

Detection of Tongue Pathologies Using Computer Vision Based on MobileNetV2

Ríos Colque, Pablo Martín¹ and Helguero Velásquez, Hernán Luis²[0000-0002-3952-697X]

^{1,2} Facultad Nacional de Ingeniería, Universidad Técnica de Oruro, Oruro, Bolivia

¹pablo.rios.c@mail.pucv.cl

²herman.helguero@sistemas.edu.bo

Abstract. This paper presents the development of an intelligent system based on computer vision and deep learning for the automatic detection of tongue pathologies. The proposed approach employs the MobileNetV2 architecture to classify tongue images into six categories: healthy, fissured, white, geographic, bright red, and black-spotted tongue. The model was trained using data augmentation techniques and integrated into a web-based platform developed with TensorFlowJS, enabling real-time image analysis. The obtained results demonstrate satisfactory model performance under the experimental conditions considered, highlighting the potential of the proposed system as a support tool for preliminary diagnosis.

Keywords: Computer vision, deep learning, MobileNetV2, image classification, computer-aided diagnosis

Detección de Patologías Linguales mediante Visión por Computador basada en MobileNetV2

Resumen. En este trabajo se presenta el desarrollo de un sistema inteligente basado en Visión por Computador y aprendizaje profundo para la detección automática de patologías linguales. El enfoque propuesto utiliza la arquitectura MobileNetV2 para clasificar imágenes correspondientes a seis categorías: lengua sana, agrietada, blanca, geográfica, roja brillante y con manchas negras. El modelo fue entrenado utilizando técnicas de aumento de datos y se integró en una plataforma web desarrollada con TensorFlowJS, permitiendo el análisis en tiempo real. Los resultados obtenidos muestran un desempeño satisfactorio del modelo bajo las condiciones experimentales consideradas, evidenciando el potencial del sistema como herramienta de apoyo en el diagnóstico preliminar.

Palabras clave: Visión por Computador, aprendizaje profundo, MobileNetV2, clasificación de imágenes, diagnóstico asistido.

1 Introducción

La lengua es un órgano fundamental cuya apariencia puede ofrecer información valiosa sobre la salud de una persona. Diversas patologías, como infecciones, deficiencias nutricionales y enfermedades sistémicas, pueden manifestarse a través de cambios en el color, la textura o la forma de la lengua (Neville et al., 2015). Sin embargo, el diagnóstico visual por parte de especialistas puede ser subjetivo y depender de la experiencia clínica individual. Se consideraron seis categorías de patologías linguales debido a que presentan características visuales diferenciables, permitiendo evaluar la aplicación de técnicas de visión por computador en tareas de clasificación automática. Diversos estudios han demostrado que las técnicas de aprendizaje profundo aplicadas al análisis de imágenes médicas ofrecen resultados prometedores en tareas de diagnóstico asistido (Goodfellow et al., 2016; Ravì et al., 2017). En este contexto, las CNN han demostrado una alta capacidad para identificar patrones visuales en imágenes clínicas.

Entre las arquitecturas utilizadas para este tipo de aplicaciones se encuentra MobileNetV2, debido a su equilibrio entre precisión y costo computacional, lo que facilita su implementación en dispositivos con recursos limitados. (Sandler et al., 2018). En este trabajo presenta se desarrolló un sistema inteligente para identificar automáticamente seis categorías de patologías linguales mediante visión por computador y técnicas de aumento de datos. Como principal contribución, se propone la integración de un modelo de clasificación basado en MobileNetV2 con una plataforma web para realizar la detección automática de patologías linguales en tiempo real como apoyo al diagnóstico preliminar.

2 Metodología

2.1 Recolección de datos

En esta etapa se recopilieron imágenes correspondientes a seis categorías de patologías linguales, utilizadas para el entrenamiento y validación del modelo de clasificación. Las imágenes fueron organizadas de acuerdo a sus características clínicas: sana, agrietada, blanca, geográfica, roja brillante, manchas negras. El conjunto de datos estuvo conformado por un total de 744 imágenes. En la tabla 1 se muestra cada una de las patologías analizadas en el proyecto.

Tabla 1. Patologías de Lengua utilizadas en el modelo

Patología de Lengua	Imagen	Posibles causas	Nº Imágenes
Lengua Agrietada		Síndrome de Sjögren (enfermedad autoinmune que causa sequedad en la boca y ojos).	107
Lengua Blanca		Candidiasis (infección por hongos). Problemas digestivos y hepáticos. Deshidratación. Mala higiene bucal.	112

Lengua Geográfica		Alergias. Psoriasis. Problemas en el sistema inmunológico.	132
Lengua Roja Brillante		Deficiencias de vitaminas, especialmente de B12 y ácido fólico. Fiebre escarlatina en casos de infección bacteriana.	102
Lengua Sana		Buena higiene bucal.	116
Lengua con Manchas Negras o Marrones		Lengua vellosa ocurre por acumulación de células muertas, sobre por higiene bucal, fumadores o que consumen ciertos medicamentos.	175

Fuente: Adaptado de Neville et al. (2015).

2.2 Preparación de Datos

Los datos recolectados se procesaron y normalizaron para garantizar su calidad antes del entrenamiento del modelo. Posteriormente, se aplicaron técnicas de Data Augmentation, además de rotación, desplazamiento, recorte, zoom y ajustes de brillo y contraste, sobre las imágenes del conjunto de datos de las categorías de patologías linguales. Estas transformaciones permitieron incrementar la diversidad del conjunto de entrenamiento mediante nuevas muestras derivadas de las imágenes originales, favoreciendo la variabilidad de los datos utilizados durante el proceso de aprendizaje.



Fig. 1. Data Augmentation aplicado en imágenes de lenguas agrietadas.

2.3 Selección del Modelo

En esta etapa se seleccionó la arquitectura de CNN MobileNetV2 debido a su eficiencia computacional y adecuado equilibrio entre precisión y costo computacional, características que la hacen apropiada para dispositivos móviles y sistemas con recursos limitados (Sandler et al., 2018). Para la clasificación de las seis categorías de patologías linguales, se utilizó la versión preentrenada sobre ImageNet, reemplazando la capa de salida por una nueva capa adaptada a las categorías definidas en este trabajo.

2.4 Entrenamiento

El modelo fue entrenado durante 100 épocas utilizando el conjunto de datos aumentado. Para evaluar el proceso de aprendizaje se monitorearon las métricas de precisión (accuracy) y pérdida (loss) tanto en el conjunto de entrenamiento como en el de validación. El conjunto de datos fue dividido en dos subconjuntos: El 70% de las imágenes se destinó para el entrenamiento del modelo y el 30% restante a la validación, permitiendo evaluar su desempeño sobre datos no utilizados durante el aprendizaje.



Fig. 2. Procesamiento de imágenes para los tipos de patologías.

La figura 2 muestra ejemplos del procesamiento aplicado a las imágenes correspondientes a las diferentes categorías de patologías linguales utilizadas durante el entrenamiento.

2.5 Evaluación del Modelo

Una vez finalizado el entrenamiento, el modelo fue evaluado utilizando el conjunto de datos de validación, analizando su desempeño mediante las métricas de precisión (accuracy) y pérdida (loss). La figura 3 muestra la evolución de ambas métricas durante las 100 épocas de entrenamiento. Se observa un comportamiento estable de las curvas de entrenamiento y validación, lo que evidencia un proceso de aprendizaje consistente bajo las condiciones experimentales consideradas.

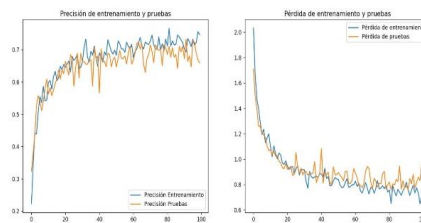


Fig. 3. Precisión y pérdidas durante el entrenamiento.

3 IMPLEMENTACIÓN

Se desarrolló una plataforma web para integrar el modelo de clasificación y permitir el análisis automático de imágenes linguales en tiempo real. La interfaz implementada ofrece las siguientes funcionalidades:

- Análisis en tiempo real: utiliza la cámara del dispositivo para capturar imágenes de la lengua y realizar la clasificación automática.
- Carga de imágenes: permite al usuario subir fotos de la lengua para su análisis.
- Resultados: muestra la categoría predicha entre las seis patologías consideradas, junto con la probabilidad asociada a cada clase.

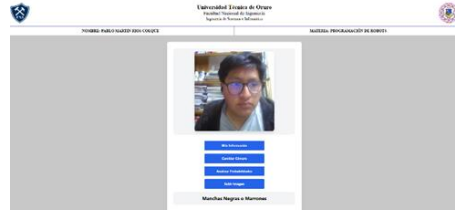


Fig. 4. Página web implementada para la detección de las patologías linguales.

4 RESULTADOS

Los resultados obtenidos evidencian un desempeño satisfactorio del modelo en la tarea de clasificación de patologías linguales. Durante el entrenamiento, se observó una convergencia progresiva de las métricas de precisión y pérdida, que indica un proceso de aprendizaje estable. Bajo las condiciones experimentales consideradas, el modelo alcanzó una precisión superior al 90% durante las etapas de entrenamiento y validación. Las pruebas realizadas con imágenes externas y mediante una cámara en tiempo real, mostraron el potencial del sistema para identificar las diferentes categorías de patologías linguales. La figura 5 presenta algunos ejemplos de clasificación obtenidos mediante la plataforma web, evidenciando la aplicabilidad del enfoque propuesto como herramienta de apoyo al diagnóstico preliminar.



Fig. 5. Detección de patologías con imágenes externas y en tiempo real.

5 Conclusiones y Trabajo Futuro

Los resultados obtenidos permiten concluir que el modelo basado en la arquitectura MobileNetV2 constituye una alternativa adecuada para la clasificación de patologías linguales, alcanzando un desempeño satisfactorio bajo las condiciones experimentales consideradas.

La aplicación de técnicas de Data Augmentation permitió incrementar la diversidad del conjunto de datos de entrenamiento mediante la generación de nuevas muestras a partir de las imágenes originales, contribuyendo al proceso de aprendizaje del modelo.

La división del conjunto de datos en 70% para entrenamiento y 30% para validación permitió evaluar el desempeño del modelo utilizando datos no empleados durante el entrenamiento. Asimismo, el monitoreo de las métricas de precisión (accuracy) y pérdida (loss) facilitó el seguimiento del proceso de aprendizaje durante las 100 épocas de entrenamiento.

La implementación del sistema mediante una plataforma web permitió realizar la clasificación automática de imágenes linguales tanto a partir de imágenes externas como mediante adquisición en tiempo real, evidenciando el potencial del enfoque propuesto como herramienta de apoyo al diagnóstico preliminar.

Como trabajo futuro, se plantea ampliar el conjunto de datos con nuevas imágenes y categorías de patologías linguales, así como incorporar métricas adicionales de evaluación que permitan analizar con mayor detalle el desempeño del modelo.

References

- Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.
- Neville, B. W., Damm, D. D., Allen, C. M., & Chi, A. C. (2015). Oral and Maxillofacial Pathology (4th ed.). Elsevier.
- Ravi, D., Wong, C., Deligianni, F., Berthelot, M., Andreu-Perez, J., Lo, B., & Yang, G. Z. (2017). Deep learning for health informatics. *IEEE Journal of Biomedical and Health Informatics*, 21(1), 4–21. <https://doi.org/10.1109/JBHI.2016.2636665>
- Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. <https://arxiv.org/abs/1801.04381>
- TensorFlow.js. (2024). A JavaScript library for training and deploying machine learning models in the browser and on Node.js. <https://www.tensorflow.org/js>