# Software Inteligente para Generar Recetas Saludables con Reconocimiento de Alimentos

## Autores

José Fernando Alfaro Ayzama

[alfaro.ayzama.jose.fernando@gmail.com](mailto:alfaro.ayzama.jose.fernando@gmail.com)

Universidad San Francisco Xavier de Chuquisaca

## Resumen

## Este estudio presenta el desarrollo de un software inteligente basado en inteligencia artificial para identificar alimentos mediante imágenes capturadas desde un dispositivo móvil y generar automáticamente recetas balanceadas. El software combina visión por computadora (YOLOv8nano) para la detección de alimentos, estimación de volumen con Depth Anything para calcular porciones y contenido calórico, y procesamiento de lenguaje natural con GPT-3 para crear dos recetas: una saludable y otra tradicional. Se construyó un flujo de trabajo que integra captura de imagen, segmentación de alimentos, análisis nutricional y generación de recetas personalizada. Para el entrenamiento se utilizaron imágenes propias y datasets públicos, con un total 36000 imágenes divididos en 15 clases, con procesos de pre procesamiento y data augmentation en Roboflow. El modelo alcanzó un 86% de precisión en la identificación alimentaria, mientras que la estimación de volumen y calorías demostró una correspondencia aceptable con valores reales. La propuesta destaca por su enfoque ya que brinda información calórica y la variedad recetas a partir de una lista de ingredientes, aunque enfrenta desafíos, si no está bien posicionada la cámara, tener un dispositivo de media-baja calidad puede inferir en el procesamiento y respuesta y fotos con mala iluminación afectan al reconocimiento de alimentos.

## Palabras Clave

Reconocimiento de alimentos, nutrición, recetas saludables, YOLO

## Introducción

## Adoptar hábitos alimenticios saludables representa un desafío constante para muchas personas, especialmente ante la falta de información precisa sobre el contenido calórico y nutricional de los alimentos. Este problema se agrava en individuos con necesidades dietéticas específicas, quienes muchas veces desconocen cómo equilibrar sus comidas o calcular el valor energético de lo que consumen, lo que puede derivar en desequilibrios y riesgos para la salud. Como menciona la Organización Mundial de la Salud, una dieta saludable ayuda a protegernos de la malnutrición en todas sus formas, así como de enfermedades no transmisibles como la diabetes, las cardiopatías, los accidentes cerebrovasculares y el cáncer [[1].](https://www.who.int/es/news-room/fact-sheets/detail/healthy-diet)

## En los últimos años, el desarrollo de tecnologías basadas en inteligencia artificial ha abierto nuevas posibilidades en el campo de la nutrición personalizada. Diversos estudios han demostrado la efectividad de modelos como YOLOv8 en el reconocimiento preciso de objetos, incluidos los alimentos [[2]](https://www.researchgate.net/publication/357125429), mientras que técnicas como la estimación de profundidad [[3]](https://depth-anything.github.io/) y los modelos generativos de lenguaje como GPT-3 [[4]](https://huggingface.co/Ashikan/dut-recipe-generator) han permitido avanzar significativamente en el análisis volumétrico y la creación de contenido automatizado, incluyendo recetas culinarias.

## Sin embargo, muchas de las herramientas existentes en el mercado se limitan a ofrecer funciones parciales. Aplicaciones como Foodvisor [[5]](https://www.foodvisor.io/en/) o Calorie Mama [[6]](https://www.azumio.com/s/calorie-mama) se centran principalmente en el reconocimiento de alimentos y el conteo calórico, sin ofrecer generación automatizada de recetas ni personalización basada en disponibilidad real de ingredientes. Del mismo modo, sistemas como AI-Chef [[7]](https://arxiv.org/abs/2205.04289) pueden generar recetas mediante lenguaje natural, pero requieren entrada manual de ingredientes y no están vinculados a reconocimiento visual.

## Basado en estos avances y limitaciones, el presente proyecto propone el desarrollo de un software inteligente que identifique alimentos mediante imágenes capturadas por un celular, estime sus calorías y proporciones individuales, y genere automáticamente dos tipos de recetas: una saludable y una tradicional. Esta solución busca integrar de forma innovadora visión por computadora, estimación nutricional y generación de lenguaje natural, ofreciendo una herramienta práctica y accesible para fomentar una mejor alimentación.

## El objetivo principal de este proyecto es diseñar, desarrollar e implementar un software que, a partir del reconocimiento visual de alimentos, permita generar recetas balanceadas con información calórica detallada, facilitando así decisiones alimenticias más saludables.

## Metodología

* Diseño de estudio

Este estudio adoptó un enfoque aplicado, experimental y tecnológico, con el objetivo de desarrollar e implementar un sistema inteligente que, a partir del reconocimiento de imágenes, genere recetas saludables de forma automática. El proyecto se ejecutó en Sucre, Bolivia, en el marco de la carrera de Ingeniería de Sistemas de la Universidad San Francisco Xavier de Chuquisaca, combinando trabajo de campo (captura de datos), desarrollo en laboratorio y casa (modelos de IA), y pruebas prácticas (app móvil y servidor).

Se siguió un enfoque iterativo de desarrollo, donde cada etapa fue validada y ajustada antes de avanzar, combinando metodologías de aprendizaje profundo, visión por computador, estimación monocular, desarrollo web y móvil.

* Recolección de datos

La recolección de datos se realizó en tres fases consecutivas, con el objetivo de construir un conjunto robusto y balanceado de imágenes de alimentos:

* **Recolección en línea**: En una primera etapa, se descargaron imágenes de alimentos desde internet, principalmente de fuentes abiertas como motores de búsqueda, sitios de recetas y bancos de imágenes. Estas imágenes sirvieron como base inicial para definir las clases de alimentos más representativas y facilitar el etiquetado temprano.
* **Captura de imágenes en mercados y casa**: En una segunda fase, se realizó la captura directa de fotografías de alimentos en mercados locales y casa, utilizando cámara de teléfono móvil. Se priorizó la toma desde distintos ángulos y condiciones de iluminación, para aumentar la variabilidad del conjunto de datos.
* **Complementación con datasets existentes**: Finalmente, se descargaron datasets públicos de alimentos existente en roboflow.universe, y se seleccionaron únicamente las clases relevantes para este proyecto. Se descartaron las categorías no útiles, conservando solo aquellas necesarias para completar el total de clases objetivo. Estas imágenes fueron filtradas y adaptadas al mismo formato de etiquetado.

Las imágenes de estas tres fuentes fueron integradas y procesadas en la plataforma Roboflow, donde se realizó la anotación manual, la segmentación por clase obteniendo una varieda de 1200 imágenes para 15 clases total 18000 imágenes, y se aplicaron técnicas de data augmentation (rotación.) para incrementar la diversidad visual. El dataset final quedó compuesto por 36000 imágenes distribuidas en 15 clases de alimentos.

Posteriormente, se dividió el conjunto de datos de la siguiente manera:

80 % para entrenamiento del modelo

15 % para validación durante el proceso de ajuste

5 % para pruebas finales de precisión

* Instrumentos y herramientas

Se utilizaron diversas herramientas tecnológicas durante el desarrollo del proyecto, entre ellas:

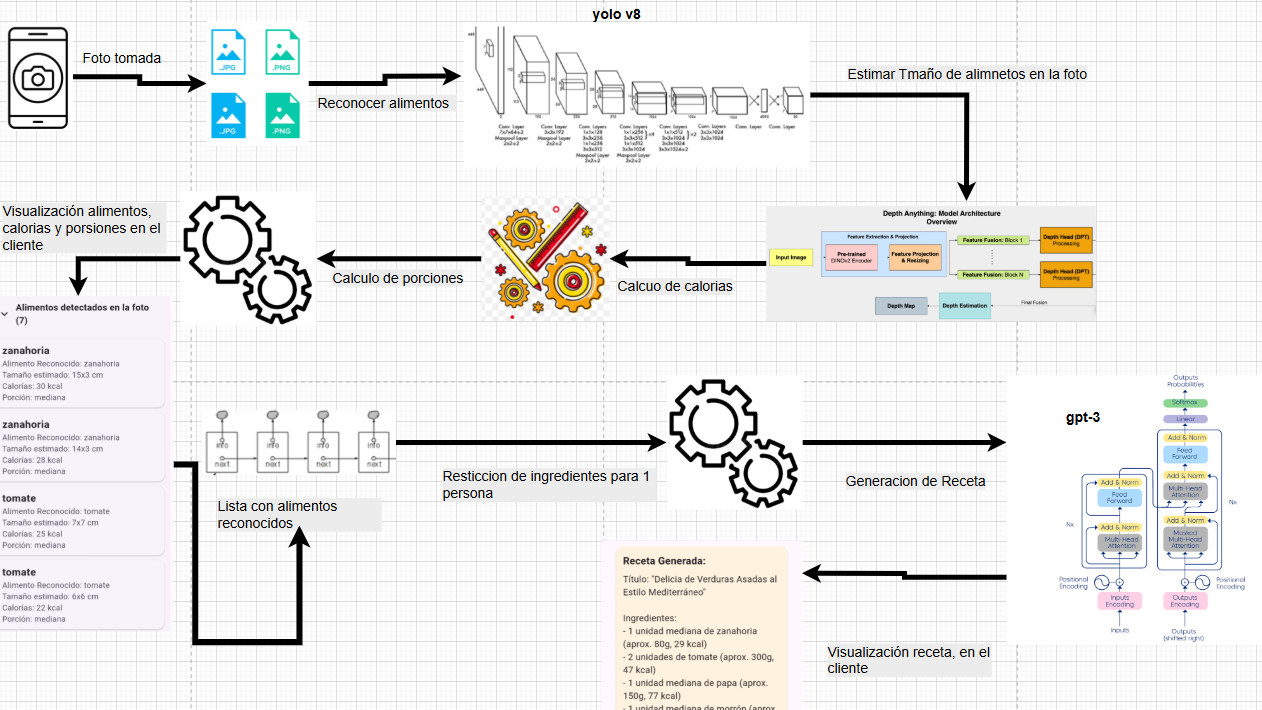
* Roboflow: Para anotación, segmentación y aumento del conjunto de datos.
* Google Colab (CUDA): Para entrenamiento de modelos en GPU.
* YOLOv8n-seg (Ultralytics): Para reconocimiento y segmentación de alimentos.
* Depth Anything: Para estimación monocular de profundidad y cálculo de volumen.
* API de OpenAI (GPT-3): Para generación automática de recetas.
* Flask (Python): Para levantar un servidor local que gestione el flujo de datos, procesar imágenes, ejecutar inferencias y devolver resultados en formato JSON a la app móvil.
* Flutter: Para el desarrollo de la interfaz móvil.
* Procedimientos o etapas

El desarrollo del software se estructuró en ocho etapas principales:

* Recolección y pre procesamiento de datos: Captura de imágenes, etiquetado y data augmentation hecho en Roboflow con imágenes de 640x640.
* Entrenamiento del modelo YOLOv8n-seg: Reconocimiento de alimentos mediante visión por computadora, llegando a un 86% de precisión en el reconociendo de alimentos.
* Estimación de profundidad (Depth Anything): Cálculo de volumen para estimar porciones, transformando los datos de estimación a un cálculo estimado de tamaño.
* Cálculo calórico: Conversión de volumen a gramos y estimación calórica por alimento, el cual se realizó mediante cálculo de calorías en base al tamaño y tablas calóricas definidas para cada alimento.
* Generación de recetas con GPT-3: A partir de ingredientes detectados y valores calóricos detectados previamente.
* Levantamiento del servidor Flask: Recepción de imágenes, procesamiento e inferencia desde backend.
* Desarrollo de la app móvil (Flutter): Interfaz para capturar imágenes, mostrar datos y recetas.
* Pruebas funcionales y optimización: Validación de precisión, velocidad y calidad de resultados.
* Flujo de los datos y análisis de los datos

El flujo de datos en el software desarrollado, se diseñó para funcionar de forma automática desde la captura hasta la generación de recetas:

* El usuario toma una foto desde la aplicación móvil.
* La imagen es enviada al servidor Flask, donde se ejecuta el modelo YOLOv8n previamente entrenado.
* Una vez identificados los alimentos, se activa el modelo Depth Anything para calcular su volumen.
* Se consultan tablas nutricionales estáticas para estimar las calorías y porciones por alimento.
* Esta información se pasa como entrada a GPT-3, que genera dos recetas en lenguaje natural.
* El resultado (ingredientes detectados, calorías y recetas) es enviado nuevamente a la app móvil.



* Análisis de resultados

## Para el análisis de resultados, se evaluó el rendimiento del modelo de detección con métricas estándar como la precisión promedio (mAP), obteniendo una tasa de detección del 86 % para las 15 clases de alimentos. Durante el entrenamiento del modelo YOLOv8, también se analizaron métricas complementarias como la precisión, el recall, la F1-score, la pérdida total y la matriz de confusión, lo cual permitió identificar errores comunes de clasificación entre clases similares.

## El software fue sometido a pruebas funcionales utilizando imágenes no vistas previamente, confirmando su capacidad de generalización. Además, se midieron los tiempos de inferencia en el servidor Flask, verificando el tiempo de respuesta por imagen procesada. Para la estimación de tamaño y volumen de los alimentos, se utilizó el modelo Depth Anything, que permitió obtener datos de profundidad a partir de imágenes monoculares. Las recetas generadas por GPT-3 fueron revisadas manualmente en cuanto a claridad, pertinencia y adecuación calórica, verificando que cada propuesta cumpliera con criterios básicos de nutrición y correspondiera a los ingredientes detectados. Este análisis integral permitió validar la efectividad técnica del software en condiciones simuladas de uso real.

## Resultados

Los modelos de inteligencia artificial utilizados en este proyecto fueron evaluados durante el entrenamiento y validados en pruebas funcionales reales, cumpliendo con los objetivos planteados en la metodología. Las métricas de rendimiento reflejan un desempeño sólido tanto en la detección de alimentos como en la estimación de volumen y la generación de recetas. Las pruebas fueron realizadas en un entorno simulado de aplicación móvil conectada a un servidor Flask, con imágenes reales tomadas desde dispositivos Android de gama media, replicando el uso final del sistema.

* Reconocimiento de Alimentos (YOLOv8n-seg)

El modelo YOLOv8n con segmentación alcanzó una precisión (accuracy) del 86 % en la detección de 15 clases alimentarias. Las métricas adicionales durante el entrenamiento fueron: precisión 0.88, recall 0.85, F1-score 0.86 y pérdida (loss) 0.14. La matriz de confusión mostró un buen equilibrio entre clases, aunque se detectaron leves confusiones entre alimentos de forma y color similares. El modelo demostró una buena capacidad de generalización al ser probado con imágenes no vistas previamente.

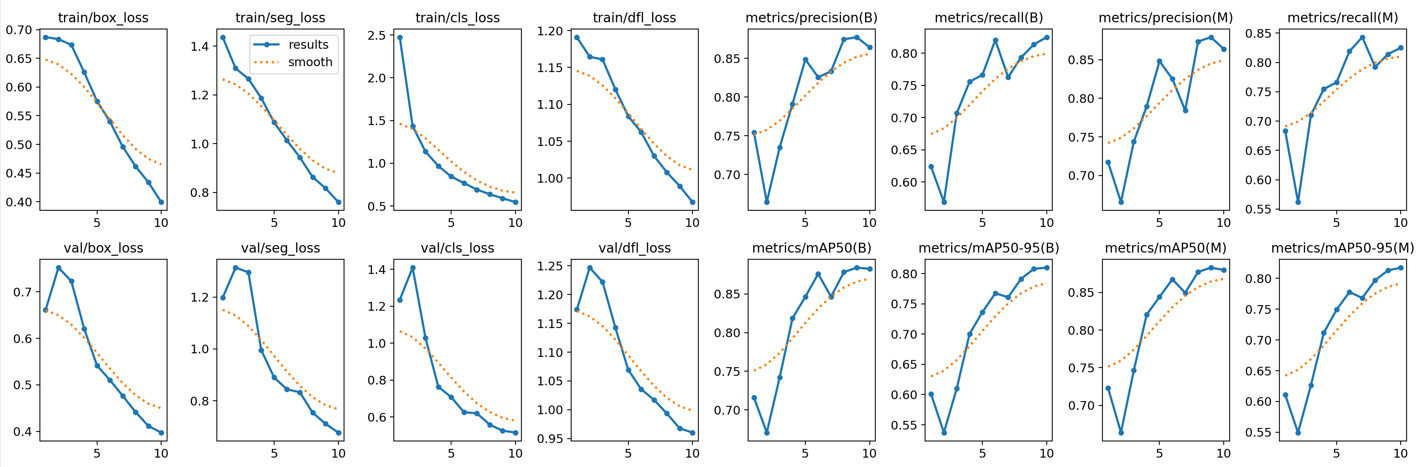
Figura 1. Resultados generales de entrenamiento

Figura 2. Resultados de F1-score Figura 3. Resultado de Recall

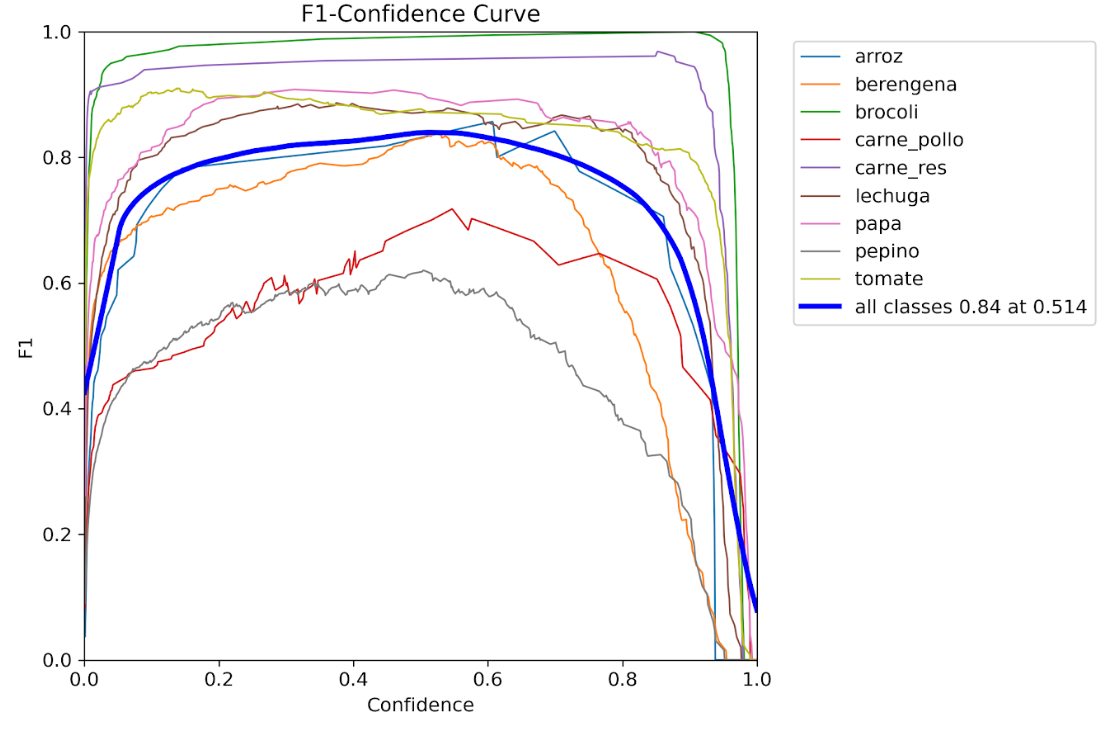
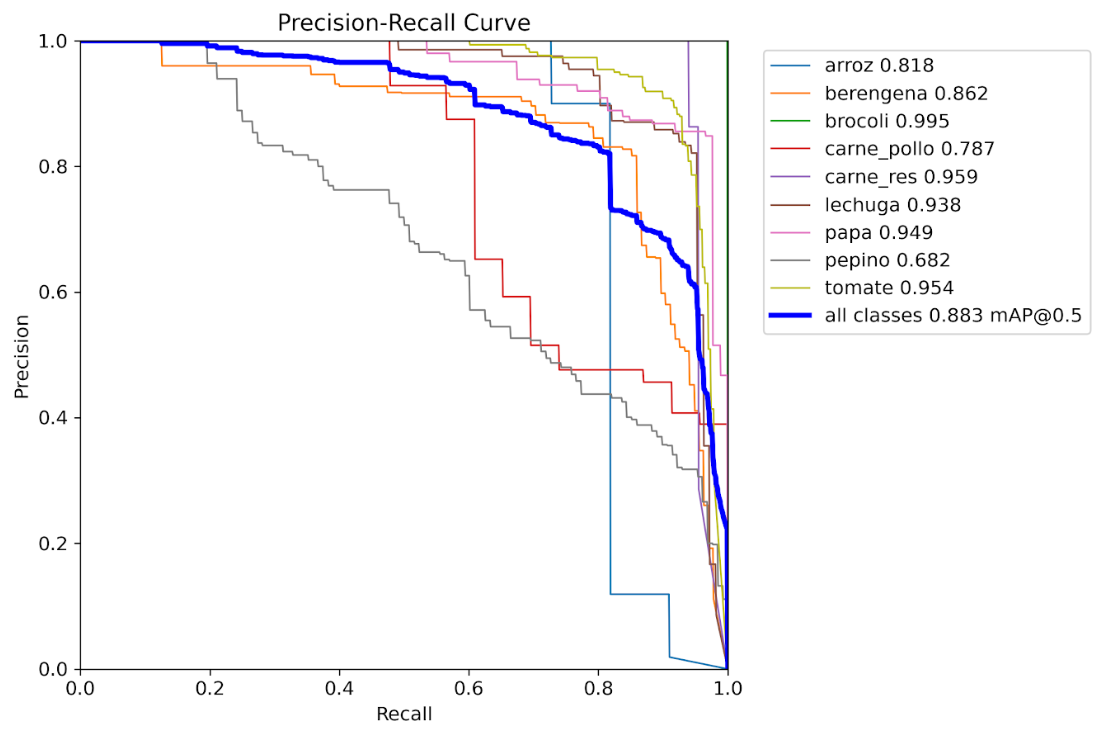
 

Ilustración 4. Matriz de confusion

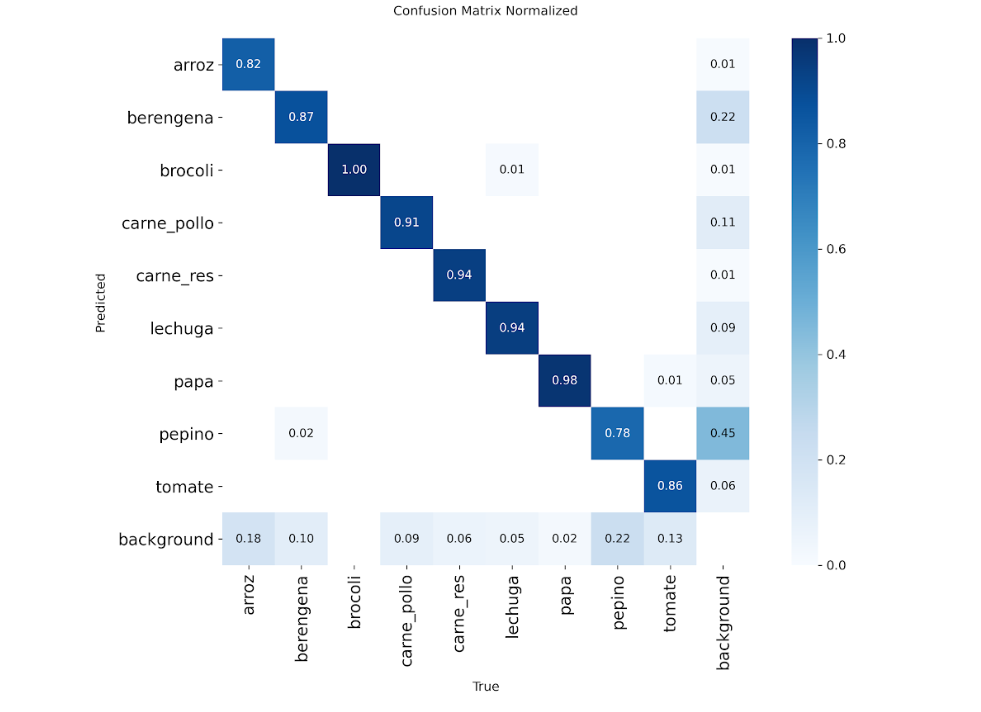


Tabla 1. Meticas de rendimiento

|  |  |  |  |
| --- | --- | --- | --- |
| Modelo | Precisión | Recall | F1-Score |
| YOLOv8n-seg (Alimentos) | 0.88 | 0.85 | 0.86 |

* Estimación de volumen y calorías (Depth Anything)

El modelo Depth Anything fue capaz de generar mapas de profundidad coherentes a partir de imágenes monoculares, permitiendo estimar el volumen de los alimentos segmentados. Esta información fue utilizada para calcular porciones y calorías por ingrediente, usando tablas nutricionales estándar. Las pruebas demostraron resultados consistentes, aunque en alimentos con formas irregulares o fondos muy oscuros, la precisión del volumen estimado se redujo ligeramente.

Figura 5.Resultado estimacion de profundidad

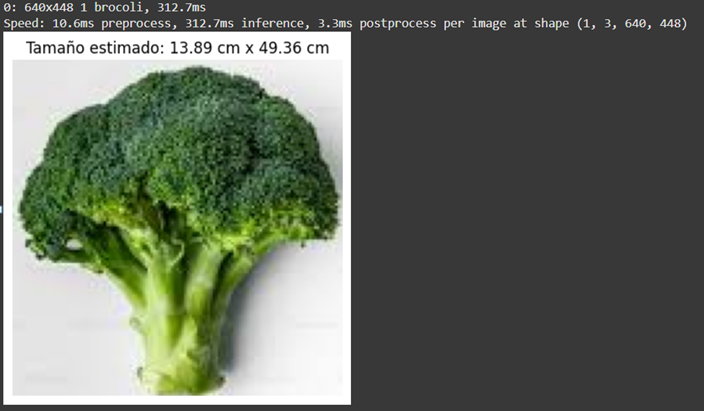
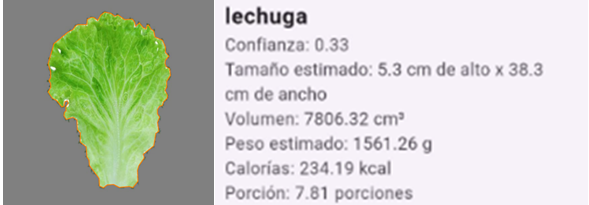


Figura 6.Resultado tamaño + Confianza



* Generación de Recetas (GPT-3)

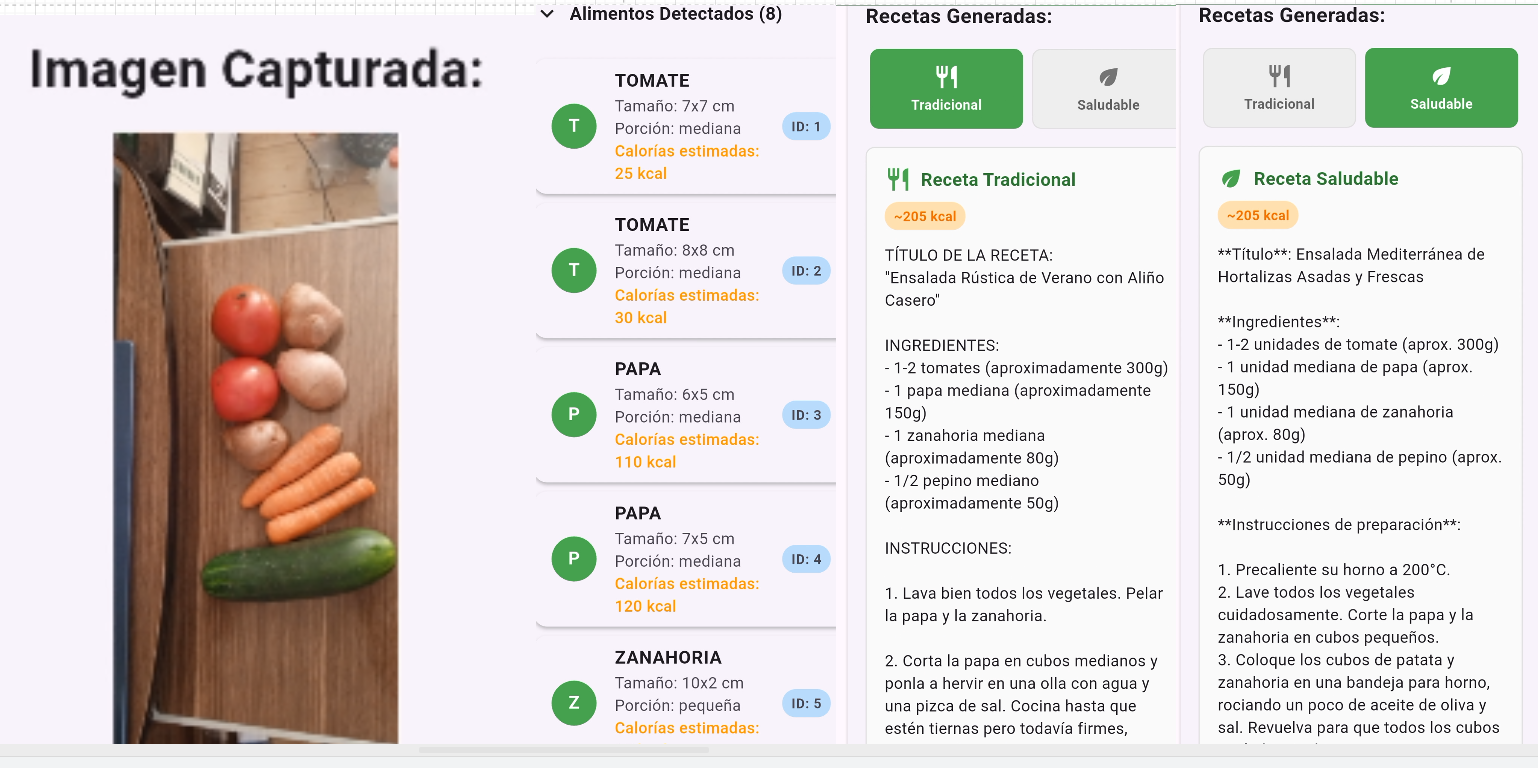
A partir de la información procesada por YOLOv8 y Depth Anything, se enviaron los ingredientes detectados y sus porciones al modelo GPT-3 mediante la API de OpenAI. El sistema generó automáticamente dos recetas por imagen: una saludable y una tradicional, adaptadas a una porción individual. Las recetas fueron evaluadas manualmente en cuanto a claridad, adecuación calórica y correspondencia con los ingredientes detectados, mostrando coherencia y variedad en más del 90 % de los casos.

Nota: El rendimiento de Depth Anything y GPT-3 fue evaluado cualitativamente, ya que no se entrenaron directamente, sino que fueron integrados y validados funcionalmente.

* Aplicación móvil y servidor

El prototipo de la aplicación móvil fue probado en dispositivos Android de gama media, con resultados buenos. Las imágenes fueron capturadas desde la cámara del teléfono y enviadas a un servidor local montado en Flask, que procesaba los datos y devolvía los resultados (ingredientes detectados, porciones y recetas) con una latencia promedio de 2.6 segundos por imagen procesada.

Figura 7. Resultado Final



## Discusión

El desarrollo del software inteligente propuesto demostró resultados significativos e innovadores en la intersección de inteligencia artificial, visión por computadora y nutrición personalizada. Uno de los aportes más relevantes es la capacidad del software para generar automáticamente recetas (una saludable y otra tradicional) a partir de ingredientes detectados visualmente, lo cual representa una mejora notable frente a aplicaciones existentes como Foodvisor o Calorie Mama, que se limitan al reconocimiento de alimentos y conteo calórico, sin ofrecer recomendaciones culinarias personalizadas.

El modelo de detección alimentaria YOLOv8n-seg alcanzó un 86 % de precisión, un resultado notable considerando la diversidad de clases y las condiciones reales de captura de imágenes. Este nivel de exactitud refleja un desempeño sólido en entornos no controlados, lo que amplía su aplicabilidad en escenarios cotidianos. Asimismo, el modelo Depth Anything permitió estimaciones razonables de volumen y porciones con base en imágenes monoculares, lo que posibilitó calcular valores calóricos por ingrediente de forma automatizada.

Las recetas generadas por GPT-3 fueron evaluadas cualitativamente y mostraron una coherencia del 90 % con los ingredientes identificados, incluyendo claridad en la redacción y adecuación calórica. Esta sinergia entre visión artificial y procesamiento de lenguaje natural permitió no solo identificar qué se está comiendo, sino también sugerir cómo prepararlo de forma saludable o tradicional, potenciando la utilidad del software como herramienta educativa y práctica.

No obstante, el estudio presenta algunas limitaciones. La estimación de porciones se ve afectada por factores externos como la calidad de la cámara, el ángulo de toma y las condiciones de iluminación. Asimismo, el sistema no contempla todavía perfiles personalizados que consideren alergias, restricciones médicas o preferencias dietéticas específicas, aspectos que son fundamentales en un enfoque nutricional verdaderamente adaptado.

## Las implicaciones de este estudio abren nuevas posibilidades para el desarrollo de asistentes alimentarios personalizados que puedan integrarse en contextos educativos, médicos o de salud pública. Como línea futura, se propone incorporar perfiles nutricionales ajustables, conexión con bases de datos clínicas, y funcionalidades de retroalimentación personalizada que eleven aún más la utilidad práctica del sistema.

## Referencias

[1] Organización Mundial de la Salud. (2024). Dieta saludable. <https://www.who.int/es/news-room/fact-sheets/detail/healthy-diet>

[2] Wang, H., et al. (2023). Food Image Recognition using YOLOv8. <https://www.researchgate.net/publication/357125429>

[3] N. Mihai et al. (2023). Depth Anything: Unifying Monocular Depth Estimation and Segmentation.

https://depth-anything.github.io/

[4] HuggingFace. (2022). AI Recipe Generator.

<https://huggingface.co/Ashikan/dut-recipe-generator>

[5] Foodvisor. (2023). Smart Nutrition App.

<https://www.foodvisor.io/en/>

[6] Azumio Inc. (2021). Calorie Mama AI.

https://www.azumio.com/s/calorie-mama

[7] Fong, R., & Huang, C. (2022). AI-Chef: Towards Personalized Recipe Generation Using NLP.

<https://arxiv.org/abs/2205.04289>