tahap identifikasi visual berpengaruh terhadap hasil estimasi atribut selanjutnya.

Seiring berkembangnya *deep learning*, berbagai penelitian memanfaatkan arsitektur Convolutional Neural Network dan Transformer untuk meningkatkan kualitas representasi visual. Wang et al. (2024) melaporkan bahwa arsitektur berbasis Transformer seperti Swin Transformer mampu menangkap informasi global dan lokal secara efektif dalam analisis komposisi nutrisi berbasis citra. Namun, sebagian besar penelitian tersebut menggunakan dataset internasional dan menerapkan pendekatan *end to end* tanpa memisahkan tahap representasi visual dan tahap perhitungan numerik.

Kajian sistematis oleh Konstantakopoulos et al. (2024) menunjukkan bahwa penelitian di bidang ini masih didominasi oleh dataset non lokal dan belum banyak mengintegrasikan referensi gizi nasional spesifik suatu negara. Selain itu, keterbatasan anotasi nutrisi langsung serta minimnya eksplorasi pendekatan modular juga menjadi perhatian. Dalam konteks algoritma boosting, Otchere et al. (2022) menunjukkan bahwa metode gradient boosting efektif dalam memodelkan hubungan nonlinier pada data berdimensi tinggi.

Konstantakopoulos et al. (2023) juga melaporkan bahwa XGBoost menunjukkan performa yang konsisten pada estimasi atribut kuantitatif berbasis fitur citra, meskipun fokusnya pada estimasi berat. Temuan ini mengindikasikan potensi penggunaan boosting dalam memproses representasi visual hasil ekstraksi fitur. Pendekatan hibrida yang menggabungkan ekstraktor fitur berbasis *deep learning* dan algoritma *machine learning* lanjutan mulai banyak dikaji. Raju et al. (2024) menunjukkan bahwa kombinasi Vision Transformer dan XGBoost dapat digunakan untuk klasifikasi pada citra medis.

Meskipun berada pada domain berbeda, pendekatan tersebut menunjukkan bahwa pemisahan antara tahap representasi visual dan tahap klasifikasi dapat diterapkan secara modular. Sebaliknya, penelitian seperti Agarwal et al. (2023) dan Mvd & Kishore Gattim, (2024) mengembangkan estimasi kalori menggunakan pendekatan regresi langsung dari citra berbasis CNN. Pendekatan ini bergantung pada ketersediaan anotasi kalori pada dataset dan umumnya tidak mengaitkan hasil estimasi dengan standar gizi nasional tertentu.

Berdasarkan telaah tersebut, dapat dilihat bahwa penelitian terdahulu umumnya menggunakan pendekatan end to end atau regresi berbasis fitur untuk memperkirakan atribut nutrisi langsung dari citra. Sementara itu, penelitian yang secara eksplisit memisahkan tahap klasifikasi bahan pangan dan tahap perhitungan kalori berbasis referensi gizi nasional masih terbatas. Ruang inilah yang menjadi landasan pengembangan sistem dalam penelitian ini. Ringkasan penelitian terdahulu yang relevan disajikan pada Tabel 2.1.

Tabel 2. 1 Penelitian Terdahulu

Peneliti Tahun Judul	Metode / Model Utama	Dataset	Kontribusi & Keterbatasan	Relevansi terhadap Penelitian
(Konstantakopoulos et al., 2024) A Review of Image-Based Food Recognition and Volume Estimation AI Systems	Review sistematis metode berbasis citra (CNN, Transformer)	Food-101, UEC-Food, VIREO Food-172, UNIMIB 2016	Mengkaji perkembangan sistem pengenalan makanan dan estimasi volume. Menyoroti dominasi dataset internasional dan keterbatasan anotasi nutrisi langsung.	Menjadi dasar teoretis mengenai tantangan dataset dan kebutuhan pendekatan modular. Menunjukkan relevansi penggunaan dataset lokal dan referensi nasional seperti TKPI.

Peneliti / Tahun / Judul	Metode / Model Utama	Dataset	Kontribusi & Keterbatasan	Relevansi terhadap Penelitian
(Wang et al., 2024) Nutritional Composition Analysis Using Swin Transformer	Swin Transformer dan Vision transformer	Dataset citra makanan beranotasi nutrisi	Menunjukkan efektivitas Transformer dalam menangkap fitur global dan lokal. Keterbatasan pada penggunaan dataset non-Indonesia dan pendekatan end-to-end.	Menguatkan pemilihan Vision transformer sebagai ekstraktor fitur visual. Berbeda karena penelitian ini tidak melakukan prediksi langsung dari citra.
(Konstantakopoulos et al., 2023) Estimation of Food Weight Using Boosting Algorithms	Ekstraksi fitur citra + XGBoost, CatBoost, LightGBM	MedGRF food	Menunjukkan stabilitas XGBoost pada estimasi atribut kuantitatif berbasis fitur visual. Fokus pada estimasi berat.	Menjadi dasar pemilihan XGBoost sebagai <i>classifier</i> berbasis fitur <i>embedding</i> . Pada penelitian ini digunakan untuk klasifikasi multiclass, bukan regresi atribut numerik.
(Mvd Kishore Gattim, 2024) Deep learning-based Food Calorie Estimation	CNN + SVR, Random Forest, XGBoost	Dataset makanan beranotasi kalori	Menggunakan regresi langsung dari citra untuk memprediksi kalori. Bergantung pada anotasi numerik dataset.	Berfungsi sebagai pembanding pendekatan end-to-end regresi. Penelitian ini berbeda karena memisahkan klasifikasi bahan dan perhitungan kalori deterministik berbasis TKPI.
(Agarwal et al., 2023) Hybrid Deep learning for Food Recognition and Calorie Estimation	Hybrid CNN untuk klasifikasi dan regresi kalori	Dataset makanan global	Menggabungkan beberapa CNN untuk meningkatkan akurasi estimasi kalori. Kompleksitas model relatif tinggi.	Menjadi pembanding pendekatan <i>deep learning</i> murni. Penelitian ini menggunakan arsitektur hibrida yang memisahkan ekstraksi fitur dan klasifikasi.

Peneliti / Tahun Judul	Metode / Model Utama	Dataset	Kontribusi & Keterbatasan	Relevansi terhadap Penelitian
(Raju et al., 2024) Vision transformer and XGBoost Ensemble	Vision transformer + XGBoost	Citra medis	Menunjukkan efektivitas pendekatan hibrida dalam klasifikasi citra. Fokus pada domain medis.	Memberikan justifikasi metodologis untuk penggunaan arsitektur ViT–XGBoost pada tugas klasifikasi bahan pangan.
(Sultana et al., 2023) Food Value Estimation from Images	Review taksonomi metode estimasi nilai makanan	Food-101, UECFO OD, PFID	Mengidentifikasi dominasi dataset Barat dan keterbatasan generalisasi pada makanan Asia.	Menunjukkan relevansi penggunaan bahan pangan Indonesia dan relevansi referensi lokal seperti TKPI.
(K.-S. Lee & Lee, 2024) Multi-Spectral Food Classification	Deep learning berbasis citra multispektral	Dataset makanan multispektral	Menunjukkan peningkatan akurasi dengan informasi spektral tambahan. Membutuhkan sensor khusus.	Menjadi pembanding jenis data. Penelitian ini menggunakan citra RGB standar agar aplikatif pada kondisi umum.
(Louro et al., 2024) Recognition of Food Ingredients Dataset Analysis	Deep learning untuk klasifikasi bahan pangan dan analisis kualitas dataset	Dataset citra bahan pangan tunggal, Applied Sciences (open access)	Mengkaji pengenalan bahan pangan tunggal dan tantangan distribusi kelas serta konsistensi label. Tidak mengintegrasikan perhitungan nutrisi atau referensi nasional	Relevan karena fokus pada ingredient recognition, bukan makanan siap saji. Memperkuat pendekatan klasifikasi multiclass bahan tunggal dalam penelitian ini

Peneliti Tahun Judul	Metode / Model Utama	Dataset	Kontribusi & Keterbatasan	Relevansi terhadap Penelitian
(Nogay et al., 2025) Image-based Food Groups and Portion Prediction by Using Deep learning	Deep learning berbasis CNN untuk klasifikasi kelompok makanan dan estimasi porsi	Dataset citra makanan dengan anotasi kelompok pangan dan porsi	Menunjukkan bahwa klasifikasi berbasis citra dapat mencapai performa tinggi sebelum estimasi kuantitatif dilakukan. Pendekatan masih end to end dan tidak berbasis referensi nutrisi nasional	Mendukung pentingnya tahap klasifikasi sebelum estimasi atribut kuantitatif. Berbeda karena penelitian ini memisahkan klasifikasi bahan dan perhitungan kalori deterministik berbasis TKPI

Berdasarkan Tabel 2.1, sebagian besar penelitian terdahulu menggunakan pendekatan end to end untuk mengestimasi atribut nutrisi langsung dari citra atau memanfaatkan fitur visual untuk regresi kuantitatif. Arsitektur berbasis Transformer dan algoritma boosting banyak digunakan untuk meningkatkan kualitas representasi dan pemodelan. Namun, mayoritas studi menggunakan dataset internasional dan belum secara eksplisit mengintegrasikan referensi gizi nasional tertentu. Selain itu, pemisahan antara tahap klasifikasi bahan dan tahap perhitungan kalori masih relatif jarang dibahas. Penelitian ini mengadopsi pendekatan hibrida dengan memanfaatkan Vision transformer sebagai ekstraktor fitur, XGBoost sebagai *classifier*, serta perhitungan kalori berbasis referensi TKPI dalam satu sistem terintegrasi.

2.3 Analisa Kebaruan Penelitian

Berdasarkan telaah penelitian terdahulu, sistem analisis pangan berbasis citra umumnya berfokus pada dua pendekatan utama, yaitu klasifikasi jenis makanan atau estimasi atribut kuantitatif secara langsung melalui model regresi end-to-end. Sebagian besar penelitian menggunakan dataset internasional serta merujuk pada basis data nutrisi global (Kurniawardhani, 2020). Sementara itu, penelitian yang secara eksplisit memisahkan tahap identifikasi visual bahan pangan dan tahap perhitungan energi berbasis referensi gizi nasional masih terbatas, khususnya dalam konteks Indonesia. Berdasarkan kesenjangan tersebut, penelitian ini menawarkan beberapa aspek kebaruan sebagai berikut:

a. Fokus pada Bahan Pangan

Sebagian besar studi terdahulu menggunakan dataset makanan siap saji atau hidangan kompleks yang terdiri atas berbagai komponen. Pendekatan tersebut menghadapi tantangan dalam pemetaan komposisi bahan dan perhitungan energi secara akurat. Selain itu, dataset yang digunakan umumnya berasal dari konteks non-Indonesia, sehingga karakteristik visual pangan lokal belum terwakili secara memadai (Wibisono, 2020).

Penelitian ini membatasi objek pada bahan pangan yang umum dikonsumsi masyarakat Indonesia. Pembatasan ini bertujuan menjaga konsistensi pelabelan dan meminimalkan ambiguitas komposisi. Dengan menggunakan citra bahan pangan serta mengacu pada Tabel Komposisi Pangan Indonesia, penelitian ini menempatkan sistem dalam konteks pangan lokal yang lebih terstruktur dan terstandar (Mahaputri et al., 2022).

b. Pemisahan Tahap Klasifikasi dan Perhitungan Kalori

Berbeda dari pendekatan end-to-end yang memodelkan nilai kalori secara langsung dari citra, penelitian ini menerapkan skema dua tahap. Tahap pertama berupa klasifikasi multiclass untuk mengidentifikasi jenis bahan pangan. Tahap kedua berupa perhitungan kalori secara deterministik berdasarkan nilai energi per 100 gram yang tercantum dalam TKPI (Kementerian Kesehatan RI, 2020) dan berat yang dimasukkan pengguna.

Pendekatan ini tidak memosisikan kalori sebagai variabel prediktif yang dipelajari oleh model. Sebaliknya, model hanya bertugas mengenali jenis bahan, sedangkan estimasi energi dihitung menggunakan rumus proporsional berbasis standar nasional. Struktur ini memungkinkan pemisahan analisis kesalahan pada tahap identifikasi visual tanpa memengaruhi mekanisme perhitungan energi yang bersifat referensial.

c. Arsitektur Hibrida Vision Transformer – XGBoost

Penelitian ini menggunakan arsitektur hibrida yang mengintegrasikan Vision transformer sebagai ekstraktor fitur visual dan XGBoost sebagai *classifier* multiclass. Vision transformer memanfaatkan mekanisme self-attention untuk menangkap representasi global citra (Dosovitskiy, 2020), sementara XGBoost memetakan vektor *embedding* ke dalam kelas bahan pangan melalui pendekatan ensemble berbasis pohon keputusan.

Pendekatan hibrida ini memisahkan proses representasi visual dan proses pengambilan keputusan klasifikasi. Pada domain citra makanan, sebagian besar studi masih menggunakan arsitektur CNN end-to-end atau Transformer dengan lapisan klasifikasi linear. Integrasi Vision transformer dan XGBoost sebagai *classifier* pada konteks bahan pangan masih relatif terbatas dalam literatur yang ditinjau.

d. Integrasi TKPI sebagai Lapisan Pengetahuan Deterministik

Aspek kebaruan lain terletak pada integrasi Tabel Komposisi Pangan Indonesia sebagai basis pengetahuan gizi dalam sistem. Banyak penelitian terdahulu merujuk pada basis data nutrisi internasional atau menggunakan anotasi kalori langsung dalam dataset. Penelitian ini secara eksplisit menggunakan TKPI sebagai referensi resmi nasional untuk menghitung nilai energi setelah bahan pangan teridentifikasi.

Dengan pendekatan ini, sistem tidak hanya melakukan analisis visual, tetapi juga mengaitkannya dengan standar gizi nasional secara terstruktur. Perhitungan energi dilakukan menggunakan rumus proporsional berbasis berat aktual dan nilai kalori referensial, sehingga sistem menggabungkan komponen berbasis pembelajaran mesin dan komponen berbasis pengetahuan (knowledge-based).

2.4 Parameter Visual Bahan Pangan

Sebagai landasan dalam perancangan sistem klasifikasi bahan pangan, penelitian ini merangkum karakteristik visual setiap kelas bahan pangan berdasarkan penelitian terdahulu. Karakteristik tersebut meliputi aspek morfologi (bentuk), warna, dan tekstur yang umum digunakan untuk mendeskripsikan perbedaan visual antarobjek bahan pangan. Selain karakteristik visual, beberapa penelitian juga melaporkan representasi kuantitatif yang digunakan untuk mengukur karakteristik tersebut, seperti ukuran dimensi, rasio bentuk, nilai pada ruang warna (RGB, HSV) serta parameter tekstur. Ringkasan karakteristik visual beserta representasi kuantitatif dari masing-masing kelas bahan pangan disajikan pada Tabel 2.2.

Tabel 2. 2 Parameter Bahan Pangan

Kelas	Karakteristik Visual	Representasi Parameter
Beras Putih	Berwarna putih keabuan terang, intensitas warna pucat dan homogen. (Frengky Hanas et al., 2026)	RGB memiliki rentang nilai 0–255 pada setiap kanal warna, sedangkan HSV terdiri atas Hue (0–360°) serta Saturation dan Value (0–255). (Tanaka et al., 2023)
Telur Ayam Ras	Berwarna cokelat, permukaan bertekstur halus dengan ketajaman rendah. Berbentuk oval. (Prasetia, 2022)	RGB berada pada rentang R (103–206), G (67–189), dan B (41–182). HSV memiliki rentang H (21–151), S (50–183), dan V (116–213). Nilai kontras berada pada rentang 35–76. Sharpness (22.74–1416.13) (Maimunah et al., 2017)

Kelas	Karakteristik Visual	Representasi Parameter
Kentang Segar	Umbi berbentuk oval hingga lonjong, Berwarna coklat keabuan khas kulit kentang (Zarzecka et al., 2024)	RGB berada pada rentang R (100–180), G (70–140), dan B (20–100). Hue (H) berada pada rentang 20–30°, sedangkan Saturation (S) dan Value (V) menggunakan rentang 0–255. (Feldman et al., 2024b)
Wortel Segar	Umbi akar berbentuk silindris hingga kerucut memanjang, Berwarna jingga (oranye) dengan permukaan bertekstur kasar dan guratan yang terlihat jelas. (Sitkey et al., 2024).	RGB berada pada rentang R (112–241), G (67–203), dan B (20–186). HSV berada pada rentang H (9–110°), S (65–221), dan V (115–241). (microbiology et al., 2024)
Bawang Merah	Umbi berbentuk bulat hingga elips, berlapis (papery scales), berwarna merah keunguan hingga ungu tua mengilap. (Akbudak et al., 2020)	RGB memiliki rentang 0–240, Hue (H) berada pada rentang 0–360°, dan kontras berada pada rentang 0–100. (Kim et al., 2018)
Bawang Putih	Berwarna putih tulang kusam cenderung krem kekuningan, tingkat ketajaman tekstur permukaan sedang.	RGB memiliki rentang 0–180, HSV berada pada rentang 0–179, dan kontras berada pada rentang 1–128 (Mengistu et al., 2024)
Cabai Merah	Buah berbentuk memanjang dan ramping, berujung runcing, berpermukaan halus mengilap, serta berwarna merah cerah saat matang (Gurung et al., 2020; Hendrawan et al., 2021).	RGB dengan rentang 0–255, HSV (H, S, dan V) pada rentang 0–255, serta kontras yang berkisar dari 0 (gelap) hingga nilai tinggi (terang) untuk menggambarkan variasi warna dan intensitas pada citra. (Gurung et al., 2020b)
Jagung Kuning	Warna biji jagung manis berwarna kuning, berbeda dengan jagung biasa yang berwarna oranye cerah (biji jagung manis memiliki warna kuning berbeda dengan jagung biasa yang memiliki warna orange cerah (Astiningrum et al., 2020).	RGB berada pada rentang R (30–243), G (24–234), dan B (14–207), yang menunjukkan variasi warna yang luas. HSV berada pada rentang H (16–166°), S (43–207), dan V (30–243), yang merepresentasikan warna dari jingga hingga hijau dengan tingkat kejenuhan dan kecerahan yang bervariasi. (Zakariya et al., n.d.)

Kelas	Karakteristik Visual	Representasi Parameter
Tomat Merah	Berwarna merah terang, tekstur permukaan halus. Bentuk bulat/oval, tanpa hijau signifikan (Hidayatulloh & Riyanto, 2020).	RGB berada pada rentang R (51–233), G (26–208), dan B (23–195), yang menunjukkan variasi warna yang luas. HSV berada pada rentang H (17–270°), S (46–218), dan V (56–233), Circularity (0.75 – 0.95), Aspect Ratio (0.85 – 1.15) (Nahiduzzaman et al., 2017)
Terung Ungu	Warna kulit ungu mendominasi (terutama terong bulat ungu), permukaan mengkilap, bentuk bulat/oval. (Arie Rahady & Dhiya Ayuni, 2024)	R (23,64–213,51), G (21,80–205,99), dan B (18,56–210,35). HSV berada pada rentang H (24,88–287,90°), S (17,79–143,65), dan V (24,08–213,86) (Rahady et al., n.d.)

Tabel 2.2 merangkum karakteristik visual dan representasi parameter dari sepuluh kelas bahan pangan yang digunakan pada penelitian ini.

