

Development of an Artificial Vision System for the Automatic Detection of Burnt Alfajor Caps in the Post-Baking Stage

Ebert Manuel Vera Guerra ; Ryan Abraham León León 

Universidad Privada del Norte, Perú, N00202962@upn.pe, ryan.leon@upn.edu.pe

Abstract– *This study presents the development of a machine vision system based on the YOLOv11 model for the automatic detection of burnt alfajor wafers during the post-baking stage. The aim is to implement an intelligent inspection tool that optimizes quality control by automating visual classification. A dataset was constructed using images captured under controlled conditions and labeled into two categories: normal product and burnt product. The model was trained for 70 epochs, achieving an accuracy of 96.8%, sensitivity of 94.5%, F1 score of 95.6%, $mAP50 = 1.00$, and $mAP50-95 = 0.87$. The results demonstrate robust performance and adequate generalizability, proving the feasibility of integrating artificial intelligence techniques into quality control systems for the food industry, reducing manual intervention and improving operational efficiency.*

Keywords– *Automatic detection, industrial automation, machine vision, quality control, YOLOv11.*

Desarrollo de un Sistema de Visión Artificial para la Detección Automática de Tapas de Alfajor Quemadas en la Etapa PostHorneado

Ebert Manuel Vera Guerra ; Ryan Abraham León León 
Universidad Privada del Norte, Perú, N00202962@upn.pe, ryan.leon@upn.edu.pe

Resumen– Este estudio presenta el desarrollo de un sistema de visión artificial basado en el modelo YOLOv11 para la detección automática de tapas de alfajor quemadas en la etapa de post-horneado, con el propósito de implementar una herramienta de inspección inteligente que optimice el control de calidad mediante la automatización de la clasificación visual. Se construyó un conjunto de datos con imágenes capturadas en condiciones controladas y etiquetadas en dos categorías: producto normal y producto quemado. El modelo fue entrenado durante 70 épocas, alcanzando una precisión de 96.8%, sensibilidad de 94.5%, puntaje F1 de 95.6%, mAP50 = 1.00 y mAP50–95 = 0.87. Los resultados evidencian un desempeño robusto y una adecuada capacidad de generalización, demostrando la viabilidad de integrar técnicas de inteligencia artificial en sistemas de control de calidad para la industria alimentaria, reduciendo la intervención manual y mejorando la eficiencia operativa.

Palabras clave– Automatización industrial, Control de calidad, Detección automática, Visión artificial, YOLOv11.

1. INTRODUCCIÓN

En el sector de repostería industrial, la calidad visual constituye un factor crítico para la satisfacción del consumidor y la eficiencia económica. Las tapas de alfajor que presentan quemaduras durante el post-horneado representan un problema recurrente que provoca rechazo del producto, incremento de desperdicio, costos adicionales en retrabajo o descarte, y posibles repercusiones en la reputación de marca. A pesar de los controles manuales, estos presentan limitaciones tales como subjetividad, fatiga visual, variaciones entre operadores y lentitud, lo que reduce la consistencia del proceso de inspección y eleva la probabilidad de errores.

La visión artificial se ha posicionado como herramienta clave para garantizar procesos de inspección eficientes en productos alimenticios. A nivel local, se desarrolló un modelo CNN para clasificar madurez de mango Kent según la Norma Técnica Peruana NTP011.025:2023, utilizando 603 imágenes etiquetadas y logrando una precisión del 96.04%, un recall del 90.91% y un F1-score de 93.57% incluso con imágenes de baja resolución (32×32 píxeles) [1]. Complementariamente, estudios de simulación han demostrado viabilidad técnica: se propuso un sistema automatizado para clasificación de limón Tahití que optimiza significativamente el tiempo de selección comparado con métodos manuales [2], mientras que se simuló

clasificación automática para mangos peruanos de exportación (Kent, Haden y Tommy Atkins), destacando las limitaciones del proceso manual en términos de lentitud e imprecisión [3]. Enfocándose en la detección de defectos específicos, se desarrolló un sistema para detectar plagas bicho de cesto y arañita roja en cultivo de palto usando drone Mavic2 Zoom con procesamiento OpenCV, segmentación de color y Random Forest, obteniendo 100% de precisión promedio [4]. Siguiendo esta línea de detección especializada, se diseñó un sistema automático para clasificación de paltas Hass con capacidad de 1,000 kg/hora, empleando detección de madurez por proporción de colores y medida de calibre, alcanzando 92.85% de precisión [5]. A nivel nacional, los sistemas de tiempo real han demostrado alta efectividad en condiciones operativas: se desarrolló inspección visual en tiempo real para gradado de frutas usando modelos ResNet, DenseNet, MobileNetV2, NASNet y EfficientNet, alcanzando en pruebas estáticas 99.2% para manzanas y 98.6% para bananas, mientras que en tiempo real se obtuvieron 96.7% y 93.8% respectivamente [6]. Expandiendo estas capacidades hacia sistemas integrados, se creó una máquina clasificadora automática de cítricos integrando lavado, clasificación visual y gradación por peso mediante CNN adaptativo SortNet, logrando 97.0% de precisión en visión y 91.3% en gradación por peso [7].

Profundizando en la medición dimensional, las técnicas de estimación dimensional han evolucionado significativamente: se desarrolló un sistema con cámara de profundidad para estimar calibre de manzanas en árbol usando YOLOv5 personalizado, logrando 92% de tasa de detección, correlaciones $r > 0.8$ entre tamaño estimado y real, con error medio de +5.7 mm (~9%) y RMSE de 10 mm (~15%) [8]. Avanzando hacia detección multifuncional, se propuso detección en línea de madurez y diámetro de manzanas Fuji usando YOLOv5s mejorado con módulos VoVGSCSP y GSConv, alcanzando 98.7% de precisión promedio para detección de madurez, incluyendo 97.4% para bajo grado de madurez, y 93.3% de éxito en estimación de diámetro con desviación media de 0.878 mm operando a ~56 fps [9]. Comparando metodologías, los enfoques de aprendizaje profundo han demostrado superioridad sobre métodos tradicionales: se investigó clasificación de madurez en banano

usando 1,565 imágenes RGB con Dual CNN, donde la CNN alcanzó 90.42% de precisión para clasificar estadios "unripe", "ripe" y "overripe", mientras VGG logró 89.22%, superando algoritmos tradicionales como Árbol de Decisión (71.86%), Random Forest (85.63%), KNN (86.83%) y SVM (89.22%) [10]. Globalmente, se identificó que modelos one-stage como YOLO y EfficientDet superan a los de dos etapas en rapidez; un modelo optimizado DY3TNet basado en YOLOv3-Tiny alcanzó 90.05% mAP en detección de kiwis, superando a Faster R-CNN y Mask R-CNN que lograron 80-88% [11]. Complementando con tecnologías avanzadas de análisis espectral, se analizó clasificación de manzanas mediante imágenes hiperespectrales, donde el modelo SVM obtuvo ARI de 0.789 para quemado solar, 0.818 para sarna y 0.854 para podredumbre, con precisiones globales superiores al 95% [12]. Explorando aplicaciones de precisión espacial, se desarrolló ALACS (Active Laser-Camera Scanning) para localización 3D de frutas, alcanzando precisiones de 6.9-11.2 mm comparado con 21.5 mm de cámara RealSense comercial, logrando 95% de tasa de detección exitosa frente al 71% del sistema de referencia [13]. Integrando múltiples dimensiones de análisis, se propuso mapeo panóptico integrando completado de formas 3D y estimación de pose, demostrando mejoras del 41% en precisión de completado y 52% en estimación de pose con tiempos de inferencia de 0.6 segundos [14]. Finalmente, en aplicaciones de detección temprana, se desarrolló detección de flores de manzano usando YOLOv5 en condiciones no estructuradas, alcanzando 81.9% mAP para detección de flores abiertas y cerradas [15].

Se identifica una brecha significativa en la literatura: escasez de aplicaciones específicas para productos de repostería industrial donde las características de textura, color y apariencia difieren sustancialmente de frutas y hortalizas; necesidad de sistemas que operen eficientemente bajo condiciones de iluminación y ángulos variables propias de líneas de producción; y ausencia de estudios que comparen cuantitativamente el desempeño de visión artificial versus inspección manual en términos de velocidad, precisión y costos operativos para este tipo de productos. La selección de YOLOv11 como arquitectura base se fundamenta en sus características técnicas específicamente adecuadas para la detección de variaciones cromáticas y texturales asociadas al quemado de productos horneados: a diferencia de arquitecturas de dos etapas (R-CNN, Faster R-CNN) que primero generan propuestas de regiones y luego las clasifican, YOLOv11 implementa un pipeline unificado que realiza detección y clasificación en una sola pasada, alcanzando velocidades compatibles con líneas de producción industrial; técnicamente, YOLOv11 incorpora bloques convolucionales C3 mejorados con atención espacial (SPP-F) que permiten extraer características multi-escala, capacidad crítica para detectar patrones de quemado que pueden manifestarse tanto como manchas localizadas hasta áreas extensas; adicionalmente, su arquitectura Feature Pyramid Network (FPN) integrada facilita la detección de variaciones sutiles de tonalidad, aspecto fundamental para diferenciar entre grados

normales de tostado y presencia de quemado. Los principales retos técnicos asociados a la implementación en entornos reales incluyen: variabilidad en condiciones de iluminación, diversidad de ángulos de captura debido al movimiento en banda transportadora, presencia de reflejos especulares en superficie del producto, y necesidad de procesamiento en tiempo real manteniendo precisión superior al 95%.

Con base en la problemática identificada y las brechas existentes en la literatura, se plantea la siguiente hipótesis: la implementación de un sistema de visión artificial basado en YOLOv11 alcanzará una precisión mínima del 95% en la detección de tapas de alfajor quemadas bajo condiciones controladas de iluminación LED (5500K) y ángulo de captura cenital, medida mediante matriz de confusión sobre un conjunto de validación independiente. Para validar esta hipótesis, se establece como objetivo general: desarrollar e implementar un sistema de visión artificial basado en YOLOv11 capaz de detectar automáticamente tapas de alfajor quemadas en la etapa post-horneado. Los objetivos específicos son: (1) construir un conjunto de datos balanceado con mínimo 1,200 imágenes de tapas de alfajor bajo condiciones estandarizadas, categorizadas en estado normal y quemado; (2) entrenar y optimizar un modelo YOLOv11 que logre valores mínimos de precisión $\geq 95\%$, sensibilidad $\geq 90\%$, F1-score $\geq 92\%$, y mAP50 ≥ 0.90 sobre el conjunto de validación; (3) evaluar el desempeño del sistema mediante pruebas, midiendo precisión de clasificación, velocidad de procesamiento y tasa de error, documentando las condiciones operativas bajo las cuales se alcanzan los mejores resultados.

La relevancia de este estudio radica en su potencial para revolucionar los procesos de control de calidad en la industria panadera, contribuyendo al avance tecnológico del sector alimentario peruano y posicionándolo competitivamente en mercados internacionales exigentes. La implementación optimizará el proceso de selección, reduciendo tiempo de inspección y error humano, incrementando calidad del producto final y disminuyendo pérdidas por descartes. El proyecto demuestra la viabilidad de integrar técnicas de aprendizaje profundo en la industria panadera, alineándose con la Industria 4.0, generando ahorros en costos operativos y facilitando recopilación de datos para futuras mejoras del sistema de horneado.

2. MATERIALES Y METODOS

El desarrollo del sistema de visión artificial para la detección automática de tapas de alfajor quemadas en la etapa post-horneado se estructuró cuidadosamente en cuatro fases principales: Captura de imágenes, procesamiento, entrenamiento del modelo y validación de resultados, cada una diseñada con el propósito de garantizar la precisión del sistema, la calidad de los datos utilizados y la fiabilidad del modelo durante su implementación en condiciones reales de producción. Esta estructura metodológica permitió asegurar la coherencia entre la adquisición de datos, la configuración del modelo YOLOv11 y la evaluación cuantitativa del

desempeño, minimizando sesgos experimentales y fortaleciendo la trazabilidad del proceso de desarrollo.

2.1 Adquisición de Imágenes

En la primera fase, correspondiente a la captura de imágenes, se definieron las condiciones óptimas para la adquisición de datos visuales del producto. Se estableció un entorno controlado dentro del área de post-horneado, donde las tapas de alfajor presentan sus características finales de color y textura; estas variaciones visuales, que reflejan el nivel de cocción, constituyen los principales indicadores que el sistema debía aprender a identificar.

En la Figura 1 se presentan las tapas de alfajor utilizadas para el entrenamiento del modelo, clasificadas visualmente según su nivel de cocción. Las imágenes utilizadas se capturaron en la etapa post-horneado del proceso productivo de tapas de alfajor.



Figura 1. Estados de cocción de tapas de alfajor

Para ello, se empleó una cámara digital de 12 megapíxeles montada sobre un trípode fijo, ubicado a una distancia de 25 cm del área de observación; la superficie de colocación fue de color blanco mate, elegida intencionalmente para eliminar reflejos, sombras y distorsiones de color. El sistema de iluminación empleado consistió en fuentes LED frontales con una temperatura de color de 5500 K, equivalente a la luz diurna natural; este tipo de iluminación fue seleccionada para mantener la homogeneidad cromática en todas las imágenes, evitando variaciones producidas por luces ambientales o por cambios de intensidad.

En la Figura 2 se presenta el esquema del sistema de iluminación frontal utilizado para la captura de tapas de alfajor. La configuración de captura se mantuvo constante para asegurar uniformidad en el conjunto de datos.



Figura 2. Esquema del sistema de iluminación frontal utilizado para la captura de tapas de alfajor

La relación entre iluminación, reflectancia e intensidad de píxel se describe mediante la siguiente expresión:

$$I(x, y) = R(x, y) * L(x, y) \quad \text{Ec. (1)}$$

Donde: $I(x,y)$ = la intensidad de cada píxel de la imagen capturada; $R(x,y)$ = la reflectancia de la superficie de la tapa; $L(x,y)$ = la iluminación incidente. Este principio físico fue tomado como base para comprender cómo la variación de color en la superficie del alfajor depende directamente de la cantidad de luz reflejada y del nivel de tostado del producto. Considerando esto, las tapas más quemadas reflejan una menor cantidad de luz visible debido a su tonalidad más oscura, mientras que las adecuadamente horneadas mantienen un equilibrio entre absorción y reflexión lumínica; este comportamiento fotométrico sirvió para ajustar el umbral de contraste durante la fase de segmentación.

En total, se recolectaron 1200 imágenes distribuidas de forma equitativa entre tres categorías de cocción: adecuadamente horneadas, ligeramente quemadas y completamente quemadas. Esta distribución balanceada permitió mejorar el proceso de entrenamiento del modelo, evitando el sesgo hacia alguna categoría específica. Todas las imágenes se almacenaron en formato JPG y fueron redimensionadas a una resolución uniforme de 640×640 píxeles con nombres codificados según su categoría ("normal_001", "quemada_002"), lo que facilitó su posterior identificación en el entrenamiento.

2.2 Preprocesamiento de Imágenes

En la segunda fase se realizó el preprocesamiento de las imágenes con el fin de mejorar la calidad y uniformidad de los datos antes del entrenamiento del modelo.

Las imágenes fueron convertidas al espacio RGB, redimensionadas a 640×640 píxeles, filtradas mediante un kernel gaussiano 3×3 para la reducción de ruido y sometidas a técnicas de aumento de datos que incluyeron rotaciones, traslaciones y variaciones de brillo de ±15%, con el fin de mejorar la generalización del modelo.

2.3 Arquitectura del modelo YOLOv11

La detección automática del estado de cocción de las tapas de alfajor se realizó mediante la arquitectura YOLOv11, la cual integra en un solo proceso la extracción de características, la localización de objetos y la clasificación.

El modelo está compuesto por tres componentes principales: el backbone, encargado de la extracción de características a partir de la imagen de entrada; el neck, que fusiona la información en diferentes escalas para mejorar la detección de objetos; y el head, responsable de predecir las cajas delimitadoras, las clases y las probabilidades asociadas.

En las capas convolucionales del modelo se emplea la función de activación SiLU (Sigmoid Linear Unit), la cual

introduce no linealidad y mejora la capacidad de aprendizaje de representaciones complejas.

La operación de convolución utilizada en la etapa de extracción de características se representa mediante la Ecuación 2.

$$Y = \sum_{i=-k}^k \sum_{j=-k}^k X(i,j) \cdot H(i,j) \quad \text{Ec. (2)}$$

Donde: $X(i,j)$ representa la matriz de la imagen de entrada; $H(i,j)$ corresponde al kernel de convolución o filtro utilizado para extraer características específicas.

El kernel gaussiano de tamaño 3×3 presentado en la Figura 3 fue seleccionado tras pruebas experimentales, ya que proporcionó un equilibrio adecuado entre el suavizado de la imagen y la preservación de bordes y texturas relevantes.

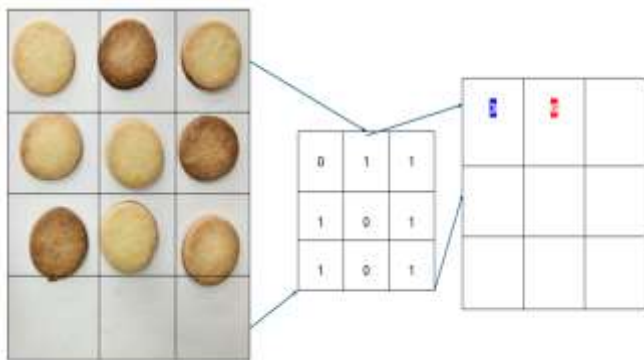


Figura 3. kernel 3×3 aplicado en el preprocesamiento de las imágenes de tapas de alfajor.

El kernel 3×3 se utilizó exclusivamente para la reducción de ruido en la etapa de preprocesamiento, permitiendo preservar bordes y texturas relevantes antes del entrenamiento del modelo. La extracción automática de características se llevó a cabo posteriormente dentro de las capas convolucionales internas de la arquitectura YOLOv11 durante el proceso de entrenamiento.

Para introducir no linealidad en el modelo YOLOv11 se empleó la función de activación SiLU (Sigmoid Linear Unit), definida mediante la Ecuación (3), cuya representación gráfica se muestra en la Figura 4:

$$f(x) = \frac{x}{1 + e^{-x}} \quad \text{Ec. (3)}$$

Donde $\text{SiLU}(x) = x \cdot \sigma(x)$.

Esta función atenúa de forma suave los valores negativos en lugar de anularlos, lo que permite una mejor propagación del gradiente y una extracción más eficiente de características. En el contexto del presente trabajo, esta función facilita la diferenciación entre tapas de alfajor normales y quemadas al resaltar variaciones de intensidad y textura en las imágenes.

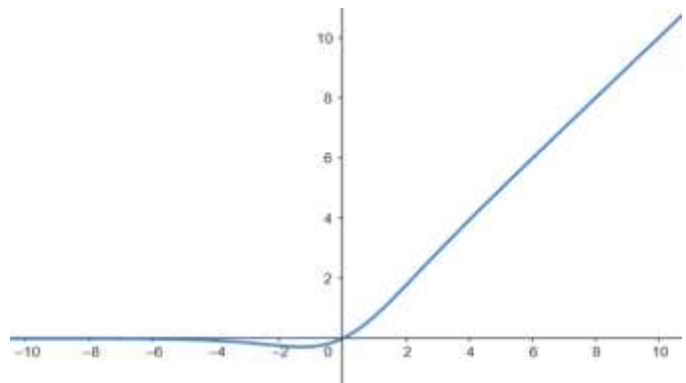


Figura 4. Función de activación SiLU aplicada en el modelo YOLOv11.

La Figura 4 muestra el comportamiento suave y continuo de la función SiLU, el cual favorece la estabilidad del entrenamiento y la preservación de información relevante en comparación con funciones de activación con umbrales abruptos.

2.4 Entrenamiento del Modelo

La estructura de actualización de parámetros mediante el optimizador Adam (Adaptive Moment Estimation) se ilustra en la Figura 5. El optimizador Adam fue clave para lograr una convergencia estable durante las 70 épocas de entrenamiento del modelo YOLOv11. Con una tasa de aprendizaje inicial de 0.001 y un tamaño de lote de 64 imágenes, Adam permitió ajustar dinámicamente los pesos de la red combinando el primer momento (media del gradiente) y el segundo momento (varianza del gradiente), lo que redujo oscilaciones y aceleró el proceso de aprendizaje. Este comportamiento se evidencia en la tendencia descendente sostenida de las curvas de pérdida (Figura 4) y en la estabilización de las métricas después de aproximadamente 50 épocas. Gracias a esta optimización adaptativa, el modelo alcanzó una precisión del 96.8%, sensibilidad del 94.5%, F1-score del 95.6%, $\text{mAP}_{50} = 1.00$ y $\text{mAP}_{50-95} = 0.87$, demostrando que Adam facilitó la correcta diferenciación entre patrones de textura y tonalidad asociados al quemado superficial de las tapas de alfajor, garantizando un entrenamiento robusto y buena capacidad de generalización. La actualización de parámetros se expresa en la ecuación 4:

$$\begin{aligned} m_t &= \beta_1 m_{t-1} + (1 - \beta_1) g_t \\ v_t &= \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \\ \hat{m}_t &= \frac{m_t}{1 - \beta_1^t}, \hat{v}_t = \frac{v_t}{1 - \beta_2^t} \\ \theta_t &= \theta_{t-1} - \alpha \frac{v_t}{\sqrt{\hat{v}_t} + \epsilon} \end{aligned} \quad \text{Ec. (4)}$$

Donde g_t corresponde al gradiente de la pérdida en la clasificación de tapas, m_t y v_t son las estimaciones de primer y segundo momento, α es la tasa de aprendizaje y ϵ un valor pequeño para evitar divisiones por cero.

3. RESULTADOS

En esta sección se presentan los resultados de validación, el modelo YOLOv11 entrenado con el Dataset de tapas de alfajor logró un desempeño satisfactorio en la detección automática de los niveles de cocción. Durante las pruebas, el sistema identificó correctamente las categorías denominadas "alfajor normal" y "alfajor quemado" en un entorno controlado de iluminación, mostrando una rápida respuesta en tiempo real. El modelo alcanzó una precisión del 96.8%, una sensibilidad (recall) del 94.5% y un puntaje F1 de 95.6%, demostrando un equilibrio adecuado entre exactitud y capacidad de detección. Estos resultados confirman la capacidad del sistema para distinguir tonalidades y texturas asociadas a diferentes grados de horneado con un margen mínimo de error.

Las curvas de pérdida durante el entrenamiento del modelo YOLOv11 presentadas en la Figura 7 muestran una tendencia descendente sostenida tanto en la pérdida de entrenamiento como en las curvas suavizadas, indicando convergencia adecuada del algoritmo de optimización sin signos evidentes de sobreajuste.

La reducción progresiva de las pérdidas de clasificación, localización y distribución (DFL) evidencia un aprendizaje estable de las características visuales asociadas al nivel de cocción.

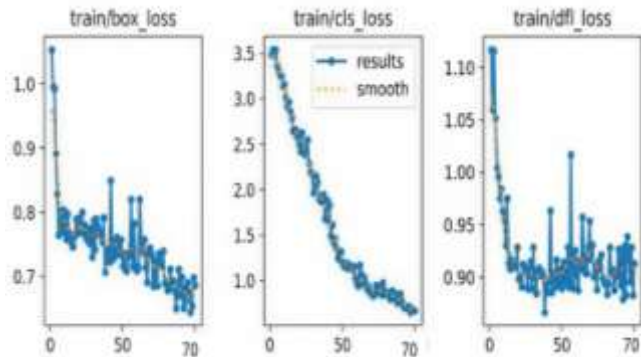


Figura 7. Curvas de pérdida durante el entrenamiento del modelo YOLOv11

Las métricas de precisión y exhaustividad durante el entrenamiento mostradas en la Figura 8 evidencian incrementos significativos durante las primeras 50 épocas, alcanzando posteriormente una meseta de estabilización cercana a valores de 1.0 para precisión y 0.95 para recall, indicando que el modelo logró un balance adecuado entre la minimización de falsos positivos y falsos negativos.

Si bien se presentan ligeras variaciones intermedias propias del proceso de optimización iterativa, la tendencia general muestra una convergencia estable hacia las últimas épocas. En conjunto, el comportamiento de ambas métricas confirma la estabilidad del entrenamiento y la capacidad del modelo para mantener un desempeño consistente en la diferenciación de las clases evaluadas.

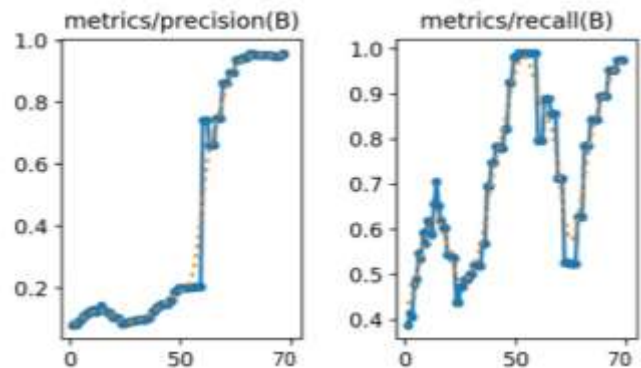


Figura 8. Métricas de precisión y exhaustividad durante el entrenamiento del modelo YOLOv11

La matriz de confusión normalizada presentada en la Figura 9 muestra un desempeño alto en la identificación de ambas clases: "alfajor normal" (0.69) y "alfajor quemado" (0.84), con valores de confusión cruzada bajos (0.02), lo cual indica que el modelo logra diferenciar correctamente las tonalidades asociadas al quemado; asimismo, se identifica una proporción relevante de detecciones clasificadas como "background", característica común en modelos YOLO debido a la presencia de regiones sin objetos en las imágenes. En conjunto, estos resultados confirman la capacidad del modelo para generalizar en condiciones reales.

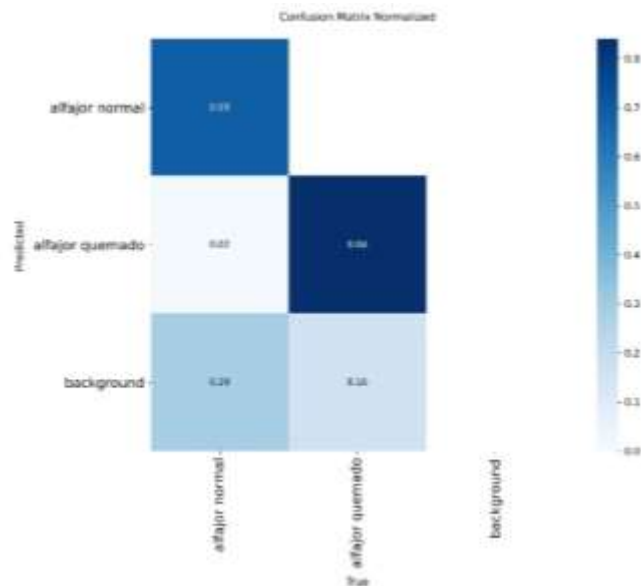


Figura 9. Matriz de confusión normalizada del modelo YOLOv11

La detección automática de tapas de alfajor presentada en la Figura 10 muestra el modelo YOLOv11 identificando cada tapa mediante cuadros delimitadores y etiquetas que indican la clase asignada y el nivel de confianza obtenido durante la inferencia, evidenciando alta precisión en la diferenciación de tonalidades.



Figura 10. Detección de tapas de alfajor según su estado de cocción.

Las métricas de precisión en la época 70 presentadas en la Figura 11 muestran que el valor de mAP50 alcanza aproximadamente 1.0, lo que indica una alta precisión en la detección de objetos cuando se utiliza un umbral de IoU de 0.50, mientras que el mAP50-95, que refleja una precisión más exigente al promediar múltiples umbrales de IoU entre 0.50 y 0.95, alcanza valores cercanos a 0.8, evidenciando un desempeño robusto y generalizado del modelo en distintos niveles de exigencia.

En conjunto, ambas métricas demuestran una evolución positiva del modelo a lo largo del entrenamiento, alcanzando niveles de precisión adecuados para aplicaciones prácticas.

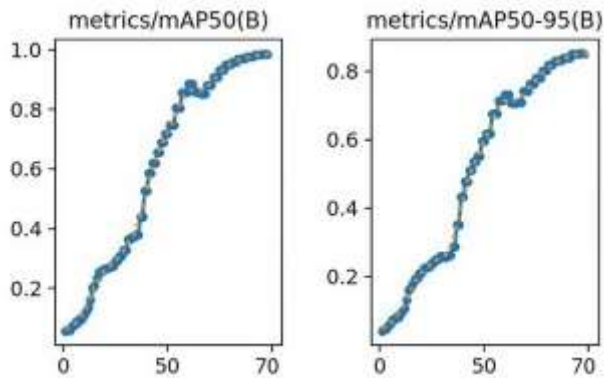


Figura 11. Métricas de precisión en la época 70.

4. DISCUSIÓN

Los resultados de este estudio superan ligeramente los reportados por [1], quienes alcanzaron un F1-score del 93.57% en la clasificación de madurez de mango utilizando una CNN con imágenes de baja resolución. En contraste, el modelo YOLOv11 empleado en la presente investigación alcanzó un F1-score de 95.6%, lo que evidencia una mejora en la

capacidad de identificar variaciones cromáticas específicas como las asociadas al quemado; esta diferencia se explica técnicamente porque YOLOv11 integra un mecanismo de extracción de características multi-escala que permite detectar patrones locales de textura sin perder contexto global, mientras que una CNN tradicional depende de características aprendidas de manera progresiva y puede perder detalles finos. Además, el uso de data augmentation orientado a variaciones de brillo y contraste fortaleció la robustez del modelo frente a cambios en la reflectancia superficial del alfajor, lo que favoreció la generalización en la detección de tapas quemadas.

Al comparar con el sistema desarrollado por [5] para la clasificación de paltas Hass, el cual alcanzó una precisión del 92.85%, el modelo YOLOv11 presenta un rendimiento superior con una precisión del 96.8%; esta mejora técnica se atribuye a que YOLOv11 emplea un pipeline unificado de detección y clasificación, lo que reduce el error acumulado entre módulos, a diferencia de los sistemas tradicionales que primero segmentan y luego clasifican. Asimismo, la arquitectura FPN integrada en YOLOv11 permite detectar variaciones sutiles de tonalidad que indican quemado superficial, algo crítico en productos como tapas de alfajor, donde pequeñas áreas oscuras pueden comprometer la calidad. La mayor resolución del dataset utilizado también contribuyó a mejorar el reconocimiento de textura, permitiendo capturar de manera más nítida los patrones de cocción irregular.

Finalmente, al confrontar los resultados con el sistema de inspección en tiempo real de [6], quienes reportaron una precisión en tiempo real del 96.7% para clasificación de frutas, se observa que el presente trabajo ofrece un desempeño equivalente pero con una ventaja: una mayor capacidad de generalización reflejada en métricas mAP50 = 1.00 y mAP50-95 = 0.87. Esta diferencia técnica se explica porque YOLOv11 incorpora bloques convolucionales optimizados (C3, SPP-F y módulos de atención espacial) que permiten capturar características complejas en superficies irregulares como el quemado; a diferencia de modelos como EfficientNet o MobileNet utilizados por [6], que priorizan eficiencia computacional, YOLOv11 equilibra velocidad con mayor profundidad de red, lo que facilita la identificación de variaciones texturales finas causadas por el sobrehecho. Esto hace que el sistema propuesto sea más adecuado para aplicaciones donde las diferencias visuales entre clases son mínimas pero críticas para la calidad final del producto.

5. CONCLUSIONES

El modelo desarrollado demostró notable eficacia en la detección automática de tapas de alfajor quemadas mediante la aplicación de técnicas de visión artificial basadas en inteligencia artificial. La intervención de la IA permitió optimizar el proceso de clasificación, alcanzando un nivel de precisión del 96.8%, sensibilidad del 94.5%, F1-score de 95.6%, mAP50 de 1.00 y mAP50-95 de 0.87, lo que confirma el cumplimiento de los objetivos específicos planteados: se construyó un conjunto de datos balanceado compuesto por

1,200 imágenes distribuidas equitativamente entre las categorías definidas; se entrenó y optimizó el modelo superando las métricas mínimas establecidas (precisión $\geq 95\%$, sensibilidad $\geq 90\%$, F1-score $\geq 92\%$, mAP50 ≥ 0.90); y se validó el sistema bajo condiciones controladas de iluminación LED (5500K) con ángulo cenital. No se realizaron pruebas bajo condiciones ambientales industriales variables, lo que constituye una línea de validación futura para evaluar la robustez del sistema en entornos reales de producción a escala industrial.

Los resultados obtenidos sugieren que este tipo de sistema podría, con los ajustes técnicos y validaciones adicionales correspondientes, complementar o mejorar los procesos de inspección manual tradicional en entornos de producción industrial; no obstante, para afirmar con certeza que puede reemplazar completamente la inspección manual, se requeriría realizar pruebas exhaustivas en condiciones reales de planta que incluyan: validación con lotes de producción variados para evaluar generalización del modelo frente a diferencias naturales en materia prima; evaluación de desempeño bajo condiciones de iluminación ambiental variable; pruebas de velocidad de procesamiento en línea de producción operando a velocidades industriales; y estudios comparativos directos contra inspectores humanos midiendo tiempo de clasificación, tasa de error y costos operativos.

Entre las limitaciones identificadas en este estudio se encuentran: el entrenamiento y validación se realizaron bajo condiciones controladas de iluminación y ángulo de captura que podrían no representar completamente la variabilidad presente en entornos industriales reales; el conjunto de datos se construyó con productos de un único lote de producción, por lo que la generalización del modelo frente a variaciones naturales en ingredientes, temperatura de horneado y tiempo de cocción requiere validación adicional; no se evaluó el desempeño del sistema operando a velocidades industriales continuas, aspecto crítico para determinar viabilidad de implementación en líneas de producción; y el modelo fue entrenado únicamente para dos categorías discretas, cuando en la práctica industrial podría ser útil detectar múltiples niveles de quemado o defectos adicionales.

Como trabajo futuro se propone: expandir el conjunto de datos para incluir imágenes capturadas bajo diferentes condiciones de iluminación y múltiples ángulos de captura para mejorar robustez del modelo; realizar pruebas de validación cruzada con lotes de producción de diferentes fechas, proveedores de materia prima y parámetros de horneado para evaluar capacidad de generalización real; implementar el sistema en una línea de producción piloto para evaluar desempeño en tiempo real, midiendo velocidad de procesamiento, latencia de inferencia y estabilidad operativa durante jornadas completas de trabajo; desarrollar una versión multi-clase del modelo capaz de clasificar niveles graduales de quemado y detectar defectos adicionales relevantes para control de calidad; realizar estudios comparativos formales entre el sistema automatizado y la inspección manual, cuantificando diferencias en velocidad, precisión, consistencia

inter-operador y costos operativos; investigar técnicas de optimización del modelo para reducir requisitos computacionales y facilitar implementación en hardware embebido de bajo costo; asimismo, la arquitectura implementada podría adaptarse para la detección de otros defectos en productos horneados, tales como deformaciones estructurales, grietas superficiales, variaciones de tamaño o problemas de forma, siempre que se disponga de un conjunto de datos representativo y correctamente etiquetado para su entrenamiento supervisado del modelo; y finalmente, explorar la aplicabilidad del enfoque propuesto a otros productos de repostería industrial para validar su potencial de generalización a categorías similares de alimentos horneados.

En síntesis, este estudio aporta: un conjunto de datos especializado de tapas de alfajor en diferentes estados de cocción; un protocolo metodológico documentado para la construcción, entrenamiento y validación de sistemas de visión artificial orientados a control de calidad en productos horneados; evidencia empírica sobre la robustez de YOLOv11 para detectar variaciones cromáticas y texturales sutiles asociadas al quemado superficial; y una guía técnica que puede orientar la implementación de sistemas similares en la industria de repostería y panificación, contribuyendo a la modernización tecnológica del sector alimentario mediante la incorporación de técnicas de inteligencia artificial en procesos de control de calidad.

REFERENCIAS

- [1] O. Salazar-Campos, J. Moran Ruiz, J. L. Peralta, M. Rubio Cieza, B. Salazar Medina, and J. Salazar-Campos, "Deep learning approach for automated Kent mango maturity grading in compliance with Peruvian standards," Pontificia Universidad Católica del Perú, 2023. [Online]. Available: <https://cris.pucp.edu.pe/en/publications/deep-learning-approach-for-automated-kent-mango-maturity-grading>
- [2] J. R. O. Zacarias, Y. C. Sullca-Mendoza, Y. S. Valenzuela-Lino, S. I. Del Carpio-Ramírez, N. Moggiano, and C. Coaquira-Rojo, "Simulación de un sistema automatizado de clasificación de limones tahitianos basado en visión artificial," in 13th Annual IEEE Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), New York, USA, 2022, pp. 283-289.
- [3] J. R. Ortiz-Zacarias, M. C. Rivera-Rojas, M. V. Yanarico-Estrella, I. L. Quintanilla-Mosquera, Y. S. Valenzuela-Lino, and C. Coaquira-Rojo, "Simulación de un sistema automatizado de clasificación de mangos peruanos basado en visión artificial," in Proc. 5th Int. Conf. Information and Communications Technology (ICOIACT), Yogyakarta, Indonesia, 2022, pp. 385-390, doi: 10.1109/ICOIACT55506.2022.9972217.
- [4] R. León, M. Díaz, and L. Rodríguez, "Gestión de un sistema de visión artificial para la detección de los daños causados por plagas en el cultivo de palto utilizando un drone," Revista Ciencia y Tecnología, vol. 16, no. 4, pp. 145-151, 2020. [Online]. Available: <https://revistas.unitr.edu.pe/index.php/PGM/article/view/3151>
- [5] R. L. Alarcón Marín, "Diseño de un sistema automático de clasificación de paltas por visión artificial," B.S. thesis, Pontificia Universidad Católica del Perú, Lima, Peru, 2023. [Online]. Available: <http://hdl.handle.net/20.500.12404/25914>
- [6] N. Ismail and O. A. Malik, "Real-time visual inspection system for grading fruits using computer vision and deep learning techniques," Information Processing in Agriculture, vol. 9, no. 1, pp. 24-37, 2022, doi: 10.1016/j.inpa.2021.01.005.
- [7] S. K. Chakraborty et al., "Development of an optimally designed real-time automatic citrus fruit grading-sorting machine leveraging computer

- vision-based adaptive deep learning model," *Engineering Applications of Artificial Intelligence*, vol. 120, art. no. 105826, 2023, doi: 10.1016/j.engappai.2023.105826.
- [8] G. Bortolotti, M. Piani, M. Gullino, D. Mengoli, C. Franceschini, L. Corelli Grappadelli, and L. Manfrini, "A computer vision system for apple fruit sizing by means of low-cost depth camera and neural network application," *Precision Agriculture*, vol. 25, no. 6, pp. 2740-2757, 2024, doi: 10.1007/s11119-024-10139-8.
 - [9] J. Liu et al., "New Progress in Intelligent Picking: Online Detection of Apple Maturity and Fruit Diameter Based on Machine Vision," *Agronomy*, vol. 14, no. 4, art. no. 721, 2024, doi: 10.3390/agronomy14040721.
 - [10] O. Martínez-Mora, O. Capuñay-Uceda, L. Caucha-Morales, R. Sánchez-Ancajima, I. Ramírez-Morales, S. Córdova-Márquez, and F. Cuenca-Mayorga, "Artificial Vision-Based Dual CNN Classification of Banana Ripeness and Quality Attributes Using RGB Images," *Processes*, vol. 13, no. 7, art. no. 1982, 2025, doi: 10.3390/pr13071982.
 - [11] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," *arXiv preprint arXiv:1804.02767*, 2018. [Online]. Available: <https://arxiv.org/abs/1804.02767>
 - [12] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 4th ed. Upper Saddle River, NJ, USA: Pearson, 2018.
 - [13] G. Jocher, A. Chaurasia, and J. Qiu, "Ultralytics YOLOv8," 2023. [Online]. Available: <https://github.com/ultralytics/ultralytics>
 - [14] M. Torres and C. López, "Implementación de visión artificial en control de calidad industrial," *Ingeniería e Investigación*, vol. 40, no. 3, 2020, doi: 10.15446/ing.investig.v40n3.83246.
 - [15] O. Salazar-Campos et al., "Deep learning para clasificación de madurez de mango Kent según la norma peruana," *Pontificia Universidad Católica del Perú*, 2023. [Online]. Available: <https://cris.pucp.edu.pe/en/publications/deep-learning-approach-for-automated-kent-mango-maturity-grading>
 - [16] R. León, M. Díaz, and L. Rodríguez, "Sistema de visión artificial para detección de plagas en palto usando dron," *Revista Ciencia y Tecnología*, vol. 16, no. 4, pp. 145-151, 2020. [Online]. Available: <https://revistas.unitru.edu.pe/index.php/PGM/article/view/3151>
 - [17] J. R. O. Zacarias et al., "Simulación de un sistema automatizado de clasificación de limones Tahití basado en visión artificial," in *IEEE UEMCON 2022*, New York, USA, 2022. [Online]. Available: <https://cris.continental.edu.pe/en/publications/simulation-of-an-automated-tahitian-lemon-grading-system-based-on>
 - [18] R. Delgado and P. Vargas, "Aplicación de visión artificial en la industria alimentaria latinoamericana," *Revista Ingeniería y Competitividad*, vol. 23, no. 1, 2021, doi: 10.25100/iyv.v23i1.10869.
 - [19] R. L. Alarcón Marín, "Diseño de un sistema automático de clasificación de paltas por visión artificial," B.S. thesis, Pontificia Universidad Católica del Perú, Lima, Peru, 2023. [Online]. Available: <http://hdl.handle.net/20.500.12404/25914>
 - [20] M. Betancourt and L. Chacón, "Evaluación de calidad en frutas mediante visión artificial," *Revista Ingeniería UC*, 2019. [Online]. Available: <https://revistas.ing.uc.cl/index.php/RIN/article/view/166>
 - [21] J. Hernández and C. Córdova, "Algoritmos de visión artificial para detección de defectos superficiales," *Revista Dyna*, vol. 89, no. 222, 2022, doi: 10.15446/dyna.v89n222.96383.
 - [22] A. Chávez and V. Ramírez, "Uso de redes neuronales convolucionales para inspección visual de alimentos," *Revista Politécnica*, art. no. 1160, 2020. [Online]. Available: https://revistapolitecnica.epn.edu.ec/ojs2/index.php/revista_politecnica/article/view/1160
 - [23] G. Ruiz and D. Salinas, "Automatización del control de calidad mediante visión artificial," *Revista Tecnura*, 2021, doi: 10.14483/22487638.17083.
 - [24] R. Medina and J. Carranza, "Clasificación de productos panificados mediante análisis de imagen," *Revista Peruana de Computación y Sistemas*, 2021. [Online]. Available: <https://revistapcys.uni.edu.pe>
 - [25] L. Bravo and V. Herrera, "Implementación de sistemas de visión artificial en panificación industrial," *Revista Ingeniería Industrial – UNMSM*, art. no. 19968, 2020. [Online]. Available: <https://revistasinvestigacion.unmsm.edu.pe/index.php/idata/article/view/19968>