

# Modelo de aprendizaje profundo para el desarrollo de parámetros visuales de valoración del labio superior en pacientes operados de fisura labial unilateral

## Deep learning model for the development of visual parameters for upper lip assessing in patients operated on for unilateral cleft lip



Rosell-Perry P.

Percy ROSSELL-PERRY\*, Jhosimar ARIAS-FIGUEROA\*\*

### Resumen

**Introducción y objetivo.** La fisura labio palatina es un defecto congénito (la segunda malformación congénita más frecuente) de la cara, que afecta a la formación del labio superior, región alveolar y palatina. Su impacto social es grande; los pacientes se ven afectados psicológica y funcionalmente, pudiendo presentar defectos del habla y la alimentación así como secuelas estéticas que generan ansiedad, depresión, baja autoestima, etc.

El objetivo principal del presente estudio es desarrollar un modelo de aprendizaje profundo (*deep learning*) que permita valorar parámetros visuales labiales superiores postoperatorios en pacientes operados de fisura labial unilateral.

**Material y método.** Coleccionamos imágenes postoperatorias de 364 pacientes operados de fisura labial unilateral, calificadas por 3 revisores expertos como resultados buenos o malos. Entrenamos 3 modelos de redes convolucionales neuronales para determinar la simetría labial y la calidad del resultado postoperatorio y los comparamos en base a su sensibilidad, especificidad y curva ROC-AUC.

**Resultados.** Los 3 modelos de redes neuronales convolucionales usados fueron eficaces en la detección de la calidad de la cirugía realizada en pacientes con fisuras labiales unilaterales, siendo los modelos MobileNet y EfficientNet los que tuvieron mejores resultados con valores por encima del 80 % en todas las métricas.

**Conclusiones.** El modelo de inteligencia artificial utilizado ha demostrado su habilidad para medir la simetría labial postoperatoria en pacientes operados de fisura labial unilateral. Los mejores modelos fueron MobileNet y EfficientNet con métricas por encima del 80 %. Esta eficiencia se puede mejorar aumentando el número de datos.

### Abstract

**Background and objective.** Cleft lip and palate is a congenital defect (the second most frecuente congenital malformation) of the face that affects the formation of the upper lip, alveolar and palatine region. The social impact that this pathology represents is great; patients are affected psychologically and functionally, and may present speech and eating defects as well as aesthetic consequences that generate anxiety, depression, low self-esteem and others.

The main objective of our study is to develop a deep learning model that allows assessing postoperative upper labial visual parameters in patients operated on for unilateral cleft lip.

**Methods.** Postoperative images of 364 patients operated on for unilateral cleft lip were collected and rated by 3 expert reviewers as good or poor results. Three convolutional neural network models were trained to determine lip symmetry and postoperative outcome quality and were compared based on their sensitivity, specificity, and ROC-AUC curve.

**Results.** The 3 convolutional neural network models used were effective in detecting the quality of surgery performed in patients with unilateral cleft lips, with the MobileNet and EfficientNet models having the best results with values above 80% in all metrics.

**Conclusions.** The artificial intelligence models used have demonstrated their ability to measure postoperative lip symmetry in patients operated on for unilateral cleft lip. The best models were MobileNet and EfficientNet with metrics above 80%. This efficiency can be improved by increasing the number of data.

**Palabras clave** Inteligencia artificial, Aprendizaje profundo, Clasificación fisuras labio palatinas.

**Nivel de evidencia científica** 5c Diagnóstico  
**Recibido (esta versión)** 21 septiembre / 2024  
**Aceptado** 1 diciembre / 2024

**Key words** Artificial intelligence, Deep learning, Cleft lip and palate classification.

**Level of evidence** 5c Diagnostic  
**Received (this version)** September 21 / 2024  
**Accepted** December 1 / 2024

**Conflicto de intereses:** Los autores declaran no tener ningún interés financiero relacionado con el contenido de este artículo.  
**Financiación:** No hubo fuentes externas de financiación para este trabajo.

\* Cirujano Plástico, Profesor Investigador Facultad de Medicina Humana Universidad Peruana Union (UpeU), Lima, Perú. Maestría en Inteligencia Artificial en Salud, Universidad Católica de Murcia, España.

\*\* Ingeniero de sistemas, Universidad Peruana de Ciencias Aplicadas, Lima, Perú.

## Introducción

La fisura labio palatina es un defecto congénito (la segunda malformación congénita más frecuente) de la cara que afecta a la formación del labio superior, región alveolar y palatina. Afecta a uno de cada 700 nacidos vivos y se le atribuyen causas multifactoriales.<sup>(1)</sup> El impacto social que representa esta patología es grande; los pacientes se ven afectados psicológica y funcionalmente, pudiendo presentar defectos del habla y la alimentación así como secuelas estéticas labio-nasales que les generan ansiedad, depresión y baja autoestima, entre otros problemas. Su tratamiento demanda una intervención temprana y oportuna por parte de un equipo interdisciplinario de profesionales tales como cirujanos plásticos, otorrinolaringólogos, anesthesiólogos, pediatras, maxilofaciales, ortodoncistas, enfermeras, terapeutas de voz y habla y psicólogos.<sup>(2)</sup>

Existen diferentes modelos de atención para esta patología dependiendo de la disponibilidad de recursos y de los sistemas de salud desarrollados. Así en países en desarrollo, la modalidad de campañas quirúrgicas es muy utilizada para tratar a este tipo de pacientes, aunque esto tiene ventajas y desventajas ya descritas en la literatura.<sup>(3,4)</sup> Organizaciones como la Fundación Operación Sonrisa (*Smile Train*) de los Estados Unidos, la más importante a nivel mundial en el apoyo de la atención a niños nacidos con fisuras labio-palatinas en países en desarrollo, tienen como principal objetivo el desarrollo de centros especializados para el tratamiento de esta patología, para lo cual es esencial poder evaluar los resultados obtenidos por los diferentes equipos a la hora de llevar a cabo la planificación de sus actividades. El apoyo de esta organización no solo se da para la realización de las operaciones, sino también para el entrenamiento de los profesionales involucrados y la investigación científica, por lo cual, determinar la calidad de los resultados obtenidos por los diferentes equipos permite planificar las actividades de entrenamiento de los diferentes profesionales que conforman el equipo interdisciplinario. Sin embargo, existe falta de recursos para obtener de forma estandarizada esta valiosa información que repercutirá directamente en la mejora de la atención brindada a estos pacientes, y esto representa hoy por hoy una necesidad no cubierta.

El aprendizaje profundo (*deep learning*) es un tipo de aprendizaje automático que utiliza redes neuronales artificiales para aprender de los datos. Estas redes neuronales artificiales están inspiradas en el cerebro humano y se pueden utilizar para resolver una amplia variedad de problemas, incluidos el reconocimiento de imágenes, el procesamiento del lenguaje natural y el reconocimiento de voz. Consideramos que el desarrollo de modelos de aprendizaje profundo (*deep learning*) puede permitir ob-

tener parámetros visuales de simetría labio-nasal postoperatoria en pacientes operados de fisura labial unilateral. Esta información servirá para determinar la calidad de la cirugía practicada y así poder planificar actividades de entrenamiento de los equipos quirúrgicos con una mayor objetividad y eficiencia.

El objetivo principal del presente estudio es desarrollar una técnica de aprendizaje profundo que permita obtener parámetros visuales labionasales postoperatorios en pacientes operados de fisura labial unilateral que determinen la calidad de la cirugía realizada. De esta forma nos proponemos convertir data no estructurada (fotos postoperatorias) en estructurada para poder utilizarla en el entrenamiento de sistemas de aprendizaje profundo; desarrollar una metodología basada en inteligencia artificial para evaluar resultados postoperatorios de cirugía de fisura labial unilateral; y obtener información basada en esta tecnología que pueda ser utilizada en la mejora de la calidad de atención de los pacientes con fisuras labio-palatinas.

Los conocimientos susceptibles de transferencia se basan en los resultados del estudio en el que la aplicación desarrollada por la inteligencia artificial permitirá establecer la calidad de resultados quirúrgicos a obtener en base al grado de simetría del labio superior tras la operación. De esta forma, se fortalecerá la competitividad al darse a conocer la calidad de resultados operatorios proporcionada por cada hospital/centro quirúrgico, así como el desarrollo tecnológico al implementar aplicativos inteligentes que permitan evaluar la calidad de atención a los pacientes fisurados. Por otro lado, se mejorará la productividad al poder utilizar mejor recursos como la capacitación a los profesionales involucrados.

Estas actividades se desarrollan actualmente sin objetivos claros y de manera indiscriminada. Con el desarrollo de esta tecnología se podrá saber qué necesidades existen y en qué áreas del tratamiento quirúrgico de estos pacientes.

De manera indirecta la reducción de cirugías secundarias de corrección (producto de los malos resultados que pudieran obtenerse), tendría un impacto sobre la reducción en la necesidad de insumos quirúrgicos y medicamentosos, los cuales de una u otra forma afectan a nuestro medio ambiente.

Además, las actividades de capacitación profesional serán significativamente mejoradas a través de la identificación específica de las necesidades de estos centros.

## Material y método

Estudio analítico, observacional, coleccionando imágenes postoperatorias de 364 pacientes operados de fisura labial unilateral calificadas por 3 revisores expertos

como resultados buenos o malos; para tal fin se asignó la etiqueta de 1 a los casos malos y de 0 a los casos buenos.

Se incluyeron un total de 226 casos buenos y 138 casos malos. Todos los pacientes fueron operados entre los 3 y 6 meses de edad utilizando diferentes técnicas de queiloplastia primaria, tales como Millard y Tennison-Randall. Los resultados postoperatorios fueron analizados tras un periodo de al menos 1 año, lo cual garantiza la estabilidad de los resultados medidos.

Se estimaron como buenos resultados aquellos con mayor simetría del labio superior comparando el lado sano con el lado fisurado operado (Fig. 1). Asimismo, se consideró como mal resultado aquel con menor grado de simetría del labio superior comparando el lado fisurado operado con el lado sano (Fig. 2).

Para evaluar la concordancia intra e inter examinador se utilizó el coeficiente kappa. Se observó una concordancia de 0.90 y 0.92 respectivamente.

Entrenamos 3 modelos de redes convolucionales neuronales para determinar la simetría labial y la calidad del resultado postoperatorio. Las redes convolucionales neuronales son tipos de red neuronal artificial donde las neuronas artificiales corresponden a campos receptivos de una manera muy similar a las neuronas en la corteza visual primaria (V1) de un cerebro biológico.<sup>(5,6)</sup> Son muy efectivas para tareas de visión artificial, como en la clasificación y segmentación de imágenes, entre otras aplicaciones.<sup>(7)</sup>

Los 3 modelos de redes convolucionales neuronales entrenados por nosotros para determinar la simetría labial y la calidad de resultado postoperatorio fueron:

- ResNet-50. Es una red neuronal convolucional con 50 capas de profundidad utilizada en clasificación de imágenes. Es un analista de imágenes que permite diseccionar una imagen y dentro de ella identificar objetos para clasificarlos.
- MobileNet. Es otra red neuronal convolucional de código abierto de Google diseñada para usarse en aplicaciones móviles. Este es un algoritmo que permite detectar partes y características deseadas de una imagen.
- EfficientNetV2-B0. Esta red neuronal convolucional tiene modelos más pequeños y de entrenamiento más rápido y de mejor eficiencia

Se realizó *fine-tuning* (proceso de ajustar un modelo pre-entrenado para una tarea o dominio específico, mejorando su rendimiento en aplicaciones concretas) de estos modelos, congelando la mayoría de capas y entrenándose las últimas.

Se utilizó un modelo de procesamiento de imágenes en aprendizaje profundo (*deep learning*), la red neuronal convolucional con técnica de clasificación. Este sistema

basándose en los datos de entrenamiento, extrae automáticamente características que luego se usarán para la clasificación de la data (Fig. 3).



Figura 1. Ejemplo de caso de tipo buen resultado.

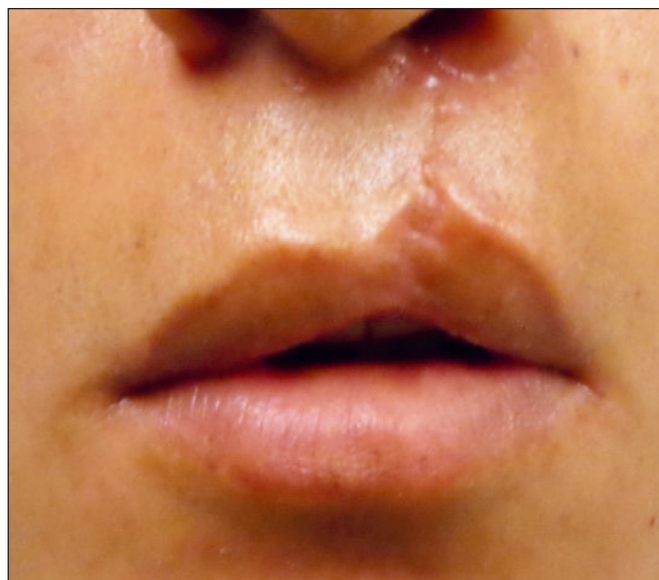


Figura 1. Ejemplo de caso de tipo mal resultado.

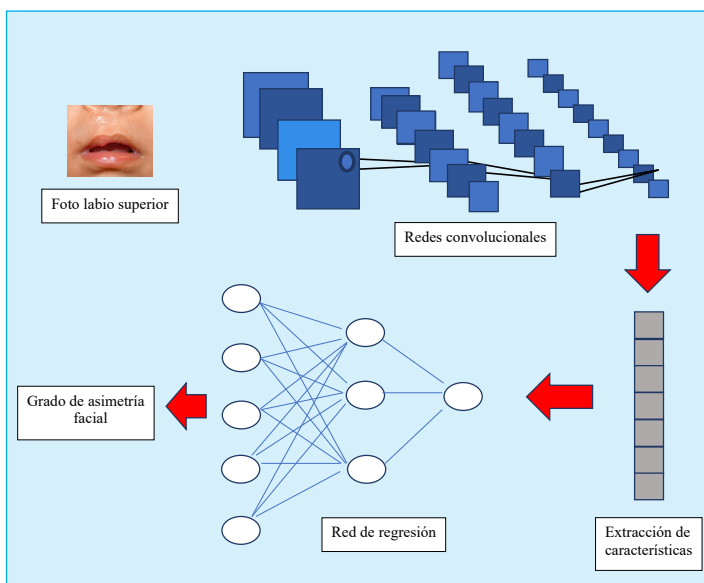


Figura 3. Sistema de redes convolucionales para el entrenamiento de modelos de clasificación de tipos de simetría labial en pacientes operados de fisura labial unilateral.



Esta técnica permitió establecer categorías en base al grado de asimetría labio-nasal, como buen y mal resultado.

La data disponible es una data no estructurada que consiste en imágenes faciales frontales estandar de al menos un año de evolución de los pacientes operados de fisura labial unilateral que compusieron el grupo de estudio. Estas fotos se registran en alta resolución para su análisis.

Las tomas fotográficas se hicieron bajo las siguientes características: paciente en bipedestación o en decúbito dorsal bajo anestesia general, posición de la cámara horizontal o vertical dependiendo de la posición del paciente, ubicada a la altura de los ojos del mismo, a una distancia de 0.5 m y con enfoque a nivel de las alas nasales.

Las características de la cámara empleado fueron: cámara digital compacta SONY® Cyber-shot DSC-RX100M7, con zoom de 24-200 mm y resolución de 20.1 megapíxeles.

### Metodología de entrenamiento de los modelos a evaluar

- Particionamiento de datos. Se dividieron en 70% (254 muestras) de entrenamiento, 15% (55 muestras) de validación y 15% (55 muestras) de prueba.
- Creación del set de datos. Utilizando las librerías de aprendizaje automático Tensorflow y Keras, se procedió a entrenar los modelos de redes neuronales seleccionados en el procesamiento de las imágenes seleccionadas.

A través del módulo tf.data de Tensorflow se realizó la creación del set de datos, el cual permite una mayor flexibilidad y personalización además de mejorar el rendimiento. Este proceso se inicia con el redimensionado de las imágenes a un tamaño standard de 224 x 224 píxeles utilizado en modelos pre entrenados.

- Balanceo de datos. Debido al desequilibrio en la distribución de las imágenes para el grupo 1 y 2, se utilizó la técnica de adición de pesos en la función de pérdida para compensar a la clase con menos muestras (malos resultados). Así se asignaron pesos diferentes a cada clase en la función de pérdida. Para obtener los pesos se aplicó la técnica de frecuencias inversas.
- Modelos pre entrenados. Se realizó primero un preprocesamiento y aumento de datos, cambiando el tamaño de las imágenes (hecho anteriormente) y normalizando los píxeles dependiendo del modelo. Luego se procedió a generar las arquitecturas basadas en redes pre entrenadas. Para este fin se extraen las características aprovechando los pesos entrenados previamente en el conjunto de datos, haciendo uso

únicamente de las capas convolucionales que extraen características. Luego se agrega una capa de agrupamiento promedio global para obtener un vector característico que es ingresado al clasificador, el cual está basado en una capa densa con una unidad de salida que usa la función sigmoide. Estas redes se entrenan durante 10 épocas con el optimizador AdamW y una tasa de aprendizaje 1e-5. Se utiliza *dropout* con una tasa de 0.5 y *weight decay* de 1e-2 para reducir sobreajuste.

Estos modelos usan las siguientes etapas de control:

*Model checkpoint*: se guarda el mejor modelo encontrado durante el entrenamiento.

*Early stopping*: para detener el entrenamiento en caso de que no se observen mejoras en el error de validación.

*Reduce LR On Plateau*: para reducir la tasa de aprendizaje en un factor de 0.5 si no hay mejoras en el error de validación ajustándola hasta un mínimo de 1e-6.

Se realizaron luego predicciones de la data de entrenamiento para verificar si el modelo generaliza correctamente. Para este fin se creó un nuevo dataset de Tensorflow aplicando el preprocesamiento que es el mismo que se usa en la data de validación.

Para evaluar los resultados de la data de prueba se creó el dataset en Tensorflow siguiendo el mismo proceso que la data de validación, obteniéndose las probabilidades y predicciones de la data de prueba en todos los modelos.

Este estudio tiene implicaciones éticas al usar imágenes faciales de una población infantil. Contamos con consentimientos informados para la realización de cirugías, pero no para la utilización de material fotográfico de los pacientes. Es por ello que el análisis fotográfico realizado solo incluye el tercio inferior facial, de tal forma que protegimos el anonimato de los pacientes al no ser identificables. Sin embargo, podemos agregar que, de acuerdo al Reglamento General de Protección de Datos, se considera legal al ser necesario el tratamiento de estos datos para el cumplimiento de una misión realizada de interés público en el ámbito de la salud pública (Artículo 6 apartado 1 letra e del RGPD) y al ser necesario con fines de archivo de interés público, fines de investigación científica (Artículo 9 apartado 2 letra j). Asimismo, el tratamiento de los datos fue llevado a cabo por un profesional sujeto a la obligación de secreto profesional. De acuerdo a la normativa vigente, basándonos en el artículo 18.1 de la Constitución Española, Reglamento General de Protección de Datos (RGPD), así como el reglamento 2016/679 del Parlamento Europeo que garantiza el derecho a la intimidad personal y protección de las personas físicas en relación al tratamiento de datos personales, se ocultó la

identidad de las mismas a través de la seudonimización de datos personales. (Artículo 32 apartado 1 letras a y b del RGPD).

Dado que la población en estudio corresponde a una población infantil se debe tomar en cuenta el Artículo 8 del RGPD, en relación al consentimiento del niño, que se delega a los padres o tutores, manteniéndose así el anonimato de las imágenes usadas en el estudio.

En relación a la normativa de protección de datos en Perú (país de donde se estudió la muestra de datos), esta se encuentra contemplada en el Decreto Supremo 003-2013-JUS Reglamento de la Ley de Protección de Datos Personales, que regula además el régimen sancionador.

## Resultados

### Modelo pre entrenado con ResNet-50

En esta data de entrenamiento, el modelo ResNet logra una clasificación casi perfecta, con muy pocos falsos positivos y falsos negativos.

En la data de validación, el modelo clasifica con algunos falsos positivos y falsos negativos, para valorar con mayor precisión se usarán los resultados de las métricas de evaluación.

Las métricas de entrenamiento arrojan resultados altos para ambas clases, aunque se tiene un sobreajuste debido a que los resultados de validación son más bajos.

En las métricas de validación se observa un buen rendimiento con una exactitud de 0.85, sin embargo, estos resultados son menores a los obtenidos en el entrenamiento, lo que sugiere un sobreajuste que podría afectar su generalización.

En relación a las curvas ROC-AUC (ROC: Curva de característica operativa del receptor; AUC: Área bajo la curva), la curva de entrenamiento es casi perfecta, siendo su valor de 0.997 (Fig. 4) En la curva de validación, no es tan perfecta como la de entrenamiento, siendo el valor del AUC de 0.915. Se confirma con estas curvas el sobreajuste del modelo (Fig. 5).

### Modelo pre entrenado con MobileNet

El error de entrenamiento y validación disminuyen de forma paralela en las primeras épocas, pero luego se separan debido a un posible sobreajuste. Las curvas de exactitud de entrenamiento y validación aumentan correctamente.

La métricas de entrenamiento usan el conjunto de datos de entrenamiento, pero con el mismo preprocesamiento que la data de validación. En esta data, este modelo demuestra una alta precisión en el entrenamiento. Comparado con el modelo anterior, este presenta más falsos negativos.

En las métricas de validación reduce el número de falsos positivos en comparación al ResNet, pero se comporta de manera similar con sobreajuste.

En las métricas de entrenamiento para este modelo se observaron resultados sobresalientes, con una exactitud promedio de 0.96, indicando un aprendizaje muy efectivo a partir de esta data. En las de validación, el modelo muestra buen rendimiento con una exactitud promedio de 0.91, aunque con discrepancia entre los resultados de

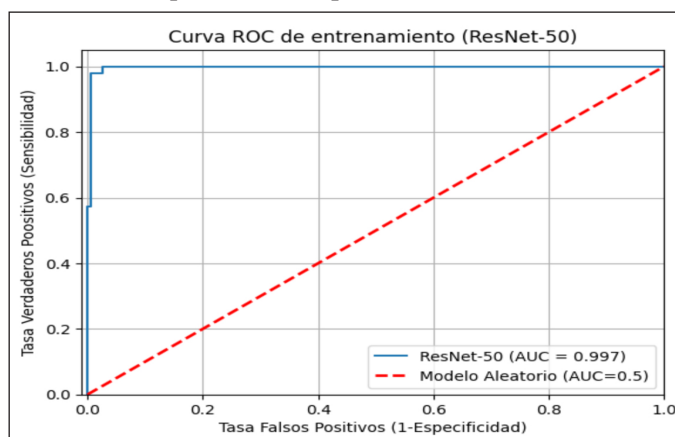


Figura 4. Curva ROC de entrenamiento para ResNet-50.

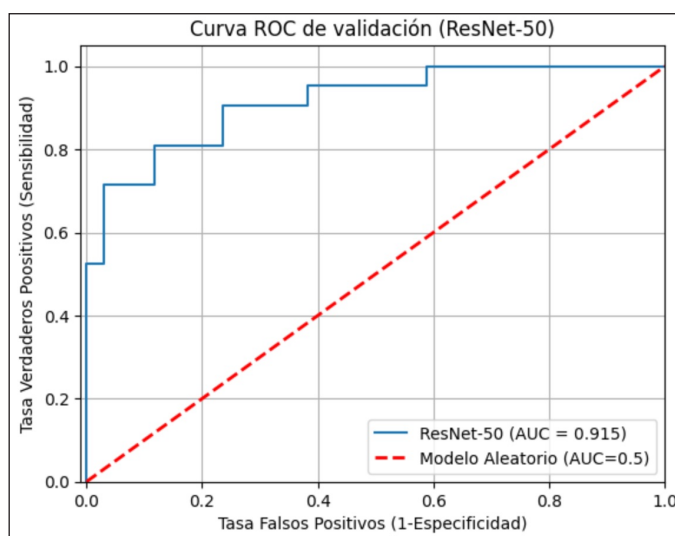


Figura 5. Curva ROC de validación para ResNet-50.

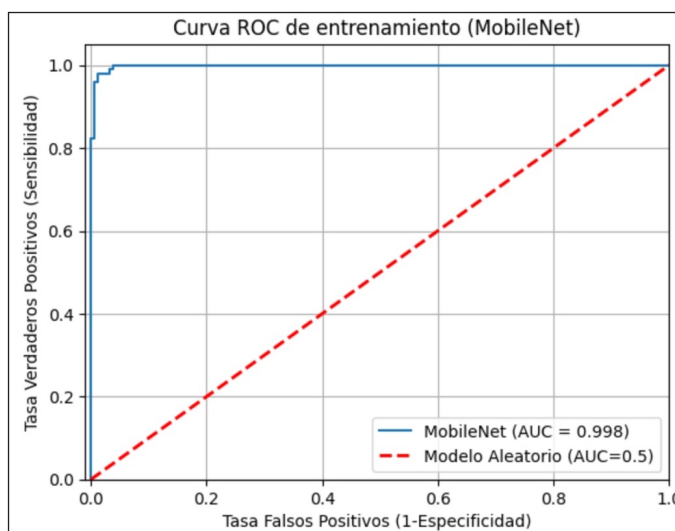


Figura 6. Curva ROC de entrenamiento para MobileNet.

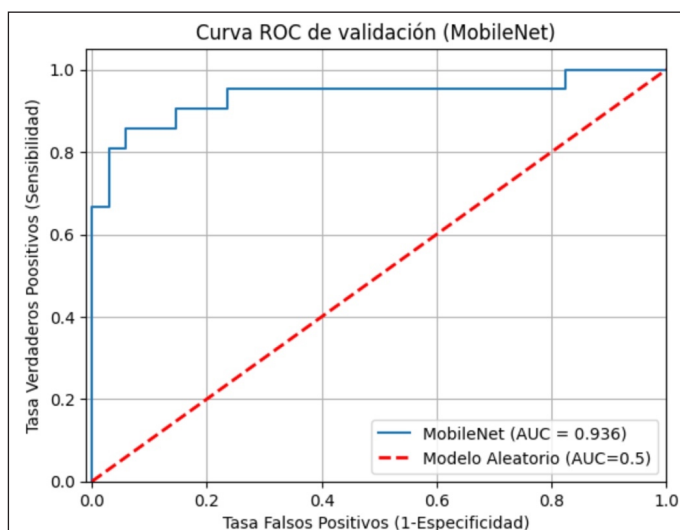


Figura 7. Curva ROC de validación para MobileNet.

entrenamiento y validación al igual que el modelo anterior. Finalmente, para este modelo, la curva ROC de entrenamiento es también casi perfecta, con un valor de 0.998. La curva ROC de validación muestra un valor promedio de 0.936 (Fig. 6 y 7).

### Modelo pre entrenado con EfficientNetV2

Las curvas de aprendizaje de entrenamiento y validación muestran que ambos errores disminuyen de forma paralela en las primeras épocas, pero luego empiezan a separarse mostrando sobreajuste.

Las métricas de entrenamiento muestran una alta precisión con este modelo. En la data de validación, se observa una mejora en comparación a los modelos anteriores con menor número de falsos positivos.

Así, el modelo EfficientNetV2-B0 logró resultados sobresalientes con una exactitud promedio de 1.00, indicando un aprendizaje muy efectivo a partir de la data de entrenamiento. También se observa un buen rendimiento en la data de validación con un valor promedio de 0.95, siendo mejor que los anteriores. Sin embargo, comparte el mismo problema de sobreajuste.

La curva ROC-AUC de entrenamiento para este modelo es casi perfecta, con un valor de 1.00 (Fig. 8). La curva de la data de validación no es tan perfecta, con un valor de 0.966 (Fig. 9).

En el modelo ResNet-50, se identificaron 29 imágenes como buenas, indicando una buena capacidad para detectar casos buenos, pero se identificaron incorrectamente 5 casos como malos (falso positivo). Adicionalmente, clasificó 9 imágenes malas como buenas (falsos negativos), mostrando una dificultad del modelo para identificar casos malos. Solo identificó 12 casos malos como tales (verdaderos positivos).

El modelo MobileNet mostró mejores resultados en los datos de prueba en comparación con el ResNet-50,

con menos falsos positivos y negativos. Sin embargo, la alta cantidad de falsos negativos hace que los resultados sean menores que la validación.

Finalmente, el modelo EfficientNetV2-B0 mostró una ligera mejora en los falsos negativos, sugiriendo que el modelo tiene menos probabilidades de error en los casos malos. También se observó un aumento de falsos positivos en comparación a MobileNet.

En relación a las métricas de evaluación, el modelo ResNet-50 mostró un rendimiento bajo en la data de prueba en comparación con lo observado en la validación. La exactitud promedio fue similar, 0.745 y 0.739. Este modelo discrimina mejor los casos buenos, pero falla en los malos, tal como se ve en la precisión por clase.

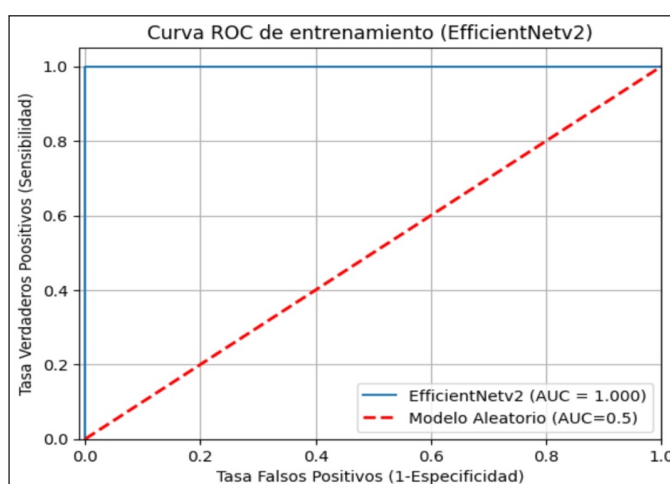


Figura 8. Curva ROC de entrenamiento para EfficientNetV2-B0.

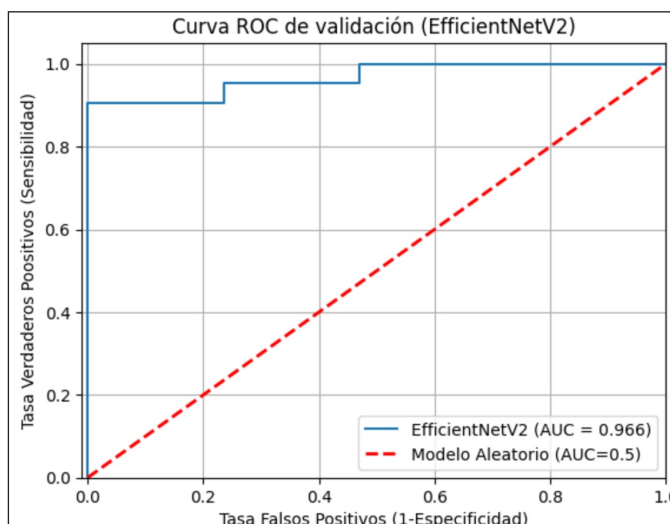


Figura 9. Curva ROC de validación para EfficientNetV2-B0.

En el modelo MobileNet se tuvo una exactitud de 0.818 en los datos de prueba, mejor que los resultados observados con ResNet-50.

Finalmente, el modelo EfficientNetV2-B0 logró un rendimiento similar al modelo MobileNet con una exactitud de 0.818, pero al igual que los modelos anteriores tiene dificultades para identificar correctamente los casos malos.

La comparación de la sensibilidad y especificidad de los 3 modelos usando la curva ROC muestra que estos superan los resultados aleatorios. De ellos, el mejor es EfficientNetV2-B0 con un AUC de 0.88, seguido de MobileNet con 0.86 y ResNet-50 con 0.817 (Fig. 10).

Para la comparación final, en la Fig. 10 observamos que el peor modelo es el ResNet-50, con valores en las métricas por debajo de 0.8. Tanto MobileNet como EfficientNetV2-B0 obtienen resultados similares, sin embargo, en cuanto al número de parámetros, el modelo MobileNet es más eficiente y menos costoso que los demás; así, este modelo es el mejor por ser más eficiente y de rendimiento similar al EfficientNetV2-B0.

Para finalizar la valoración comparativa de estos modelos, se muestra la aplicación de la técnica Grad-CAM que consiste en buscar en qué partes de la imagen la red neuronal convolucional ha basado su decisión. De esta manera se generan mapas de calor que muestran qué partes de la imagen fueron más significativas para la predicción de la clase respectiva, utilizando las ultimas capas de los modelos pre entrenados (Fig. 11).

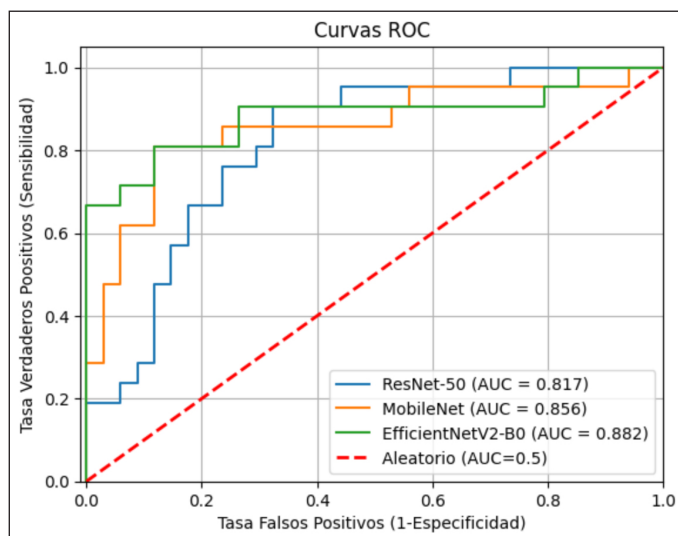


Figura 10. Curvas ROC de sensibilidad y especificidad de los 3 modelos.

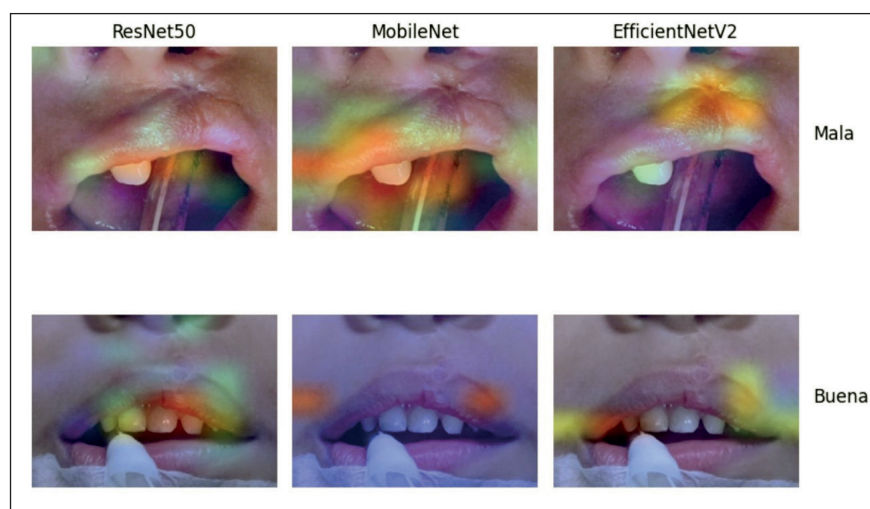


Figura 11. Mapas de calor visuales obtenidos a través de la técnica Grad-CAM mostrando las regiones de la imagen que fueron más significativas para la predicción de cada clase.

Así se puede observar que en casos malos los modelos se enfocan mayormente en el labio superior, siendo los mejores resultados los de EfficientNetV2-B0 y MobileNet.

## Discusión

La inteligencia artificial ha sido utilizada en el campo de las fisuras labio-palatinas principalmente para predecir el riesgo de tener una gestación con esta malformación congénita, así como en el diagnóstico prenatal y en la identificación de las características cefalométricas de estos pacientes. Asimismo, en el tratamiento prequirúrgico y en la terapia de rehabilitación de la voz y el habla.

En relación a los antecedentes, y hasta donde hemos podido revisar, no hemos encontrado estudios con un objetivo similar al planteado en este proyecto (valoración de resultados quirúrgicos), aunque sí existen estudios relacionados a la aplicación de inteligencia artificial en el diagnóstico y tratamiento de las fisuras labio-palatinas. Estos han dado origen a estudios de tipo revisión sistemática en la actualidad. Así, un estudio recientemente publicado por Hugh y col.<sup>(8)</sup> incluyó artículos originales de metodología observacional retrospectiva evaluando la utilidad de la inteligencia artificial en el diagnóstico o predicción del tratamiento de pacientes con fisura labio-palatina. Encontraron en 12 artículos seleccionados que las técnicas de inteligencia artificial pueden usarse en actividades específicas relacionadas a esta patología, tales como detección de puntos de referencia anatómicos, análisis cefalométrico digital rápido, toma de decisiones clínicas y predicción de tratamientos. Las variables de resultado más medidas fueron la detección de voz hipernasal y predicción de riesgo genético, así como los factores predictivos de riesgos en cirugía ortognática. Todos estos estudios se desarrollaron en un solo centro y usando data de la población local.<sup>(9-15)</sup>

Otro estudio realizado por Dhillon y col. de tipo revisión sistematizada incluyó 26 artículos que exploraban el uso de la inteligencia artificial en el diagnóstico, tratamiento y predicción de las fisuras labio palatinas.<sup>(16)</sup> McCullough y col utilizan las técnicas de inteligencia artificial, específicamente modelos de redes neuronales convolucionales, para medir la severidad preoperatoria de la fisura en pacientes con fisura labial unilateral.<sup>(17)</sup> En este estudio, recolectan imágenes preoperatorias de 800 fisuras labiales unilaterales identificando puntos anatómicos de referencia específicos que fueron eva-



luados usando una escala de severidad validada por opinión de expertos. Luego, entrenan 5 modelos de redes neuronales convolucionales para detección de los puntos anatómicos y su respectivo grado de severidad de afectación.

Por otro lado, encontramos también publicados diferentes estudios en determinación de simetría facial utilizando inteligencia artificial, aunque no en pacientes con fisuras, pero con una metodología que puede ser comparada. Un primer estudio presentado por Hsiu-Hsia Lin y col.<sup>(18)</sup> utiliza técnicas de aprendizaje profundo para evaluar resultados operatorios de corrección de mala oclusión dental mediante cirugía ortognática a través del grado de simetría facial. En otro estudio publicado por Bobrov y col.<sup>(19)</sup> se entrenaron sistemas de aprendizaje profundo usando fotos de alta resolución de la región palpebral etiquetadas con su edad cronológica. Se desarrolló de esta forma un sistema predictor de edad cronológica. Este modelo fue diseñado para aceptar fotos de diferente resolución y puede ser útil en caso de que la data no este estandarizada. Fue el modelo usado en el presente estudio.

Existen otros trabajos, como el de Wu y col.<sup>(20)</sup> y el de Mc Cullough y col.<sup>(17)</sup> donde se evaluó la simetría facial de pacientes con fisura labial pero antes de ser operados, a manera de estimación de grados de severidad de la malformación congénita. Sin embargo, no hemos encontrado ningún estudio realizado con el objetivo de establecer un diagnóstico postoperatorio que nos permita planificar mejoras en la atención interdisciplinaria. Esta es una limitación que no nos permite establecer parámetros comparativos de los hallazgos del estudio que presentamos.

La calidad de resultados obtenidos en cirugías de fisuras labio-palatinas es un tema del cual se tiene muy poca información y se mide por lo general de forma subjetiva. En el contexto del apoyo que brindan organizaciones de tipo no gubernamental a países en desarrollo, existe un interés por parte de estos en conocer la calidad de servicio que brindan sus socios, a los cuales se apoya económicamente para la realización de estas cirugías. Además, estas organizaciones brindan una importante inversión destinada a capacitación profesional e investigación, herramientas fundamentales en el mejor conocimiento del entorno relacionado a esta enfermedad.

En el análisis del presente estudio, los 3 modelos fueron entrenados a través de *fine-tuning* para adaptarlos a tareas específicas, y en todos ellos se congelaron la mayoría de capas y se entrenaron las últimas.

Una limitación importante fue el desbalance en el número de muestras por clase, para lo cual se tuvo que aplicar la técnica de frecuencias inversas aplicando pe-

sos a la función de pérdida, atenuando el problema. Esta diferencia en los grupos se explica por el menor número de malos resultados del autor, teniendo que recurrir a casos recibidos para corrección secundaria y operados por otros cirujanos.

Otra de las principales limitaciones es el sobreajuste observado debido a la poca cantidad de datos, el cual se observó en los 3 modelos usados. Esto a pesar del uso de técnicas como *dropout*, *weight decay*, *early stopping* y un aumento ligero de datos. Se pueden mejorar estas limitaciones aumentando la data para que el modelo pueda generalizar de mejor manera, y la calidad de las mismas al enfocar mejor la zona anatómica a evaluar.

Una última limitación a considerar es que no hay una correlación de los resultados evaluados con la severidad congénita de la fisura operada; se puede estimar que una fisura más severa puede limitar la posibilidad de mejores resultados. Este es un aspecto que debe ser mejorado en el proceso futuro de entrenamiento de los modelos.

El enfoque de las zonas anatómicas realizado por los modelos entrenados mostró diferencias entre los grupos; así, el uso del Grad-CAM mostró que los modelos enfocaron mejor la zona anatómica de interés en los casos malos que en los casos buenos, lo que puede explicar los resultados obtenidos cuantitativamente. Se sugiere mejorar los modelos con un mejor preprocesamiento de las imágenes enfocando mejor el labio superior, así como también probar otras arquitecturas.

El presente es un estudio preliminar, y como toda tecnología de aprendizaje automático (*machine learning*) requerirá una reingeniería continua para mejorar su eficiencia.

Dado que el presente es un proyecto personal, la data no se encuentra disponible en ningún repositorio actualmente; está en el registro personal del PC de los autores, pero se encuentra disponible. Se podrían haber usado las bases de datos de la Fundación Smile Train, pero el principal problema es que esta base de datos no registra resultados postoperatorios faciales.

## Conclusiones

El modelo de inteligencia artificial utilizado demuestra su habilidad para medir la simetría labial postoperatoria en pacientes operados de fisura labial unilateral.

En nuestro estudio, los mejores modelos fueron MobileNet y EfficientNetV2-B0, con resultados por encima del 80% en todas las métricas.

Consideramos que son necesarios estudios futuros de naturaleza multicéntrica basados en modelos de inteligencia artificial con mayor número de datos para validar estos modelos en la práctica clínica. Esta información



servirá para determinar la calidad de cirugía practicada y así poder planificar actividades de entrenamiento de los equipos quirúrgicos con una mayor objetividad y eficiencia.

## Dirección del autor

Dr. Percy Rossell Perry

Correo electrónico: prossellperry@gmail.com

## Bibliografía

1. **Rossell-Perry P.** Tratamiento de la fisura labio-palatina. Fondo Editorial Universidad San Marcos Lima, 2009.
2. **Frederick R, Hogan AC, Seabolt N, Stocks RM.** An ideal multidisciplinary cleft lip and cleft palate care team. *Oral Diseases*, 2002;28(5):1412-1417.
3. **Rossell-Perry P, Segura E, Salas-Bustinza L, Cotrina-Rabanal O.** Comparison of two models of surgical care for patients with cleft lip and palate in resource-challenged settings. *World J. Surg.* 2015;39(1):47-53.
4. **Hendriks TC, Botman M, Rahmee CN, Ket JC, Mullender MG, Gerretsen B.** Impact of short-term reconstructive surgical missions: a systematic review. *BMJ Global Health*;2019; 4(2):e001176.
5. **Celeghin A, Borriero A, Orsenigo D, Diano M. et al.** Convolutional neural networks for vision neuroscience: significance, developments, and outstanding issues. *Frontiers in Computational Neuroscience* 2023, 17.
6. **Yamashita R, Nishio M, Do R, Togashi K.** Convolutional neural networks: an overview and application in radiology. *Insights into Imaging*, 2018;9 (4): 611-629.
7. **Cruz YJ, Rivas M, Quiza R, Villalonga A, et al.** Ensemble of convolutional neural networks based on an evolutionary algorithm applied to an industrial welding process. *Computers in Industry*. 2012; 133.
8. **Hugh MZU, Abdullah JY, Wong LS, Jamayet NB, Alam MK, Rashid QF, et al.** Clinical Applications of Artificial Intelligence and Machine Learning in Children with Cleft Lip and Palate: A systematic review. *Int J of Envir Res and Pub Health*. 2022; 19(17):10860.
9. **Machado R. de Oliveira Silva C, Martelli-Junior H, Das Neves L.** Machine learning in prediction of genetic risk of non-syndromic oral clefts in the Brazilian population. *Clin. Oral. Investig*, 2021;25(3):1273-1280.
10. **Zhang SJ, Meng P, Zhang J, et al.** Machine learning models for genetic risk assessment of infants with non-syndromic orofacial cleft. *Genom. Proteom. Bioinform*, 2018;16(5):354-364.
11. **Alam MK, Alfawzan AA, Haque S, et al.** Sagittal Jaw Relationship of Different Types of Cleft and Non-cleft Individuals. *Front. Pediatr*, 2021;5:9:651951.
12. **Wang X, Yang S, Tang M, Yin H, Huang H.** Hypernasalitynet: Deep recurrent neural network for automatic hypernasality detection. *Int. J. Med. Inform*, 2019;129: 1-12.
13. **Golabbakhs M, Abnavi F, Kadkhodaei Elyaderani M, et al.** Automatic identification of hypernasality in normaland cleft lip and palate patients with acoustican analysis of speech. *J. Acoust. Soc. Am* 2017;141(2): 929-935.
14. **Wang X, Tang M, Yang S, Yin H, Huang H, He L.** Automatic hypernasality detection in clef tpalate speech using cnn. *Circuits Syst. Signal Process*, 2019;38(8):3521-3547.
15. **Alam M, Alfawzan A.** Dental Characteristics of Different Types of Cleft and Non-cleft Individuals. *Front. Cell Dev. Biol*, 2020; 8: 789.
16. **Dhillon H, Chaudhari P, Dhingra K, Kuo R, Sokhi R, et al.** Current applications of artificial intelligence in cleft care: A scoping review. *Front Med*, 2021; 8: 676490.
17. **McCullough M, Ly S, Auslander A, et al.** Convolutional Neural Network Models for Automatic Preoperative Severity Assessment in Unilateral Cleft Lip. *Plast Reconstr Surg*. 2021;148(1): 162-169.
18. **Hsiu-Hsia Lin, Lun-Jou Lo, Wen-Chung Chiang.** A novel assessment technique for the degree of facial symmetry before and after Orthognatic surgery based on three dimensional contour features using deep learning algorithms. *Int J of Advances in Science Engineering and Technology*, 2018;6(2):2321-2991.
19. **Bobrov E, Georgievskaya A, Kiselev K, Sevastopolsky A, et al.** PhotoAge Clock: deep learning algorithms for development of non-invasive visual biomarkers of aging. *Aging*, 2018;10(11): 3249-3259.
20. **Wu J, Heike C, Birgfeld C, Evans K, Maga M, et al.** Measuring symmetry in children with unrepaired cleft lip: defining a standard for the three-dimensional midfacial reference plane. *Cleft Palate Craniofac J*. 2016; 53(6): 695-704.

