Clasificación automatizada de camisetas mediante redes neuronales convolucionales

Roger A. QUEREVALU-GALAN
Escuela Ingeniería de Sistemas, - Universidad Nacional de Trujillo
Trujillo, La Libertad, C.P. 13001, Perú

Giovani. SALCEDO-QUISPE
Escuela Ingeniería de Sistemas, - Universidad Nacional de Trujillo
Trujillo, La Libertad, C.P. 13001, Perú

Juan P. SANTOS-FERNÁNDEZ
Departamento de Ingeniería de Sistemas, - Universidad Nacional de Trujillo
Trujillo, La Libertad, C.P. 13001, Perú

RESUMEN

Este estudio desarrolló un sistema automatizado de clasificación de camisetas mediante redes neuronales convolucionales (CNN), aplicado al dataset Fashion Product Images de Kaggle, con el objetivo de mejorar la categorización multi-atributo de productos textiles en el comercio electrónico. Se utilizaron 8,734 imágenes de camisetas con una arquitectura CNN personalizada basada en ResNet-50 modificado. Las imágenes en bruto se filtraron del conjunto original de 44,441 productos según el tipo de articulo: "camisetas", junto con atributos de género y uso. Para la clasificación por el tipo género y uso, nuestra metodología arrojó 95% y 89% de precisión, respectivamente, con un puntaje F1 promedio de 0.918. Se cree que el marco sugerido supera el estándar y posibilita un enfoque escalable y granular para plataformas de comercio electrónico, con potencial para impulsar la búsqueda visual y las recomendaciones de productos y la administración del inventario
Palabras Claves: Clasificación de imágenes, aprendizaje profundo, análisis de moda, comercio electrónico, redes neuronales.

1. INTRODUCCIÓN

El comercio electrónico en el ámbito de la moda ha revolucionado la industria textil, generando la necesidad de una gestión automatizada de productos en inventarios y catálogos extensos. Estudios como Chugh y Jain (2024) avanzan, pues esta técnica está alterando las prácticas del sistema logístico y las recomendaciones de productos de las corporaciones, como Amazon, H&M y Shein. Igualmente, Kempele (2023) menciona que las camisetas, al ser una de las prendas más populares y adaptables, constituyen un sector importante dentro del vestuario casual a nivel mundial. Se proyecta que el mercado global de camisetas llegará a los 46 990 millones de dólares en 2025 y podría incrementarse hasta los 52 800 millones de dólares para 2029, de acuerdo con proyecciones del sector. Además, en 2023 se estimaron las importaciones mundiales de camisetas en cerca de 44 930 millones de USD, lo que resalta su importancia en el comercio global de ropa. La adecuada clasificación de estos productos influye directamente en la experiencia del usuario, la eficacia en las operaciones y el lucro de las plataformas de comercio electrónico.

Por otro lado, Archana y Jeevaraj (2024) afirman que los avances en inteligencia artificial aplicada a la clasificación de productos de moda han sido significativos en la última década. Las técnicas tradicionales basadas en descriptores manuales (como SIFT, HOG y LBP) han quedado ampliamente rezagadas frente a los métodos de aprendizaje profundo, especialmente las redes neuronales convolucionales. Abbas et al. (2024) demostraron que las redes neuronales convolucionales, empleando arquitecturas como ResNet-50 y EfficientNet-B0, pueden alcanzar precisiones superiores al 90 % en la clasificación de prendas de vestir en conjuntos de datos reales de e-commerce, logrando hasta un 92.2 % de exactitud.

Sin embargo, la mayoría de las investigaciones se han concentrado en la clasificación de categorías individuales (tipo de prenda, color o género), lo que limita su aplicabilidad en sistemas reales que requieren la predicción simultánea de múltiples atributos. Por ejemplo, Guo et al. (2019) destacan que los conjuntos de datos convencionales están diseñados para etiquetas únicas y objetos a nivel general, en comparación con datasets de moda como iMaterialist, que contienen múltiples atributos finos por imagen, lo que evidencia la necesidad de métodos multitarea

Zhang et al. (2020) mencionan que las arquitecturas de redes neuronales convolucionales más destacadas en la clasificación de moda incluyen variantes de ResNet mejoradas con mecanismos de atención, los cuales han demostrado un rendimiento sobresaliente; por ejemplo, ResNeSt que introduce bloques Split-Attention en ResNet50 logró un top 1 accuracy de 81.13 % en ImageNet, mientras que variantes similares con atención han mejorado notablemente tareas de clasificación de prendas. Además, Vision Transformers (ViT) han alcanzado precisiones superiores al 95 % en Fashion-MNIST y han superado a modelos CNN tradicionales como ResNet-50 (84.5 %) Abd Alaziz et al. (2023)

La clasificación automática de camisetas en contextos reales presenta desafíos técnicos relevantes aún no resueltos de manera

efectiva Liang et al. (2023) mencionan como la alta variabilidad en presentación fotográfica, condiciones de iluminación, poses de modelos y fondos, creando inconsistencias que afectan el rendimiento de los modelos. Además, Reddi et al. (2019) plantean que, en entornos de producción, los sistemas necesitan tiempos de inferencia inferiores a 50 ms por imagen para permitir un procesamiento en tiempo real y cumplir con los requisitos de experiencia del usuario, lo que restringe la complejidad de las arquitecturas que pueden emplearse, como lo indican los estándares establecidos por MLPerf para aplicaciones interactivas

Las redes neuronales convolucionales representan la solución más alentadora para abordar estos desafíos por las siguientes razones:

Según LeCun et al. (2015) las CNN pueden aprender automáticamente características visuales relevantes para cada atributo, eliminando la necesidad de ingeniería manual de características y adaptándose a la variabilidad visual de productos reales

Igualmente, Ruder (2017) menciona que permiten implementar arquitecturas multitarea que pueden predecir simultáneamente múltiples atributos, optimizando la eficiencia computacional y aprovechando las correlaciones inter-atributos

Se planteo como objetivo desarrollar un sistema de clasificación multiatributo de camisetas basado en redes neuronales convolucionales utilizando el Fashion Product Images Dataset, que logre alta precisión en la predicción simultánea de género y uso, con viabilidad para implementación en plataformas de comercio electrónico. Para lograr este objetivo, se plantean los siguientes objetivos específicos

Diseñar e implementar una arquitectura CNN multitarea optimizada para clasificación de camisetas que supere el 80% de precisión en cada atributo individual.

Evaluar comparativamente el rendimiento del modelo propuesto contra arquitecturas de referencia (ResNet-50, EfficientNet, Vision Transformer) utilizando métricas de precisión, recall, F1-score y tiempo de inferencia.

2. MATERIALES Y MÉTODOS

Este estudio es un experimento práctico que se basa en números y datos. Para ello, diseñamos un experimento en el que modificamos la variable principal — la red neuronal convolucional — para ver cómo afecta a otras variables, como la clasificación por género y el tipo de camiseta. La investigación se enmarca en el campo del aprendizaje automático supervisado, usando técnicas avanzadas de aprendizaje profundo para clasificar imágenes. Para la muestra, usamos todas las imágenes del conjunto de datos llamado Fashion Product Images Dataset, que está disponible en Kaggle y contiene 44,441 fotos de productos de moda con sus detalles. De esas, seleccionamos intencionadamente solo las que son camisetas, siguiendo ciertos criterios específicos para asegurarnos de que solo esas imágenes entraron en el estudio. Tabla 1: Criterios para elegir las imágenes de camisetas.

Tabla 1: Criterios de selección de imágenes para camisetas

Criterios de inclusión	Criterios de exclusión
Etiquetadas como “camisetas” en tipo de articulo	Etiquetas ambiguas o incompletas
Metadatos completos para género y uso	Imágenes corruptas o con errores de formato
Resolución ≥ 60 × 80 píxeles	Productos híbridos (camisetas con chaquetas)
Formato JPG, JPEG o PNG	

La muestra final resultó en 8,734 imágenes de camisetas, distribuidas de la siguiente manera: 5,421 imágenes de género masculino (62.1%), 3,313 imágenes de género femenino (37.9%); y por uso: 6,393 imágenes de uso casual (73.2%), 2,341 imágenes de uso deportivo (26.8%).

Se utilizó el Fashion Product Images Dataset, un conjunto de datos públicos desarrollado por Param Aggarwal y disponible en Kaggle bajo licencia Creative Commons. Este dataset ha sido ampliamente utilizado en la literatura científica y contiene imágenes de productos de una plataforma de comercio electrónico real. El conjunto de datos incluye los recursos descritos en la Tabla 2.

Tabla 2: Recursos de datos de camisetas

Recurso	Descripción
styles.csv	Archivo de metadatos con 44,441 registros conteniendo id, gender, masterCategory, subCategory, articleType, baseColour, season, year, usage, productDisplayName
images/	Directorio con 44,441 imágenes en formato JPG con resoluciones variables (60x80 a 2400x3200 píxeles)

La selección de este conjunto de datos se justifica por su representatividad del dominio real de comercio electrónico, la calidad de las anotaciones, y su uso previo en investigaciones de referencia que permiten comparabilidad de resultados.

El desarrollo del sistema se realizó utilizando Python 3.11.3 como lenguaje de programación principal, ejecutado en el editor de código Visual Studio Code 1.101 con las siguientes especificaciones: GPU Intel Iris Xe Graphics y RAM de 12 GB.

Se creó una red CNN multitarea que usa ResNet-50, ya preentrenado en ImageNet. Se empleó una división estratificada 70-15-15 para entrenamiento, validación y prueba, respectivamente, manteniendo la distribución proporcional de

clases, con 6,114 imágenes (70%) para entrenamiento, 1,310 imágenes (15%) para validación y 1,310 imágenes (15%) para prueba. En la configuración de entrenamiento, se utilizó el optimizador Adam con una tasa de aprendizaje inicial de 0.001, un scheduler ReduceLROnPlateau con un factor de 0.5 y paciencia de 5, un tamaño de lote de 64, un máximo de 30 épocas, early stopping con paciencia de 5 épocas basado en la pérdida de validación, y un callback ModelCheckpoint para guardar el mejor modelo.

Para evaluar cómo funciona cada tarea de clasificación, usamos estas métricas principales: accuracy (o precisión), que nos dice qué tan correctas son las predicciones en general; precision, que nos muestra qué tan bien evita falsos positivos; recall (o sensibilidad), que indica qué tan bien detecta los verdaderos positivos; y F1-score, que es un promedio entre precision y recall para tener una idea más balanceada.

Se usaron pruebas estadísticas, en particular el prueba de McNemar, para comparar cómo les fue a diferentes modelos (ResNet-50, EfficientNet-B0 y Vision Transformer) en el mismo conjunto de datos. La idea era ver si las diferencias en la precisión son realmente importantes o si podrían haber sido por azar, considerando un nivel de confianza del 95% ($p < 0.05$).

3. DISCUSIÓN Y RESULTADOS

4. CONCLUSIONES

5. REFERENCIAS

Chugh, P., & Jain, V. (2024). *Artificial Intelligence Empowerment in E-Commerce: A Bibliometric Voyage*. *Journal of Contemporary Retail and E-Business*, advance online publication. https://doi.org/10.1177/09711023241303621

Ķempele, S. (2023). *20 T-Shirt industry statistics and trends*. Printful Blog. https://www.printful.com/blog/t-shirt-industry-statistics

Archana, R., & Jeevaraj, P. E. (2024). Deep learning models for digital image processing: a review. *Artificial Intelligence Review*, *57*(1), 11. https://doi.org/10.1007/s10462-023-10631-z

Abbas, W., Zhang, Z., Asim, M., Chen, J., & Ahmad, S. (2024). Ai-driven precision clothing classification: Revolutionizing online fashion retailing with hybrid two-objective learning. *Information*, *15*(4), 196. https://doi.org/10.3390/info15040196

Guo, S., Huang, W., Zhang, X., Srikhanta, P., Cui, Y., Li, Y., ... & Belongie, S. (2019). The imaterialist fashion attribute dataset. In *Proceedings of the IEEE/CVF international conference on computer vision workshops* (pp. 0-0). https://doi.org/10.48550/arXiv.1906.05750

Zhang, H., Wu, C., … Li, M. & Smola, A. (2020). *ResNeSt: Split-Attention Networks*. arXiv. https://doi.org/10.48550/arXiv.2004.08955

Abd Alaziz, H. M., Elmannai, H., Saleh, H., Hadjouni, M., Anter, A. M., Koura, A., & Kayed, M. (2023). Enhancing fashion classification with vision transformer (ViT) and developing recommendation fashion systems using DINOVA2. *Electronics*, *12*(20), 4263. https://doi.org/10.3390/electronics12204263

Liang, J., Liu, Y., & Vlassov, V. (2023). *The impact of background removal on performance of neural networks for fashion image classification and segmentation*. *arXiv*. https://doi.org/10.48550/arXiv.2308.09764

Reddi, V. J., Cheng, C., Kanter, D., …, & Zhong, A. (2019). *MLPerf Inference Benchmark*. arXiv. https://doi.org/10.48550/arXiv.1911.02549

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature,* *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Ruder, S. (2017). *An overview of multi-task learning in deep neural networks*. *arXiv preprint*, arXiv:1706.05098. https://doi.org/10.48550/arXiv.1706.05098