

UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE INFORMÁTICA
VALPARAÍSO - CHILE



“PROPUESTA DE MODELO DE ML QUE APOYE LA
DETECCIÓN TEMPRANA DE CÁNCER BUCAL A TRAVÉS
DE TÉCNICAS DE APRENDIZAJE AUTOMÁTICO”

JORGE KEVIN BECKER OYARZO

MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL EN INFORMÁTICA

Profesor Guía: Carlos Valle Vidal
Profesor Correferente: Ricardo Ñanculef

Marzo - 2025

DEDICATORIA

Dedicado a mi familia y mis abuelos, Rubén, Ana, Rina y Yuri.

AGRADECIMIENTOS

Quiero agradecer primeramente a aquellos que han sido mi pilar fundamental todos estos años, a mis padres que se han sacrificado para que yo haya llegado a este momento, inculcándome los valores, virtudes y cualidades que me han guiado hasta aquí. A mis hermanas que me acompañaron en cada aventura, han sacudido mis temores y me han apoyado en cada tragedia y logro en mi vida.

Ellos han sido mis impulsores y les tengo mi eterna gratitud y amor.

A mis abuelos, que fueron aquellos que me brindaron la sabiduría para enfrentar cada obstáculo, aprender de cada derrota y celebrar cada victoria. Un beso al cielo a los que ya no encuentran conmigo y ojala pueda hacerlos sentir orgullosos del nieto que los extraña cada día.

Quiero agradecer a mis amigos y compañeros que me han acompañado desde el primer día, ayudándome a ser el profesional que hoy nace en mí, acompañándome en las largas noches de estudio y brindándome de su compañía, amistad y confianza: Josías, Pablo, Carlos, Gustavo, Diego, Ricardo, Natalia, Paulina, Josefa, Claudia, Amanda, Patryk.

Mención honrosa a Matías, que se ha convertido en mi hermano de mil batallas, brindándome su amistad, apoyo, inteligencia y sabiduría todos estos años.

Y finalmente, quería dedicarle unas palabras a mis profesores: Carlos Valle, Mauricio Figueroa y Pedro Godoy que me han acompañado todos estos años, enseñando, apoyando y brindándome su disposición (y paciencia) para construir el profesional que soy hoy.

¡Gracias Totales!

RESUMEN

El cáncer se ha consolidado como la segunda causa de muerte en Chile, y el diagnóstico tardío es un factor determinante que reduce significativamente la sobrevivencia del paciente. En particular, se prevé que la incidencia del cáncer oral aumente en los próximos años. Esta enfermedad, de carácter silencioso, suele detectarse en etapas avanzadas, cuando ya ha hecho metástasis, lo que disminuye drásticamente las posibilidades de tratamiento efectivo y recuperación.

Hasta la fecha, los métodos tradicionales de diagnóstico —como los clínicos, ópticos, histológicos y las biopsias— han sido las principales herramientas utilizadas para la detección precisa de esta patología. Sin embargo, con el avance de la tecnología y el desarrollo de técnicas de Machine Learning, han surgido nuevas alternativas que permiten optimizar y agilizar los procesos diagnósticos mediante el análisis de imágenes médicas. Los estudios revisados en el estado del arte han explorado distintas fuentes de imágenes, incluyendo muestras de tejido obtenidas a través de biopsias y registros visuales de zonas específicas de la cavidad oral, como la lengua.

En este trabajo, se adoptó un enfoque más generalizable al utilizar imágenes de diversas secciones de la cavidad bucal, abarcando no solo la lengua, sino también mejillas, paladar y labios, en conjunto con la implementación de un ensamble de modelos (ResNet50, Xception e InceptionV3). Esto permite desarrollar modelos más robustos capaces de detectar patrones asociados a lesiones malignas en distintas áreas de la boca. Los resultados obtenidos destacan el potencial del enfoque propuesto, alcanzando un F1-score de 0,878 y un AUC-ROC de 0,87.

Estos resultados demuestran la viabilidad del aprendizaje automático como una herramienta complementaria en el diagnóstico médico, con el potencial de integrarse en los servicios de salud y contribuir significativamente a la detección temprana del cáncer oral.

—**Palabras Clave:** redes neuronales, cáncer, clasificación de imágenes, detección

ABSTRACT

Cancer has become the second leading cause of death in Chile, with late diagnosis being a critical factor that significantly reduces patient survival. In particular, the incidence of oral cancer is expected to increase in the coming years. This silent disease is often detected at advanced stages when metastasis has already occurred, drastically reducing the chances of effective treatment and recovery.

To date, traditional diagnostic methods—such as clinical, optical, histological techniques, and biopsies—have been the primary tools used for the accurate detection of this pathology. However, with technological advancements and the development of Machine Learning techniques, new alternatives have emerged to optimize and accelerate diagnostic processes through medical image analysis. The studies reviewed in the state of the art have explored various image sources, including tissue samples obtained through biopsies and visual records of specific areas of the oral cavity, such as the tongue.

This study adopts a more generalizable approach by utilizing images from various sections of the oral cavity, covering not only the tongue but also the cheeks, palate, and lips. Additionally, it implements an ensemble of models (ResNet50, Xception, and InceptionV3). This approach enables the development of more robust models capable of detecting patterns associated with malignant lesions in different areas of the mouth. The results highlight the potential of the proposed approach, achieving an F1-score of 0,878 and an AUC-ROC of 0,87.

These findings demonstrate the feasibility of Machine Learning as a complementary tool in medical diagnostics, with the potential to integrate into healthcare services and make a significant contribution to the early detection of oral cancer.

—**Keywords:** neural networks, cancer, image classification, detection

GLOSARIO

DI: Departamento de Informática.

IA: Inteligencia Artificial.

UTFSM: Universidad Técnica Federico Santa María.

ML: Machine Learning (Aprendizaje Automático)

MINSAL: Ministerio de Salud

OPMD: Oral Potentially Malignant Disorders (Trastornos Orales Potencialmente Malignos).

AIRC: Agencia Internacional para la Investigación en Cáncer.

DSS: Sistemas de Soporte a las Decisiones.

RNN: Recurrent Neural Networks (Redes Neuronales Recurrentes).

CNN: Convolutional Neural Networks (Redes Neuronales Convolucionales).

ReLU: Rectified Linear Unit (Unidad lineal rectificadora).

Vision Computer: Visión por computador.

RGB: Red, Green, Blue.

SAM: Segment Anything Model.

SGD: Stochastic Gradient Descent.

ADAM: Adaptive Moment Estimation.

FPR: False Positive Rate (Tasa de Falsos Positivos)

TPR: True Positive Rate (Tasa de Verdaderos Positivos)

ÍNDICE DE CONTENIDOS

RESUMEN	IV
ABSTRACT	V
GLOSARIO	VI
ÍNDICE DE FIGURAS	IX
ÍNDICE DE TABLAS	X
INTRODUCCIÓN	1
CAPÍTULO 1: DEFINICIÓN DEL PROBLEMA	2
1.1 CONTEXTO	2
1.2 TRASTORNOS DE MUCOSA	2
1.3 PROCEDIMIENTO ACTUAL	3
1.4 PROBLEMA DETECTADO	4
1.5 OBJETIVOS	7
1.6 CONTRIBUCIÓN	7
CAPÍTULO 2: MARCO CONCEPTUAL	8
2.1 INTELIGENCIA ARTIFICIAL	8
2.2 APRENDIZAJE AUTOMÁTICO	8
2.3 REDES NEURONALES ARTIFICIALES	10
2.3.1 CONCEPTOS A CONSIDERAR	11
2.3.2 REDES NEURONALES CONVOLUCIONALES	15
2.3.3 ENSAMBLE DE MODELOS	15
2.3.4 MÉTRICAS PARA EVALUACIÓN DE MODELO	17
2.4 HERRAMIENTAS DE SOPORTE DE DECISIONES CLÍNICAS	18
2.4.1 CARACTERÍSTICAS COMUNES IDENTIFICADAS	18
2.4.2 BENEFICIOS Y LIMITACIONES	19
2.4.3 CONSIDERACIONES DEL MODELO	20
2.5 TRABAJOS RELACIONADOS	20
CAPÍTULO 3: PROPUESTA DE SOLUCIÓN	23
3.1 CARACTERÍSTICAS RELEVANTES DEL PROBLEMA	23
3.1.1 ESPECIFICACIONES DEL PROBLEMA	23
3.1.2 DESAFÍOS DEL PROYECTO	23
3.1.3 RELEVANCIA DEL PROYECTO	23
3.1.4 VARIABLES CLAVE Y PREPROCESAMIENTO	24
3.1.5 MODELOS SELECCIONADOS	24
3.1.6 EVALUACIÓN DEL RENDIMIENTO	24
3.2 CONSIDERACIONES ÉTICAS Y DE PRIVACIDAD	25

3.3	DESCRIPCIÓN GENERAL DE LA SOLUCIÓN PROPUESTA	27
3.4	ALGORITMOS Y TÉCNICAS DE ENTRENAMIENTO Y EVALUACIÓN DEL MODELO	29
3.4.1	XCEPTION	29
3.4.2	INCEPTIONV3	30
3.4.3	RESNET	30
3.4.4	MÉTRICAS, EVALUACIÓN Y COMPARACIÓN	31
3.5	ARQUITECTURA DEL MODELO DE APRENDIZAJE AUTOMÁTICO PROPUESTO .	32
3.5.1	PROCESAMIENTO DE DATA	32
3.5.2	MODELOS DE CLASIFICACIÓN	34
3.5.3	ETAPA DE ENTRENAMIENTO	38
3.6	DESCRIPCIÓN DE DATOS	41
CAPÍTULO 4: VALIDACIÓN DE LA SOLUCIÓN		59
4.1	RESULTADOS Y MÉTRICAS DEL MODELO PROPUESTO	59
4.1.1	ARQUITECTURAS Y MODIFICACIONES	59
4.1.2	EJEMPLOS PREDECIDOS	74
4.2	COMPARACIÓN DE DESEMPEÑO	75
4.2.1	MÉTODOS TRADICIONALES - CLÍNICOS	75
4.2.2	MÉTODOS BASADOS EN MACHINE LEARNING	79
CAPÍTULO 5: CONCLUSIONES		83
REFERENCIAS BIBLIOGRÁFICAS		85

ÍNDICE DE FIGURAS

1	<i>Número de muertes por cáncer en Chile (período año 1986 y 2016).</i>	5
2	<i>Ejemplo de Red Neuronal.</i>	11
3	<i>Ejemplo de Red Neuronal FeedForward</i>	12
4	<i>Ejemplo de Red Neuronal Recurrente</i>	12
5	<i>Ejemplo de Red Neuronal Convolutacional</i>	13
6	<i>Representación de Soft Voting en Machine Learning</i>	16
7	<i>Diagrama general del proceso</i>	27
8	<i>Diagrama de arquitectura agregada a Xception.</i>	35
9	<i>Diagrama de arquitectura agregada a ResNet50.</i>	36
10	<i>Diagrama de arquitectura agregada a InceptionV3.</i>	37
11	<i>Diagrama general del ensamble.</i>	37
12	<i>Diagrama general de votación del ensamble.</i>	38
13	<i>Gráficos de Proporciones (Train, Valid, Test, Total)</i>	42
14	<i>Comparación entre ejemplos de lesiones benignas y malignas.</i>	42
15	<i>Gráficos de Proporciones (Train, Valid, Test, Total)</i>	44
16	<i>Comparación entre ejemplos de lesiones benignas y malignas.</i>	44
17	<i>Gráficos de Proporciones (Train, Valid, Test, Total)</i>	46
18	<i>Comparación entre ejemplos de lesiones benignas y malignas.</i>	46
19	<i>Gráficos de Proporciones (Train, Valid, Test, Total)</i>	48
20	<i>Comparación entre ejemplos de lesiones benignas y malignas.</i>	48
21	<i>Gráficos de Proporciones (Train, Valid, Test, Total)</i>	50
22	<i>Comparación entre ejemplos de lesiones benignas y malignas.</i>	50
23	<i>Gráficos de Proporciones (Train, Valid, Test, Total)</i>	52
24	<i>Comparación entre ejemplos de lesiones benignas y malignas.</i>	52
25	<i>Gráficos de Proporciones (Train, Valid, Test, Total)</i>	54
26	<i>Comparación entre ejemplos de lesiones benignas y malignas.</i>	54
27	<i>Gráficos de Proporciones (Train, Valid, Test, Total)</i>	56
28	<i>Comparación entre ejemplos de lesiones benignas y malignas.</i>	56
29	<i>Distribución total en el dataset.</i>	57
30	<i>Distribución de clases en cada subconjunto del dataset.</i>	58
31	<i>Distribución de Train, Valid y Test.</i>	58
32	<i>Matriz de confusión del modelo Xception.</i>	60
33	<i>Gráfico de Loss del modelo Xception.</i>	61
34	<i>Gráfico de ROC-AUC de Validación del modelo Xception.</i>	62
35	<i>Matriz de confusión del modelo ResNet50.</i>	64
36	<i>Gráfico de Loss del modelo ResNet50.</i>	65
37	<i>Gráfico de ROC-AUC de Validación del modelo ResNet50.</i>	66
38	<i>Matriz de confusión del modelo InceptionV3.</i>	68
39	<i>Gráfico de Loss del modelo InceptionV3.</i>	69
40	<i>Gráfico de ROC-AUC de Validación del modelo InceptionV3.</i>	70
41	<i>Matriz de confusión del ensamble.</i>	72

42	<i>Imágenes con etiqueta “benigno”</i>	74
43	<i>Imágenes con etiqueta “maligno”</i>	75

ÍNDICE DE TABLAS

1	<i>Matriz de confusión</i>	17
2	<i>Métricas del modelo Xception en conjunto Test.</i>	60
3	<i>Hiperparámetros de la Arquitectura Adicional a Xception.</i>	63
4	<i>Métricas del modelo ResNet50 en conjunto Test.</i>	64
5	<i>Hiperparámetros de la Arquitectura Adicional a ResNet50.</i>	67
6	<i>Métricas del modelo InceptionV3 en conjunto Test</i>	68
7	<i>Hiperparámetros de la Arquitectura Adicional a InceptionV3</i>	71
8	<i>Métricas del ensamble en conjunto Test</i>	72
9	<i>Tabla Resumen de comparación de métricas entre modelos individuales y ensamble.</i>	73
10	<i>Métricas de los Métodos Tradicionales para la Detección de Lesiones Orales Potencialmente Malignas.</i>	75
11	<i>Métricas de los Modelos Propuestos.</i>	76
12	<i>Comparación Económica y Técnica entre Métodos Tradicionales y Modelos Basados en Aprendizaje Profundo.</i>	78
13	<i>Comparación entre tus modelos y Aubreville et al. (2017)</i>	80
14	<i>Comparación entre tus modelos y Shamim et al. (2019)</i>	81
15	<i>Comparación entre tus modelos y Koriakina et al. (2022)</i>	82

INTRODUCCIÓN

El cáncer oral representa un desafío creciente en el ámbito de la salud pública, con una incidencia en aumento y un diagnóstico que, en la mayoría de los casos, se produce en etapas avanzadas. Esta enfermedad, caracterizada por su progresión silenciosa, suele detectarse cuando las opciones terapéuticas son limitadas y la probabilidad de recuperación es reducida. En Chile, donde el cáncer es la segunda causa de muerte, la detección temprana del cáncer oral se vuelve una necesidad urgente para mejorar la supervivencia de los pacientes y optimizar las estrategias de tratamiento.

Tradicionalmente, los métodos clínicos, ópticos, histológicos y las biopsias han sido las principales herramientas para la identificación de lesiones malignas en la cavidad bucal. Sin embargo, con el avance de la inteligencia artificial y, en particular, del *Machine Learning*, han surgido nuevas alternativas que buscan complementar estos procedimientos mediante el análisis automatizado de imágenes médicas. Investigaciones recientes han explorado diversas fuentes de datos, desde muestras obtenidas a través de biopsias hasta imágenes clínicas de áreas específicas de la boca, como la lengua, demostrando el potencial de estos enfoques en la detección de anomalías.

A lo largo de este documento se desarrolla el planteamiento, diseño y evaluación de un ensemble de modelos de aprendizaje automático para el apoyo a la detección temprana del cáncer oral. En el capítulo 1, se presenta la definición del problema, abordando el contexto clínico, los métodos actuales de detección y la relevancia de una solución basada en inteligencia artificial, además de los objetivos y contribuciones del trabajo. En el capítulo 2, se introduce el marco conceptual, detallando los fundamentos del aprendizaje automático, las redes neuronales convolucionales y las métricas utilizadas para evaluar el rendimiento del modelo, junto con una revisión de trabajos previos en el área. El capítulo 3 describe la propuesta de solución, incluyendo las especificaciones del problema, el procesamiento de datos, la selección de modelos y los algoritmos utilizados en el entrenamiento y evaluación. Posteriormente, en el capítulo 4, se presentan los resultados obtenidos, analizando las métricas de desempeño del modelo y comparándolo con métodos tradicionales y enfoques basados en *Machine Learning*, además de ilustrar ejemplos de predicciones realizadas. Finalmente, en el capítulo 5, se exponen las conclusiones del estudio, resaltando los aportes del trabajo, sus limitaciones y posibles líneas de investigación futura para la mejora e implementación del sistema en entornos clínicos.

CAPÍTULO 1

DEFINICIÓN DEL PROBLEMA

1.1. CONTEXTO

En los últimos 20 años, las patologías oncológicas se han convertido en el principal enemigo de la salud pública, posicionándose como una de las principales causas de mortalidad en el mundo. Chile no es un caso especial, en el 2018 fue lanzado el Plan Nacional del Cáncer en respuesta al avance de esta enfermedad en el país, que se ha convertido en la segunda causa de muerte, solo siendo superada por las enfermedades cardiovasculares (Fernandez, 2022). Si bien existen numerosos tipos de patologías oncológicas, el cáncer bucal es uno de los menos conocidos, a pesar estar entre los 11 tipos de cáncer más comunes; el 90 % de estos casos derivan del epitelio de mucosa oral (carcinoma espinocelular), un tipo de tumor agresivo con alta capacidad metastásica, invasivo y asociado a un mal pronóstico debido a su detección tardía (Colegio Dentistas, 2020).

Actualmente existen pocas publicaciones que especifiquen la magnitud de esta enfermedad en Chile, no obstante, las cifras existentes no son muy alentadoras para aquellos que sufran esta patología. En 2020, el Minsal indicó que existe un aumento en la tasa de mortalidad para el cáncer de labio, cavidad oral y orofaringe, desde 1.0 por 100 mil habitantes en 1997, a 1.4 por 100 mil habitantes en 2015. Especialistas advierten que una de las principales dificultades de esta enfermedad, tanto en Chile como a nivel mundial, es que la mayoría de los casos son descubiertos de manera tardía (Cooperativa.cl, 2020).

Esto conlleva a que alrededor del 70-80 % de los casos sean diagnosticados en etapas donde el cáncer ya se ha ramificado; esto se traduce a una baja tasa de supervivencia a la enfermedad, que se estima es del 50 % a los 5 años, esto quiere decir que solo la mitad de los pacientes sobrevivirá más de 5 años, y las cifras a los 10 años tampoco son alentadoras.

1.2. TRASTORNOS DE MUCOSA

Existen numerosos trastornos que han sido asociados a un alto potencial de patologías cancerígenas malignas, entre ellos es importante mencionar y describir brevemente para contextualizar esta memoria respecto a aquellas señales que han sido identificadas como la manifestación de enfermedades que podrían poner en peligro al paciente. Algunos de estos trastornos son:

- **Leucoplasia:** identificado como la lesión premaligna más frecuente de la cavidad oral (Escribano-Bermejo M. et al, 2009). Es una lesión predominantemente blanca en la mucosa oral. Su descubrimiento suele ser casual, por lo que el tiempo de evolución de la lesión en muchos casos es ambiguo.
- **Eritroleucoplasia:** lesión que aparece como mancha o placa blanca que alterna con zonas notablemente eritematosas¹ (rojizas), de superficie lisa o aplanada y se asocia

¹Condición donde hay enrojecimiento o enrojecimiento anormal de la piel o las mucosas.

a infección por *Candida albicans*². Generalmente es asintomática e identificada como lesión precancerosa, combinación de lesión blanca y roja presentes en la mucosa bucal.

- **Eritroplasia:** lesión roja presente en la mucosa oral que también puede estar asociada con displasia o cáncer oral. Las eritroplasias son menos comunes que las leucoplasias, pero tienen un mayor riesgo de malignidad.
- **Displasia oral:** se considera una lesión precancerosa donde las células en la mucosa oral muestran cambios anormales; esta afección puede eventualmente convertirse en cáncer bucal si no se trata oportunamente.

Estas patologías han sido materia de estudio de profesionales del área como Mateo-Sidrón y Somacarrera-Pérez (2015), López-Jornet *et al.* (1999) y Estrada Pereira *et al.* (2012), cuyas definiciones y marcos conceptuales han permitido sintetizar estas lesiones.

1.3. PROCEDIMIENTO ACTUAL

Generalmente las patologías oncológicas orales son antecedidas a un conjunto de lesiones y afecciones que son asociadas con una mayor probabilidad de progresión del cáncer y se denominan trastornos orales potencialmente malignos (OPMD), estas pueden manifestarse a través de lesiones ulcerosas, de coloración roja o blanca. Para el proceso de diagnóstico de OPMD, entre algunos de los métodos de detección se encuentran:

1. MÉTODOS CLÍNICOS

- **EXAMEN ORAL CONVENCIONAL**

El procedimiento convencional considera la inspección visual y la palpación. Usualmente, es el procedimiento más recomendado. No obstante, algunas señales de cánceres tempranos en la cavidad oral y faríngea son difíciles de detectar a simple vista y, por consiguiente, de diagnosticar de manera oportuna y eficazmente.

- **TINCIÓN VITAL CON AZUL DE TOLUIDINA**

Este procedimiento considera el uso de un colorante metacromático catiónico que se une de manera selectiva a ciertos componentes de los tejidos, específicamente, a las estructuras de ADN y ARN, como también a espacios intracelulares del epitelio displásico³ (como tejidos inflamatorios, regenerativos, entre otros).

²Tipo de hongo que forma parte de la microbiota normal en el cuerpo humano, sin embargo, en ciertas circunstancias puede proliferar y causar una infección.

³Condición donde las células que recubren los órganos del cuerpo manifiestan cambios anormales de apariencia y/o comportamiento.

2. MÉTODOS ÓPTICOS

Este método considera el uso de luz en cierta longitud de onda generada a través de diversas fuentes, como la quimioluminiscencia (Huber, 2018) o mediante un diodo emisor de luz alimentado por batería (Rangel-Rivera *et al.*, 2009) para la evaluación de los tejidos.

3. MÉTODOS HISTOLÓGICOS

Este procedimiento es generalmente bien tolerado y se considera mínimamente invasivo y seguro. Consiste en la recolección de células de la mucosa bucal para su análisis para la identificación de células cancerosas o biomarcadores potencialmente malignos.

4. MÉTODOS DE MACHINE LEARNING

Se han desarrollado modelos basados en redes neuronales convolucionales (CNNs) entrenados con distintas fuentes de imágenes médicas. Se han aplicado modelos a biopsias e imágenes histológicas, cortes ópticos de células y tratadas con colorantes con LeNet-5, y fotografías clínicas RGB con arquitecturas como ResNet50, VGG19, entre otros. Estos enfoques han optimizado el diagnóstico, reduciendo la necesidad de métodos invasivos y aumentando la precisión en la identificación de lesiones malignas.

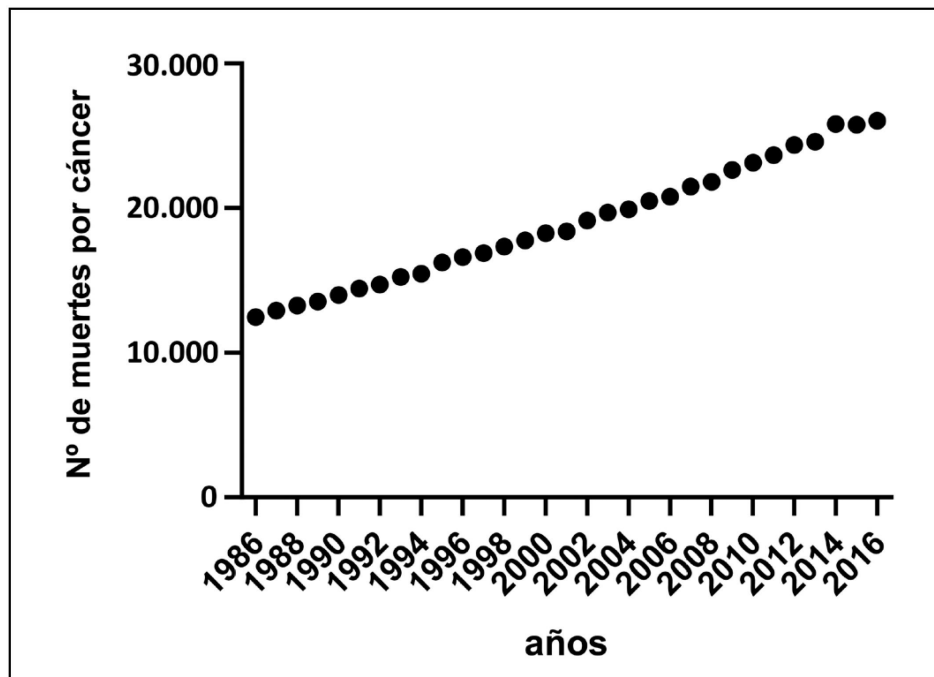
1.4. PROBLEMA DETECTADO

De acuerdo al trabajo realizado por Antonio D. *et al.* (2019), se califican los métodos descritos como bastante prometedores, sin embargo, no son un reemplazo a la biopsias de tejido y estudios histopatológicos (mencionados como “el estándar de oro” para la detección de cáncer). Estas opciones “ganadoras” a su vez, se encuentran limitadas por la morbilidad, los requisitos de tiempo y recursos, y el riesgo de sesgo de muestreo, donde a pesar de la implementación de complementos de diagnóstico, los actuales están restringidos por una precisión de diagnóstico deficiente, tal como menciona Yang E *et al.* (2018).

También se señala que a nivel país, la mortalidad por cáncer ha mostrado una notoria alza, evidenciándose que entre 1986 y 2016, las muertes por patologías oncológicas aumentaron un 109%.

Figura 1

Número de muertes por cáncer en Chile (período año 1986 y 2016).



Nota. Adaptado de *Cáncer en Chile y en el mundo: una mirada actual y su futuro escenario epidemiológico*, Fuente: Martínez-Sanguinetti et al. (2020)

Tomando en cuenta el alza descrita anteriormente, el autor menciona una proyección realizada por *Cancer Research UK* y AIRC (Agencia Internacional para la Investigación en Cáncer) que estima que para el año 2040, se diagnosticarán más de 29.5 millones de nuevos casos de cáncer en el mundo, siendo los cánceres con mayor aumento a nivel mundial: tiroides, hígado, oral y riñón; mientras que las mayores incidencias las presentarán el cáncer de mama en las mujeres y el cáncer de próstata en los hombres.

Por otro lado, estos institutos estiman que en Chile para el año 2040 habrá 94,807 nuevos casos de cáncer, lo que corresponde a un aumento del 77.6% en los próximos 21 años. En la actualidad, el cáncer ocasiona 28,443 muertes en el país, las que llegarán a 55,698 para el año 2040 según las estimaciones de AIRC.

El cáncer oral es una enfermedad cuya magnitud en Chile ha sido poco documentada. Santelices et al. (2016) destacan que “actualmente, existen muy pocas publicaciones que especifiquen la magnitud de esta enfermedad en Chile”. Esta afirmación se ve respaldada por una revisión de 27 artículos, de los cuales 16 no contenían datos relevantes sobre la estadística nacional. En este contexto, Baeza (2021) aporta información más detallada, señalando que el cáncer oral (incluyendo labio, cavidad oral y orofaringe) representa el 2.1% de la incidencia global y el 1.6% en Chile. Sin embargo, su prevalencia ha ido en aumento en la población joven debido al consumo de tabaco, alcohol y la presencia del virus del papiloma humano.

Baeza (2021) también señala que, según estudios del Instituto Nacional del Cáncer, la tasa de supervivencia a cinco años para el cáncer oral es del 57 %, reduciéndose al 46 % a los diez años. Estos datos se ven complementados por el estudio de Maraboli-Contreras *et al.* (2018), quienes analizaron la evolución de 120 pacientes y encontraron que solo el 33.9 % sobrevivió más de cinco años, evidenciando la gravedad del diagnóstico tardío.

El desconocimiento sobre el cáncer oral no solo afecta a la población general, sino también a los profesionales de la salud. Un estudio de Stillfried *et al.* (2016), realizado con 205 cirujanos dentistas de la Región Metropolitana, evidenció una baja frecuencia en la realización de exámenes diagnósticos clínicos para la detección de lesiones precancerizables y cáncer oral. Solo el 35.6 % de los participantes indicó practicar habitualmente este tipo de evaluación, mientras que el 64.4 % reconoció no realizarla de manera rutinaria, señalando diversas razones para ello.

En cuanto a la palpación de linfonodos cervicales⁴, una herramienta clave en la detección temprana, únicamente el 41.5 % de los encuestados afirmó realizarla de manera regular. Entre aquellos que sí llevaban a cabo un examen clínico diagnóstico para la identificación de lesiones orales, solo el 57.5 % complementaba la evaluación con palpación ganglionar.

Además, el estudio reveló una preocupante falta de conocimiento entre los pacientes: solo el 30.7 % de los profesionales afirmó que sus pacientes estaban informados sobre los factores de riesgo del cáncer oral, y apenas el 9.8 % indicó que estos reconocían sus signos y síntomas. Respecto a la preparación de los cirujanos dentistas, el 43.9 % manifestó sentirse capacitado para realizar un examen clínico diagnóstico, mientras que el 60.5 % declaró estar apto para la palpación de linfonodos cervicales.

Estos antecedentes ponen en evidencia la urgencia de desarrollar métodos de diagnóstico más efectivos, seguros y accesibles para la detección temprana del cáncer oral. La creciente morbilidad y mortalidad de esta enfermedad en Chile, junto con las deficiencias en su reconocimiento y diagnóstico, han sobrepasado la capacidad de respuesta del sistema de salud pública, subrayando la necesidad de soluciones innovadoras que optimicen su identificación y tratamiento oportuno.

⁴Son conocidos como los ganglios del cuello, estos filtran y transportan la linfa proveniente de los ganglios linfáticos adyacentes y los órganos de regreso hacia la circulación sanguínea (Navarro y Guzmán, 2024).

1.5. OBJETIVOS

El Objetivo General es diseñar e implementar un ensamblaje de modelos de aprendizaje automático (ML) que permita identificar y clasificar imágenes de lesiones orales, con el fin de contribuir al diagnóstico temprano de patologías oncológicas orales, especialmente cáncer bucal, mejorando así la capacidad de detección precoz y apoyando el proceso diagnóstico en el ámbito médico.

Los Objetivos Específicos son:

1. Estudiar el estado del arte en técnicas de aprendizaje automático aplicadas a la detección de cáncer bucal y otras patologías orales para comprender los avances realizados en este contexto.
2. Identificar aspectos clave del problema, como las señales indicadoras de cáncer en lesiones orales y los elementos que dificultan su correcta clasificación.
3. Recopilar y procesar datos relevantes (*datasets* de imágenes) que permitan entrenar correctamente un ensamble de redes neuronales convolucionales.
4. Diseñar y desarrollar un ensamble de redes neuronales convolucionales que pueda ser entrenada con la información recopilada.
5. Evaluar y comparar el rendimiento de los modelos y ensamble de estos con otros enfoques existentes.

1.6. CONTRIBUCIÓN

Este trabajo busca dejar un precedente de la implementación de un modelo de aprendizaje automático que permita realizar un aporte al proceso de detección precoz del cáncer bucal. De esta manera, se busca sentar las bases del desarrollo de métodos y tecnologías que puedan influir positivamente en el diagnóstico y pronóstico de pacientes con cáncer, puesto que no existen estudios robustos en Chile que abarquen este tipo de problemas a través de la implementación de la IA; Santelices *et al.* (2016) expresan de manera contundente que “la cantidad de artículos referentes al cáncer oral en Chile es baja. Además, muy pocos son específicos al cáncer oral”, agregan que actualmente el Programa Nacional del Cáncer no abarca específicamente el cáncer oral y tampoco es parte de las Garantías Explícitas en Salud (MINSAL), por lo que el acceso al diagnóstico es bastante limitado, y por ende, dificulta el tratamiento oportuno del paciente.

CAPÍTULO 2

MARCO CONCEPTUAL

2.1. INTELIGENCIA ARTIFICIAL

La Inteligencia Artificial o simplemente referida como IA es un campo de la informática que busca desarrollar sistemas capaces de realizar tareas que normalmente requieren inteligencia humana, como el reconocimiento de patrones, la toma de decisiones y el aprendizaje a partir de datos. Estas tareas se abordan a través del uso de algoritmos y modelos matemáticos para imitar ciertos aspectos del razonamiento y la percepción humana, permitiendo que las máquinas ejecuten acciones de manera autónoma o asistan a los seres humanos en la resolución de problemas complejos Ghosh y Arunachalam (2021).

Algunos autores la definen como el intento de replicar el pensamiento humano en máquinas, mientras que otros se centran en su capacidad de actuar de manera racional en función de la información que poseen. Según Russell y Norvig (2010), las principales corrientes de la IA se pueden agrupar en cuatro enfoques:

- Pensar humanamente: Intentar modelar los procesos de pensamiento humano utilizando métodos computacionales.
- Pensar racionalmente: Aplicar principios lógicos y matemáticos para imitar el razonamiento humano.
- Actuar humanamente: Construir sistemas capaces de imitar el comportamiento inteligente de las personas.
- Actuar racionalmente: Diseñar agentes que tomen decisiones óptimas basadas en la información disponible.

Hoy en día, la IA abarca múltiples ramas, como el aprendizaje automático (*Machine Learning*), las redes neuronales artificiales, la visión por computadora (*Computer Vision*), entre otros. Su impacto se ha extendido a diversas áreas, incluyendo la medicina, la industria, el entretenimiento y la seguridad, convirtiéndola en un campo fundamental en la ciencia y la tecnología moderna.

2.2. APRENDIZAJE AUTOMÁTICO

El aprendizaje automático o conocido también como *Machine Learning* (ML) es un área de la IA que consiste en el uso de modelos matemáticos de datos con el fin de que las computadoras aprendan por sí mismas, sin que estén programadas para ello directamente. De esta manera, se adaptan a la naturaleza de los problemas siendo capaz de mejorar su rendimiento con la experiencia. En la actualidad, pueden diferenciarse 3 paradigmas comunes de aprendizaje en este tópico:

- **Aprendizaje Supervisado:** a través de datos etiquetados, el modelo aprende respecto a la estructura implícita que existe en la clasificación de las imágenes y así puede aumentar su capacidad de realizar una predicción apropiada al comportamiento de nuevos datos.
- **Aprendizaje No supervisado:** mediante datos no etiquetados, de esta manera, es capaz de reconocer y aprender patrones y relaciones sin necesidad de conocer anteriormente la estructura de las variables de salida⁵.
- **Aprendizaje de Refuerzo:** se utiliza un agente que interactúa con su entorno y es recompensado o castigado de acuerdo a sus acciones, es así, que éste busca aprender a tomar decisiones para maximizar la recompensa obtenida a lo largo del tiempo.

En este contexto es necesario definir ciertos conceptos que permitirán una mejor comprensión de los trabajos previos en este campo y la solución planteada en esta memoria:

- **Datos de Entrenamiento:** refiérase a la información que se utiliza para el proceso de entrenamiento del modelo. Éstos pueden ser estructurados (con atributos o características) o no estructurados (imágenes, audios, vídeos, etc.)
- **Modelo:** representación matemática de los objetos y sus relaciones entre sí. También se refiere a la expresión de un algoritmo, cuyo fin es realizar predicciones, clasificaciones o tomar decisiones considerando ciertos datos de entrada.
- **Entrenamiento:** proceso a través del cual se proporciona un conjunto de datos al modelo para que éste, en base a dichos datos, aprenda a reconocer patrones e intente predecir ciertos resultados.
- **Validación:** evaluación del modelo en base a un conjunto de datos independiente del conjunto de datos de entrenamiento. De esta manera, es posible cuantificar la capacidad de generalización del modelo y que tan precisos son las predicciones realizadas.
- **Predicción:** estimación del modelo en base a un conjunto de datos no analizado anteriormente. La calidad de estas predicciones dependerá de factores como la calidad de los datos de entrenamiento, la adecuación del modelo al problema y la capacidad de reconocer patrones y generalizar.
- **Generalización:** la capacidad del modelo para realizar predicciones precisas en datos que no haya analizado anteriormente. Esto considera que este modelo haya sido entrenado previamente, de tal manera que éste haya aprendido de los datos de entrenamiento y sea capaz de captar atributos y las relaciones entre ellos. Una buena generalización se caracteriza por evitar el *Overfitting* y el *Underfitting*.
- **Overfitting:** Fenómeno de error de modelado donde el modelo aprende “demasiado” de los datos del conjunto de entrenamiento, memorizándose los detalles y el ruido específicos de este conjunto, mermando su capacidad de generalizar de manera precisa.

⁵Variables de salida. También se le puede llamar como *Target* (objetivo en inglés).

- **Underfitting:** Fenómeno de error de modelado donde éste es incapaz de captar adecuadamente las relaciones y patrones en los datos de entrenamiento, y por consiguiente, realizar predicciones precisas. Si se presenta underfitting, es debido a que el modelo es demasiado simple o tiene una capacidad insuficiente para aprender las relaciones complejas presentes entre los datos.

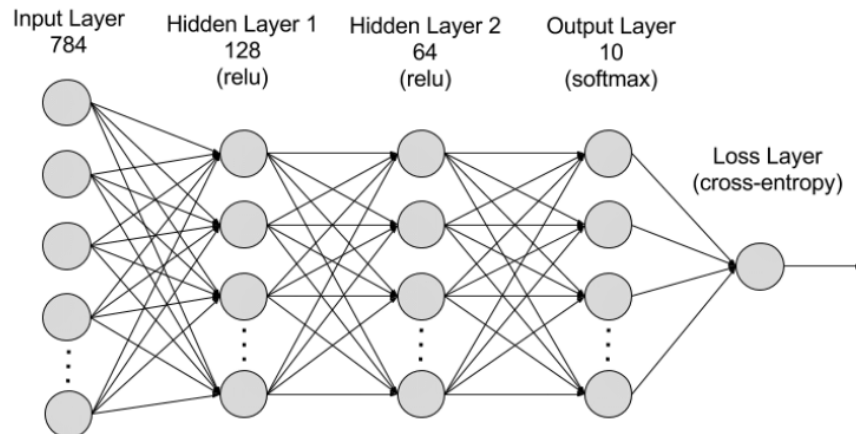
2.3. REDES NEURONALES ARTIFICIALES

Se definen como un subconjunto de conocimiento del ML que utiliza los nodos o las neuronas interconectadas en una estructura de capas inspirada en el funcionamiento del cerebro humano. Estas neuronas son identificadas como las unidades básicas de la red y se organizan en capas, ya sea en una de entrada, salida o escondida; a su vez, esta unidad posee asociada una función de activación, que determina la salida o activación de la neurona dada cierta entrada de datos, y un peso, que cuantifica la relevancia de las conexiones entre las neuronas. El entrenamiento de este tipo de estructuras considera la entrega de datos de entrada con los datos de salida deseados. Posteriormente, de manera iterativa, se van ajustando los pesos con el fin de mejorar la capacidad de generalización de la red.

Para facilitar la visualización y comprensión de una red neuronal genérica, se presenta la siguiente figura:

Figura 2

Ejemplo de Red Neuronal.



Nota. Adaptado de ¿Qué es una red neuronal?, Fuente: AWS (2024)

2.3.1. CONCEPTOS A CONSIDERAR

En este contexto, es necesario dejar en claro ciertos conceptos y procesos a considerar para sentar las bases de la comprensión de una red neuronal y su funcionamiento.

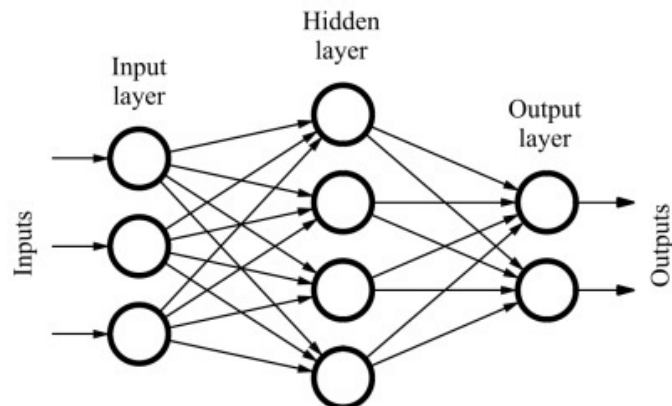
■ Arquitecturas

Algunas de las estructuras de redes neuronales más comunes son:

- **Feedforward Neural Networks:** se considera como la forma más simple de red neuronal, donde la información es procesada en una sola dirección. Esto ocurre de manera que la data puede pasar por múltiples capas ocultas, no obstante, siempre en una sola dirección y sin enviar feedback a las capas anteriores.

Figura 3

Ejemplo de Red Neuronal FeedForward

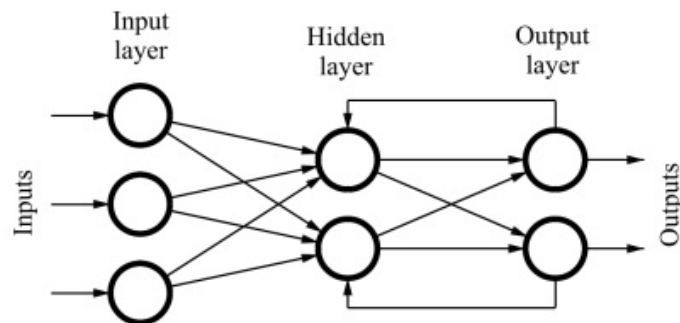


Nota. Adaptado de *Computational Methods and Optimization*, Fuente: Quiza y Davim (2011)

- **Recurrent Neural Networks (RNN):** este tipo de arquitectura se distingue por la incorporación de cierto uso de una “memoria” en los nodos, puesto que considera aspectos de la información de entrada anterior para influir sobre la data entrante actual y su correspondiente *output*.

Figura 4

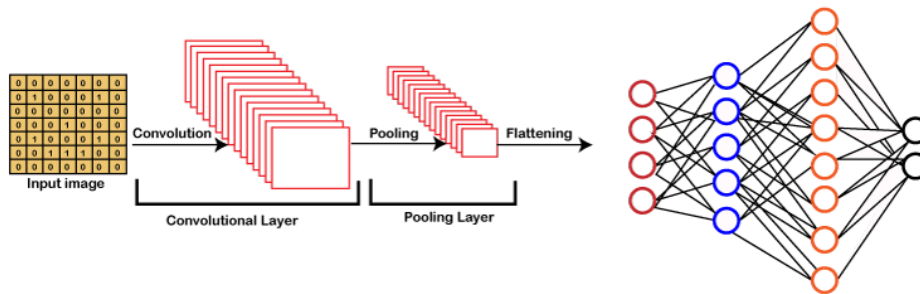
Ejemplo de Red Neuronal Recurrente



Nota. Adaptado de *Computational Methods and Optimization*, Fuente: Quiza y Davim (2011)

- **Convolutional Neural Networks:** este tipo de red neuronal emplea capas especiales denominadas convolucionales, las cuales permiten extraer información relevante y detectar patrones en los datos de entrada. Además, utiliza capas de *pooling* (agrupación) para reducir la dimensionalidad del problema, optimizando así el procesamiento de la información. Estas arquitecturas son ampliamente utilizadas en tareas como el reconocimiento de objetos, la segmentación y la clasificación de imágenes.

Figura 5
Ejemplo de Red Neuronal Convolucional



Nota. Adaptado de *Deep Learning for Cerebral Hemorrhage Detection and Classification in Head CT Scans Using CNN*, Fuente: Mahjoubi et al. (2023)

■ Entrenamiento

Para realizar el entrenamiento de estas redes, se utilizan diversos métodos, cada uno con sus beneficios y desventajas, con lo que debe considerarse el contexto y rendimiento de su aplicación en el problema. Algunas técnicas utilizadas para este proceso son: Descenso de gradiente (*Gradient Descent*), Retropropagación del error (*Backpropagation*), entre otros.

■ Regularización

Las técnicas de regularización son utilizadas como una manera de evitar el *Overfitting*, esto lo consigue a través de la penalización de nodos con pesos muy grandes o evitando que la complejidad del modelo sea muy alta. Para ello, se utilizan métodos como L1 (Lasso), L2 (Ridge), reducción de dimensionalidad, entre otros.

■ Funciones de Activación

Estas permiten determinar la salida de la neurona y el manejo de relaciones no lineales, de tal manera que sea posible captar relaciones complejas y contemplar el cálculo de ecuaciones no lineales. Generalmente se utilizan funciones como:

• Función Sigmoide

Consiste en una función logística que tiene una forma de "S" que, en términos más simples, toma una entrada real y la comprime en un rango entre 0 y 1.

$$\sigma(x) = \frac{1}{1 + e^{-x}}. \quad (1)$$

• Función Tangente Hiperbólica

Es similar a la función sigmoide (posee una figura sinuosa), pero tiene un rango ajustado entre -1 y 1. Generalmente usada en problemas de clasificación y regresión.

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}, \quad (2)$$

esta puede ser centrada para tener un rango ajustado entre 0 y 1 (para poder ser utilizada como función de predicción) de la siguiente manera:

$$\tanh^*(x) = \frac{\tanh(x) + 1}{2}. \quad (3)$$

• Función ReLU

Se define como una función no lineal que retorna cero si la entrada es negativa y retorna el valor de la misma función, si es positiva. Es usada en contextos donde existan problemas de gradiente en algunas capas de la red.

$$\text{ReLU}(x) = \max(0, x). \quad (4)$$

• SoftMax

Esta función convierte las salidas de las neuronas en una distribución de probabilidad (rango entre 0 y 1), de esta manera, se puede indicar las probabilidades de las distintas predicciones que realice la red neuronal.

$$\text{SoftMax}(x_i) = \frac{e^{x_i}}{\sum_j e^{x_j}}. \quad (5)$$

• Swish

Función no lineal definida como $f(x) = x \sigma(x)$, donde $\sigma(x)$ es la función sigmoide. A diferencia de ReLU, Swish es suave y no monótona, lo que mejora la propagación del gradiente y el rendimiento en redes profundas.

$$\text{Swish}(x) = x \cdot \sigma(x) = x \cdot \frac{1}{1 + e^{-x}}. \quad (6)$$

2.3.2. REDES NEURONALES CONVOLUCIONALES

Las redes neuronales convolucionales se definen como un subconjunto de conocimiento del ML, que consisten en capas de neuronas que pueden ser de entrada, ocultas o de salida; además, cada neurona posee asociada a ella, un peso y umbral. De esta manera, si una neurona es activada, esta envía los datos a la siguiente capa.

Algo a destacar de este tipo de redes neuronales, es que poseen un mejor rendimiento al percibir entradas de audio o de imagen.

Dicho esto, es posible identificar 3 tipos principales de capas en esta estructura de red neuronal:

- **Capa Convolutiva (*Convolutional layer*):** Es considerada la unidad central en las CNN (*Convolutional Neural Network*), ya que realiza la mayoría de los cálculos.
- **Capa de Agrupación (*Pooling*):** Lleva a cabo una reducción de dimensionalidad, lo que reduce el número de parámetros en la entrada. Ayuda a reducir la complejidad, mejorar la eficiencia y limitar el riesgo de sobreajuste, pudiendo llevar a la pérdida de información por error.
- **Capa completamente conectada o densa (*Fully-connected layer*):** se le llama así puesto que cada nodo de la capa de salida se conecta directamente a un nodo de la capa anterior.

2.3.3. ENSAMBLE DE MODELOS

Los ensambles de modelos constituyen una técnica avanzada en *Machine Learning*, donde se combinan múltiples modelos para mejorar la precisión y estabilidad de las predicciones. En lugar de depender de un único modelo, el ensamble busca reducir el error y la variabilidad, aprovechando las fortalezas individuales de cada arquitectura.

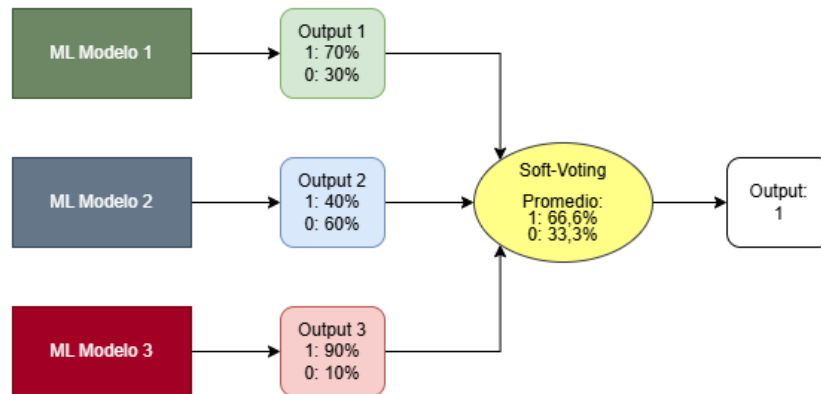
Este enfoque es ampliamente utilizado en tareas de clasificación y regresión, especialmente cuando se desea minimizar errores y lograr una mejor generalización.

Existen diversas estrategias de ensamblaje, entre las que destacan:

1. **Bagging (*Bootstrap Aggregating*):** Técnica que entrena varios modelos con subconjuntos diferentes de los datos y promedia sus predicciones para reducir la varianza. Un ejemplo popular es *Random Forest*.
2. **Boosting:** Método que ajusta modelos secuenciales, donde cada modelo corrige los errores del anterior. Esto permite mejorar el rendimiento de clasificadores débiles. Algunos algoritmos representativos incluyen *AdaBoost* y *XGBoost*.
3. **Stacking:** Estrategia en la que se combinan múltiples modelos base y sus predicciones se utilizan como entrada para un metamodelo, el cual genera la predicción final.
4. **Voting (*Hard Voting* y *Soft Voting*):** Método en el que varios modelos emiten una predicción y se selecciona la clase mayoritaria (*Hard Voting*) o se promedian las probabilidades de cada clase (*Soft Voting*).

Figura 6

Representación de Soft Voting en Machine Learning



Nota. Existe particular interés por su representación, puesto que es uno de los conceptos clave para la solución al problema planteado, Fuente: Autoría Propia.

En este trabajo, se implementó un ensamble basado en promediado de predicciones (*Soft Voting*), donde los modelos ResNet50, Xception e InceptionV3 aportan sus predicciones individuales y la decisión final se obtiene a partir de su combinación ponderada. Esta técnica permitió mejorar la capacidad de detección de cáncer oral, mitigando los sesgos individuales de cada modelo y optimizando la robustez del sistema.

2.3.4. MÉTRICAS PARA EVALUACIÓN DE MODELO

Para medir el rendimiento de nuestro modelo, tomaremos en consideración el trabajo realizado por Warin *et al.* (2021), quienes realizaron la tarea de clasificar y detectar lesiones cancerosas en cavidades orales con el mismo objetivo que la presente memoria. En su estudio emplearon los modelos DenseNet121 y Faster R-CNN, evaluando métricas como *precision*, *recall*, F1-score y AUC-ROC para medir el desempeño de los algoritmos en la clasificación y detección automatizada de dichas lesiones. No obstante, la cantidad de datos utilizada en su investigación fue bastante acotada, ya que contaron con un total de 700 imágenes divididas entre casos de carcinoma escamocelular oral y mucosa oral normal, lo que resalta los desafíos asociados a trabajar con conjuntos de datos limitados. Un factor llamativo en el trabajo mencionado fue la utilización de métricas de “Precision”, “Recall”, “F1-Score” y “AUC-ROC”, explicados como:

- **Matriz de Confusión:** esta herramienta organiza las predicciones del modelo, para de esta manera visualizar el desempeño del modelo permitiendo calcular diversas métricas, que en esta memoria, se seleccionó *Precision*, *Recall*, F1-Score y AUC-ROC.

Tabla 1

Matriz de confusión

	Positivo Real	Negativo Real
Positivo predecido	TP (Positivo Verdadero)	FP (Falso Positivo)
Negativo predecido	FN (Falso Negativo)	TN (Negativo Verdadero)

- **Precision:** medida de la exactitud del modelo en términos de predicciones correctas. Este se cuantifica de la siguiente manera:

$$Precision (P) = \frac{TP}{TP + FP} \cdot \quad (7)$$

- **Recall:** capacidad del modelo para lograr identificar todos los ejemplos positivos reales. Se cuantifica de la siguiente manera:

$$Recall (R) = \frac{TP}{TP + FN} \cdot \quad (8)$$

- **F1-Score:** evalúa el rendimiento de un modelo de clasificación combinando las métricas de *precision* y *recall* en una sola medida para obtener una evaluación más equilibrada. Definiéndose como la media armónica entre ambas, dada por la siguiente fórmula:

$$F1 - Score (F1) = 2 \cdot \frac{P \cdot R}{P + R}, \quad (9)$$

esta varía entre 0 y 1, siendo 1 el mejor rendimiento posible.

- **AUC-ROC:** Es una métrica que evalúa el rendimiento de un modelo de clasificación binaria al medir su capacidad para distinguir entre clases positivas y negativas. La métrica se basa en la curva ROC, que traza la relación entre la tasa de verdaderos positivos (TPR)⁶ y la tasa de falsos positivos (FPR)⁷ a diferentes umbrales de decisión. El valor del AUC varía entre 0 y 1, donde:
 - 1 representa un modelo perfecto que clasifica correctamente todas las instancias.
 - 0.5 indica un modelo que no tiene capacidad de clasificación (equivalente a adivinar al azar).
 - Valores más cercanos a 1 reflejan un mejor rendimiento del modelo.

2.4. HERRAMIENTAS DE SOPORTE DE DECISIONES CLÍNICAS

El apogeo del ML en los últimos años ha permitido deslumbrar el potencial que posee para apoyar las decisiones clínicas y ayudar en el proceso de realizar una atención óptima y oportuna al paciente.

Por ello es que se han generado diversas herramientas para servir de soporte en la toma de decisiones clínicas.

Por ejemplo, Suárez-Obando y Ordóñez Vásquez (2012) menciona los medios expertos de apoyo diagnóstico como:

Sistemas de interpretación de grandes volúmenes de datos, con la capacidad de “aprender” de resultados históricos, de la inclusión de nuevos datos y de la modificación de las reglas de inferencia. (pp. 199–208)

De manera general, estos sistemas inteligentes analizan las variables que se consideran en los procesos de construcción de un cuadro clínico y obtienen probabilidades para el diagnóstico y pronóstico del paciente.

2.4.1. CARACTERÍSTICAS COMUNES IDENTIFICADAS

Abstrayendo el concepto de un sistema de apoyo clínico, se tiene el sistema DSS (Sistema de apoyo a las decisiones), que incluye una variedad de herramientas, sistemas y tecnología (Lucena y Porras D., 2006); y es en este contexto que se identifican algunas características comunes que deben poseer:

1. Son sistemas computarizados, flexibles y adaptables según las situaciones.
2. Manejan problemas.

⁶Proporción de ejemplos positivos reales que el modelo clasifica correctamente como positivos. También se conoce como recall

⁷Proporción de ejemplos negativos reales que el modelo clasifica incorrectamente como positivos.

3. Soportan la solución de problemas.
4. Utilizan datos y modelos.
5. Asocian los recursos intelectuales del usuario.
6. Mejoran la calidad de las decisiones.
7. Emplean interfaces de usuario para su comunicación.
8. Poseen herramientas para la recuperación y procesamiento de datos, y análisis de la información.

Una vez identificadas las bases de un DSS, los autores ponen énfasis en que el propósito de estos sistemas es ampliar las capacidades de los tomadores de decisiones sin reemplazar su juicio, y a su vez relaja los alcances de cómo interactúa el sistema con el usuario.

Esto queda expuesto a través de la siguiente frase:

Un sistema de soporte a las decisiones se puede considerar como un programa de software que opere sobre un computador que pueda ayudar a los administradores en cualquier nivel organizacional, durante el proceso de toma de decisiones, facilitando la organización del conocimiento en temas semi estructurados y no estructurados. (Lucena y Porras D., 2006, p. 16)

Esta interpretación da entender que mientras el sistema pueda ofrecer un apoyo que permita tomar decisiones de mejor manera, éste ya satisface la característica 7, que significa que se puede dejar de lado el atributo 8 para iteraciones más avanzadas cuando el sistema tenga más cohesión con el área de trabajo, puesto que requiere de expertos que puedan asesorar y ayudar a que el sistema realice un análisis pertinente y lógico. En apoyo a esta interpretación se dan ejemplos como: sistemas de soporte ejecutivos, sistemas de información geográfica, inteligencia de negocios, sistemas de administración de conocimiento, redes neuronales artificiales, entre otros.

2.4.2. BENEFICIOS Y LIMITACIONES

Tomando en cuenta lo anteriormente expuesto respecto a los DSS, los autores Lucena y Porras D. (2006) determinan los beneficios y limitaciones que poseen estos sistemas, y que por consecuencia, poseerá el modelo de ML de apoyo para la detección temprana de cáncer bucal:

Fortalezas:

- Mejoran la calidad en la toma de decisiones.
- Mejoran la eficiencia del personal, que conlleva a un incremento en la productividad.

- Es posible analizar y procesar grandes volúmenes de información para la correcta toma de decisiones.
- Extienden las habilidades del tomador de decisiones para procesar información y conocimiento (Siendo esto, básicamente el propósito intrínseco de este tipo de sistemas).
- Disminuyen el tiempo asociado con la toma de una decisión (algo fundamental para el contexto en el que se sitúa el problema abordado).
- Se genera nueva evidencia para el soporte de una decisión (contribuyendo al estado del arte) o permite validar suposiciones previas.

Debilidades:

- No están diseñados aún para considerar talentos humanos para la toma de decisión tales como la creatividad, la imaginación o la intuición.
- El poder de un DSS está limitado por su diseño; por la capacidad del computador sobre el cual corre y el conocimiento que posee a la hora de utilizarlo.
- Algunos diseños resultan imperfectos gracias a la representación inadecuada de la realidad. Esto puede ser por no incorporar expertos en el tema o por incluir suposiciones incorrectas del trabajo.

2.4.3. CONSIDERACIONES DEL MODELO

Tomando en cuenta lo descrito anteriormente sobre los DSS, es necesario señalar que en esta memoria se desea plantear un modelo de ML que apoye la detección temprana de cáncer bucal. Y es que, los mismos autores (Lucena y Porras D., 2006) señalan este modelo como un DSS, que esta inmerso en el contexto de la informática médica, para apoyar el proceso de diagnóstico temprano, alertando si existen señales en la mucosa bucal que sean alarmantes y deban ser revisados por un profesional a través de una biopsia si así lo amerita. De este modo, el alcance de esta memoria se limitará a la investigación, búsqueda de datos y métodos que se consideren pertinentes para una solución, diseño y evaluación de un modelo de *Machine Learning*, dejando un precedente para fomentar la investigación y desarrollo de soluciones para combatir el cáncer en Chile.

2.5. TRABAJOS RELACIONADOS

En los últimos años, el uso de *Machine Learning* ha mostrado un gran potencial en la detección temprana del cáncer oral. Diversos estudios han aplicado modelos de redes neuronales convolucionales (CNNs) para analizar imágenes médicas y mejorar la precisión del diagnóstico. Los enfoques varían según la fuente de datos utilizada, los métodos de preprocesamiento aplicados y las arquitecturas implementadas.

Uno de los primeros trabajos destacados en este ámbito es el estudio realizado por Aubreville *et al.* (2017), quienes desarrollaron un sistema basado en LeNet-5⁸ para la clasificación de tejido canceroso en imágenes obtenidas mediante *Confocal Laser Endomicroscopy* (CLE)⁹. Su enfoque utilizó un método de clasificación basado en parches, donde las imágenes CLE fueron divididas en regiones de 80x80 píxeles, analizadas individualmente y combinadas posteriormente a través de una fusión probabilística para generar una predicción final. El dataset empleado consistió en 7894 imágenes CLE, logrando un rendimiento de 88,3 % de precisión, 86,6 % de sensibilidad y 90,0 % de especificidad. Sin embargo, los autores destacaron que la variabilidad en la calidad de las imágenes CLE y la necesidad de validación externa son desafíos para la implementación clínica.

Por otro lado, Shamim *et al.* (2019) exploraron un enfoque basado en redes preentrenadas como VGG19¹⁰ y ResNet50 para la clasificación de lesiones linguales en imágenes fotográficas. Su estudio abordó dos tareas: una clasificación binaria entre lesiones benignas y precancerosas, donde VGG19 obtuvo un *precision* del 98 %, y una clasificación multiclase de cinco tipos de lesiones linguales, donde ResNet50 alcanzó un 97 % de *precision*. El *dataset* empleado fue considerablemente más pequeño en comparación con otros trabajos, compuesto por 200 imágenes para clasificación binaria y 300 para clasificación multiclase, lo que representa una limitación importante en términos de generalización, esta restricción es respaldada por Shahinfar *et al.* (2020), quien enfatiza la necesidad de un mayor volumen de datos para mejorar la robustez del modelo, especialmente en problemas de clasificación multiclase. Además, los autores subrayan la importancia de complementar los modelos de IA con la evaluación de un especialista (“AI + Physician”), logrando una *precision* del 100 % cuando los errores del modelo eran corregidos mediante supervisión humana.

En un enfoque distinto, Koriakina *et al.* (2022) investigaron la aplicación de *Machine Learning* en muestras citológicas, comparando dos métodos: aprendizaje de instancias simples (SIL) y aprendizaje de múltiples instancias (ABMIL). Su *dataset* consistió en muestras de 24 pacientes y un *dataset* sintético llamado PAP-QMNIST¹¹, utilizado para mejorar la diversidad de los datos. Sus hallazgos mostraron que SIL con ResNet18 alcanzó un *precision* del 96,7 % a nivel celular, superando a ABMIL en términos de *precision* y consumo de memoria. No obstante, la escasez de datos clínicos y la dependencia de muestras sintéticas representaron limitaciones significativas para su aplicación en entornos clínicos reales. Esto coincide con las recomendaciones de Shahinfar *et al.* (2020), quienes enfatizan la importancia de priorizar la diversidad de imágenes sobre la mera acumulación de muestras similares provenientes de

⁸Fue una de las primeras arquitecturas de redes neuronales convolucionales (CNN), usada originalmente para la tarea de reconocimiento de dígitos escritos a mano en imágenes.

⁹Dicho en terminos simples, el CLE se basa en la obtención de un corte óptico de muestra celular o de tejido grueso utilizando marcadores fluorescentes, capaces de identificar receptores específicos en la célula. Posteriormente, se utiliza un rayo láser ultravioleta muy fino y paralelo que incide en un punto del tejido, provocando una fluorescencia que es captada por un filtro selectivo (Úrquiza, 2010).

¹⁰Es una CNN desarrollada por Simonyan y Zisserman (2015). Se compone de 19 capas, con una arquitectura profunda basada en múltiples capas convolucionales de 3x3, seguidas de capas de *pooling* y completamente conectadas. Es utilizada en clasificación de imágenes y otras tareas de visión por computadora debido a su capacidad para capturar características jerárquicas.

¹¹Desarrollado específicamente para imitar las características de los datos reales de muestras citológicas utilizados en la detección de cáncer oral.

un mismo entorno.

Estos estudios reflejan la diversidad de enfoques en la detección del cáncer oral mediante *Machine Learning*. Desde el uso de imágenes CLE (Aubreville *et al.*, 2017), hasta imágenes fotográficas de lesiones linguales (Shamim *et al.*, 2019) y muestras citológicas (Koriakina *et al.*, 2022), cada uno ha aportado avances significativos en la identificación de lesiones malignas. Sin embargo, persisten desafíos como la disponibilidad y variabilidad de datos de alta calidad, la necesidad de validación externa y la integración de estas soluciones en entornos clínicos.

CAPÍTULO 3

PROPUESTA DE SOLUCIÓN

3.1. CARACTERÍSTICAS RELEVANTES DEL PROBLEMA

3.1.1. ESPECIFICACIONES DEL PROBLEMA

Para lograr los objetivos planteados se utilizarán redes neuronales convolucionales (CNN) con un enfoque de ensamble, donde varios modelos contribuirán a la clasificación de las imágenes, decidiendo por votación si una lesión es maligna o benigna. Los datos utilizados serán imágenes públicas obtenidas de la plataforma RoboFlow, las cuales cumplen con normativas legales, médicas y académicas. Estas imágenes, que presentan diferentes escalas, han sido uniformemente ajustadas a una resolución de 224x224 píxeles para garantizar consistencia y reducir el costo computacional. El uso de modelos preentrenados ligeros permitirá realizar un *fine-tuning* eficiente y crear un sistema de ensamble confiable.

3.1.2. DESAFÍOS DEL PROYECTO

Entre los principales desafíos se encuentran la variedad de dimensionalidades y posibles desequilibrios en los datos. Para mitigar los efectos de este último, se seleccionaron *datasets* lo más equilibrados posibles entre las clases de interés, minimizando así sesgos hacia alguna de las etiquetas (maligno o benigno). Otro desafío importante es la naturaleza crítica del problema, ya que es fundamental detectar de manera efectiva los casos de cáncer. En este sentido, se prioriza la detección de casos positivos, es decir, es preferible incurrir en falsos positivos a tener falsos negativos que podrían impactar negativamente en la salud del paciente. Este enfoque asegura que se maximice la identificación de casos de cáncer, minimizando el riesgo de que un diagnóstico negativo incorrecto pase desapercibido.

3.1.3. RELEVANCIA DEL PROYECTO

Este trabajo tiene el potencial de generar un impacto significativo en los campos científico, tecnológico y médico. Su implementación podría contribuir al avance de la oncología, especialmente en la detección del cáncer oral, un área que aún no está suficientemente desarrollada en Chile. La escasez de datos nacionales sobre este tipo de cáncer es un problema relevante, y este proyecto busca ser un primer paso hacia la creación de herramientas tecnológicas confiables que apoyen el diagnóstico clínico en esta área de la medicina.

3.1.4. VARIABLES CLAVE Y PREPROCESAMIENTO

Se han reunido imágenes que representan tres estados: bocas sanas, bocas con heridas o patologías no cancerígenas, y bocas con patologías cancerígenas o precancerígenas. Estas clases permiten identificar cómo luce una cavidad oral en condiciones normales, frente a lesiones leves, y cuando existen signos que ameritan preocupación y la revisión de un experto. El preprocesamiento de los datos incluye la escala de imágenes, cambios en el número o nombre de las clases, entre otros aspectos, que serán detallados más adelante para cada *dataset*.

3.1.5. MODELOS SELECCIONADOS

Para este proyecto, se han seleccionado los modelos ResNet50, InceptionV3 y Xception, debido a su uso frecuente en investigaciones previas en este campo. La elección de estos modelos permitirá compararlos con el estado del arte y evaluar su rendimiento dentro del enfoque de ensamble propuesto. Las principales limitaciones del estudio están relacionadas con los recursos computacionales, lo que motivó la decisión de emplear Google Colab Pro para aprovechar las capacidades avanzadas que ofrece, optimizando el tiempo y costo de los experimentos.

3.1.6. EVALUACIÓN DEL RENDIMIENTO

Se ha seleccionado las métricas de *precision*, F1-score, *recall* y AUC-ROC, poniendo un énfasis especial en *recall* y F1-score. El *recall* es crucial porque el principal objetivo es asegurar que todos los casos de cáncer sean detectados. Es preferible cometer falsos positivos que arriesgarse a no identificar un caso real de cáncer, lo que podría tener consecuencias graves para el paciente. Al mismo tiempo, el F1-score permite mantener un equilibrio entre precisión y *recall*, al evaluar la relación armónica entre ambas. Esto es importante, ya que aunque es prioritario detectar todos los casos de cáncer, también buscamos evitar un exceso de falsos positivos.

El F1-score proporciona una visión más equilibrada del rendimiento del modelo, especialmente en escenarios con cierto desequilibrio entre las clases (cáncer y no cáncer), lo que lo convierte en una métrica clave para asegurar que el modelo mantenga un desempeño robusto. Al combinar el *recall* con el F1-score, el modelo no solo será efectivo en la detección de cáncer, sino que también mantendrá un buen equilibrio en las predicciones, evitando sesgos hacia una sola clase.

Finalmente, el uso del AUC-ROC permite medir la capacidad del modelo para discriminar entre las clases positivas (cáncer) y negativas (no cáncer) en diferentes umbrales de decisión, asegurando que el modelo pueda diferenciar correctamente los casos de cáncer de los que no lo son en una amplia variedad de escenarios.

3.2. CONSIDERACIONES ÉTICAS Y DE PRIVACIDAD

Después de haber abordado ciertos aspectos técnicos iniciales de la solución propuesta, es esencial, tanto desde un punto de vista moral como legal, considerar los aspectos legislativos relacionados con la obtención, procesamiento y tratamiento de los datos utilizados en esta memoria. Esto es crucial dado el contexto médico de la investigación, donde la confidencialidad y el cumplimiento de las normativas vigentes se convierten en responsabilidades ineludibles dentro del rol de estudiante e investigador.

Esta reciente política nacional establece:

El desarrollo de la IA debe contribuir al bienestar integral de todas las personas. La estrecha relación de esta tecnología con las personas, que cada vez tiene un impacto más grande en nuestras vidas, requiere que dichos sistemas se desarrollen de manera ética e inclusiva. (Gobierno de Chile, 2023)

Esto demarcando que el fin último de sistemas de este tipo, es el beneficio de todas las personas, tomando en cuenta prácticas que no infrinjan de manera alguna los principios éticos establecidos por la sociedad y sean inclusivos con todos los segmentos de esta.

Además, se establece que:

Los sistemas de IA se desarrollarán y se utilizarán de conformidad con las normas vigentes en materia de privacidad y protección de datos, y sólo tratarán datos que cumplan con la normativa en términos de calidad e integridad. (Gobierno de Chile, 2023)

Esto pone especial énfasis en el cómo se han obtenido los datos: deben provenir de una fuente de confianza y producidos de una manera ética, legal y privada, respetando las condiciones del contexto del que estas constituyen; por ejemplo, en el presente trabajo, es fundamental que la data obtenida del paciente, considere el respeto a la privacidad del paciente, posea una buena calidad, y lo más importante, sea con el conocimiento y consentimiento del paciente.

Esto se relaciona con lo siguiente:

Aunque la existencia de grandes volúmenes de datos no es garantía de buenos algoritmos de IA, es necesario para el entrenamiento adecuado de los algoritmos. Además, se requiere que los datos representen adecuadamente las realidades (sean de buena calidad), sin sesgos nocivos, interoperables, reutilizables, correctamente anonimizados y agregados (si corresponde), no discriminatorios, entre otras características. (Gobierno de Chile, 2023)

Donde se menciona la misma importancia de la ética en proyectos de este tipo, pero a su vez, introduce un nuevo factor, la cual indica que a pesar de que existan grandes volúmenes de datos, estos no implican que el modelo aumente su calidad, aunque sí establece la necesidad de estos datos para el entrenamiento de modelos.

Esto se complementa con el siguiente párrafo:

Para que exista un despliegue efectivo de IA en Chile es necesario un ecosistema donde existan repositorios de datos abiertos, diversos y de alta calidad, pero

que también sean seguros y resguarden los derechos de privacidad y protección de los datos de las personas. (Gobierno de Chile, 2023)

A pesar de que se establece la existencia de un repositorio abierto de datos, lo específico y transversalmente complejo del tema médico conlleva a que sea difícil que estos estén disponibles preliminarmente, considerando que solo recientemente se ha vuelto una práctica presente para los doctores obtener imágenes de la manifestación de las patologías físicas de pacientes.

Los datos utilizados en esta investigación cumplen con las siguientes indicaciones:

- **Cumplimiento de la política nacional de IA:** esta memoria se compromete a adherirse rigurosamente a los principios éticos y legales establecidos en la Política Nacional de Inteligencia Artificial de Chile (Gobierno de Chile, 2023). Dichos principios promueven una IA centrada en el bienestar de las personas y en el respeto a sus derechos fundamentales. Para garantizar la dignidad y privacidad de los pacientes cuyos datos se utilizaron, se implementaron estrictas medidas de anonimización y protección de datos. Además, se prioriza la transparencia en cada fase del proceso, desde la recolección y procesamiento de los datos hasta la creación y validación de los modelos de IA. Esto implica documentar y comunicar claramente los métodos utilizados, así como los resultados obtenidos, asegurando que todas las decisiones y algoritmos puedan ser auditados y replicados por otros investigadores.
- **Responsabilidad ética y sin fines de lucro:** se asume la responsabilidad de que esta investigación es realizada con fines no lucrativos, enfocada exclusivamente en contribuir al avance de la ciencia, la mejora de la educación universitaria, el desarrollo del país y el fortalecimiento del ámbito médico en Chile. En concordancia con la Política Nacional de IA, se garantiza que los resultados de esta memoria son utilizados de manera ética y beneficiosa para la sociedad, promoviendo un desarrollo tecnológico que respete y potencie los derechos y el bienestar de todas las personas involucradas.
- **Protección de datos personales y Propiedad Intelectual:** se cumplió estrictamente con las normativas chilenas sobre protección de datos personales, asegurando la confidencialidad y seguridad de la información médica. Los datos que pudieron incluir información personal fueron anonimizados y manejados bajo los más altos estándares de confidencialidad y seguridad. Adicionalmente, se revisaron las condiciones de uso de todos los *datasets* utilizados, garantizando que todos ellos fueran citados adecuadamente, respetando las leyes de propiedad intelectual vigentes en Chile.
- **Evitar sesgos y discriminación:** se llevaron a cabo evaluaciones para asegurar que los análisis realizados no perpetuaran sesgos que pudieran generar resultados discriminatorios o excluyentes. Se implementaron medidas para mitigar cualquier sesgo identificado, asegurando que los análisis fueran justos y equitativos.
- **Responsabilidad en el tratamiento de datos médicos:** los datos médicos fueron manejados con el más alto nivel de confidencialidad y ética, en estricta conformidad con las leyes y regulaciones aplicables, garantizando así la protección de la privacidad y la integridad de la información de los pacientes.

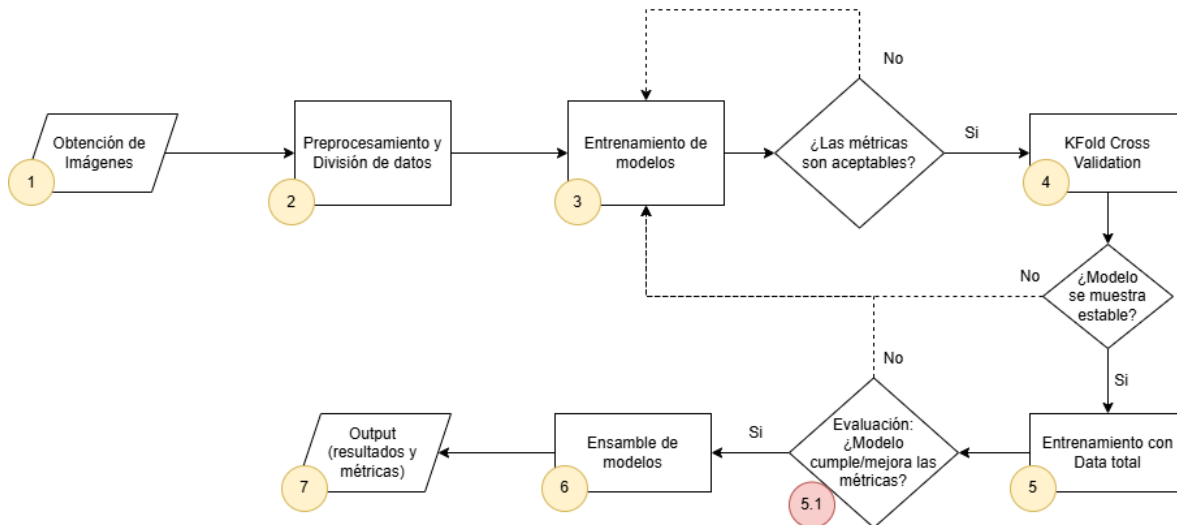
3.3. DESCRIPCIÓN GENERAL DE LA SOLUCIÓN PROPUESTA

Para la clasificación, se utilizan modelos ResNet50, InceptionV3 y Xception, cada uno pre-entrenado en el conjunto de datos ImageNet¹². Estos modelos fueron seleccionados por su eficiencia en términos de rendimiento y recursos computacionales, así como por su capacidad para generalizar en tareas de clasificación de imágenes médicas.

El flujo de trabajo general del ensamble de modelos de aprendizaje automático propuesto se puede resumir en los siguientes pasos:

Figura 7

Diagrama general del proceso



Nota. Un diagrama para visualizar el flujo que se siguió para la obtención de imágenes, entrenamiento de modelos, el ensamblado y la posterior evaluación, Fuente: Autoría propia.

1. **Obtención de imágenes:** acumulación de imágenes de diversos datasets, gran parte de estos ya poseían transformaciones y aumentos de datos realizados por sus correspondientes autores.
2. **Preprocesamiento de los datos:** aplicación de transformaciones para estandarización de los datos y unificación de los datasets.
3. **Entrenamiento:** se utiliza el conjunto de entrenamiento (*train*) para entrenar los modelos y evaluar su desempeño tanto en los conjuntos de entrenamiento como en validación (*train* y *valid*). Esta evaluación preliminar permite detectar posibles deficiencias en el modelo antes de aplicar *K-Fold Cross-Validation*, evitando así un proceso de validación innecesario en modelos que aún requieren ajustes.

¹²ImageNet es una base de datos de imágenes que contiene millones de imágenes etiquetadas y es utilizada ampliamente para entrenar y evaluar modelos de visión por computadora.

4. **K-fold Cross Validation:** el conjunto de datos de entrenamiento (*train* y *valid*) se divide en K partes, utilizando K-1 para entrenamiento y la restante para validación en cada iteración. Esto mejora la evaluación del modelo y ayuda a reducir el sobreajuste.
5. **Entrenamiento con Data Total:** usando los mejores hiperparámetros obtenidos, se utilizan todas las imágenes del conjunto *train* y *valid* para entrenar con el conjunto total de data de entrenamiento.
- 5.1 **Evaluación:** se obtienen las métricas de cada modelo en relación al conjunto test. Si estas no mejoran, se ajustan los hiperparámetros y se reinicia el proceso de entrenamiento hasta obtener un modelo más preciso y generalizable.
6. **Ensamblaje de modelos:** promediar las predicciones de los modelos para obtener una predicción final más robusta.
7. **Resultados y métricas:** se generan las métricas y las predicciones finales con las etiquetas correspondientes para facilitar la interpretación.

3.4. ALGORITMOS Y TÉCNICAS DE ENTRENAMIENTO Y EVALUACIÓN DEL MODELO

3.4.1. XCEPTION

Es una red neuronal convolucional basada en el concepto de extrema separabilidad (*Extreme Inception*). Propuesta por François Chollet en 2017, que fue quien desarrolló esta arquitectura como una extensión del modelo Inception, reemplazando las convoluciones tradicionales por convoluciones separables en profundidad (*Depthwise Separable Convolutions*), que propone mejorar significativamente la eficiencia computacional sin sacrificar rendimiento.

El modelo Xception aprovecha el principio de que la separación espacial y de canales en una convolución estándar puede mejorar la capacidad de aprendizaje de las redes neuronales convolucionales (Chollet, 2017). Este enfoque permite un procesamiento más eficiente y una mayor flexibilidad en la extracción de características.

Entre los aspectos más relevantes de Xception destacan:

1. Depthwise Separable Convolutions (Convoluciones Separables en Profundidad)

- a) **Depthwise Convolution:** realiza una convolución espacial independiente para cada canal, reduciendo el costo computacional.
- b) **Pointwise Convolution:** aplica una convolución 1×1 para combinar las características entre los canales. Este enfoque reduce drásticamente el número de parámetros del modelo, permitiendo redes más ligeras y rápidas.

2. Bloques de Convoluciones Repetitivas

Xception organiza su arquitectura en bloques repetitivos compuestos por capas convolucionales separables en profundidad y conexiones residuales. Estas conexiones permiten que la información fluya sin problemas entre las capas, ayudando a entrenar redes más profundas de manera estable.

3. Uso de Global Average Pooling (GAP)

Al final de la arquitectura, Xception utiliza una capa de *Global Average Pooling* en lugar de capas completamente conectadas (*Fully Connected*), lo que reduce aún más los parámetros y mejora la generalización. Esto también ayuda a disminuir el riesgo de sobreajuste.

Xception fue seleccionada como arquitectura base debido a su capacidad para procesar imágenes clínicas RGB de diversas superficies de la cavidad oral, como lengua, mejillas, paladar y labios. La eficiencia computacional de sus convoluciones separables permite manejar un *dataset* amplio (9507 imágenes) sin comprometer el rendimiento. En conjunto a la atención multi-cabeza (*Multi-Head Attention*) y capas densas, se incrementa la capacidad de detección de lesiones malignas.

3.4.2. INCEPTIONV3

Es una arquitectura de red desarrollada por Google, cuyo fin es mejorar la eficiencia y aumentar la precisión en la clasificación de imágenes a través de la combinación de múltiples tamaños de filtros en una sola capa, de esta manera se pueden captar *features* en distintas escalas.

Esto conlleva a una reducción importante en las dimensiones de los vectores de parámetros, disminuyendo la complejidad, aumentando el rendimiento sin perder precisión.

Entre los principales métodos utilizados por InceptionV3 (Szegedy *et al.*, 2015), se encuentran:

1. **Módulos Inception:** se aplican simultáneamente filtros de distintos tamaños en la misma capa, de esta manera, se capturan características a distintas escalas.
2. **Convoluciones Factorizadas:** se optimizan las convoluciones grandes dividiendo en secuencias más pequeñas. Esto reduce en gran medida el número de parámetros y el costo computacional, a la vez que se mantiene la capacidad de representación de la red. También puede ocurrir que esta división se realice para generar convoluciones asimétricas, que reduce aún más el costo.
3. **Clasificadores Auxiliares:** para mejorar la convergencia, se utilizan clasificadores auxiliares que contienen gradientes que mitigan el desvanecimiento del gradiente.
4. **Reducción de Parámetros:** se evitan cuellos de botella respecto a la capacidad de representación del modelo a través de la reducción de tamaño del conjunto de datos.

3.4.3. RESNET

Es una red neuronal profunda que introduce el concepto de “Skip connections”, que permite que una capa salte una o más capas y se añada a la salida de una capa posterior. Esto permite mitigar el desvanecimiento de gradiente, de manera que la red pueda entrenarse de manera más eficiente, al dejar fluir la información y gradientes a lo largo de esta, esto se materializa en una mayor precisión en la clasificación de imágenes.

Generalmente, se utiliza ResNet como red *backbone*¹³, puesto que debido a su capacidad de extracción de características (Gu *et al.*, 2022) permite obtener representaciones eficientes y adecuadas para la profundidad de la red, siendo esto crucial para un buen rendimiento en problemas de Computer Vision.

¹³Estructura principal de una red para extraer características, que luego son utilizadas por otras partes de la red.

3.4.4. MÉTRICAS, EVALUACIÓN Y COMPARACIÓN

- **Recall** (o referido también como sensibilidad): es especialmente importante en el diagnóstico de cáncer, ya que se desea minimizar los falsos negativos, es decir, que el modelo no pase por alto casos de cáncer. Un falso negativo puede ser fatal, ya que el paciente no recibiría tratamiento a tiempo, lo que podría permitir que el cáncer avance sin ser detectado.
- **F1-score**: en la clasificación de imágenes médicas, es útil porque ofrece un equilibrio entre la *precision* (evitar falsos positivos) y la sensibilidad (evitar falsos negativos). Esto es particularmente importante en el contexto del cáncer oral, donde tanto sobrediagnosticar (falsos positivos) como no diagnosticar (falsos negativos) tienen consecuencias serias.
- **Precision**: en este contexto, es relevante porque evita los falsos positivos, es decir, casos en los que el modelo clasifica incorrectamente una lesión benigna como cancerosa. Un alto número de falsos positivos puede generar estrés innecesario en los pacientes, conducir a pruebas adicionales invasivas (como biopsias) y aumentar los costos médicos.
- **AUC-ROC**: mide la capacidad del modelo para distinguir entre clases (por ejemplo, benigno y maligno) en distintos umbrales.
- **Matriz de confusión**: aunque no es una métrica en sí, nos permite visualizar de mejor manera el rendimiento del modelo, ya que contrasta los verdaderos positivos (TP), falsos positivos (FP), falsos negativos (FN) y verdaderos negativos (TN), permitiendo ver dónde exactamente el modelo está cometiendo errores. Es especialmente útil para entender cómo el modelo está manejando ambas clases (benignas y malignas).
- **Tasa de Verdaderos Negativos (TNR, también conocido como especificidad)**: Mide qué tan bien un modelo identifica correctamente las muestras negativas, en palabras simples: ¿qué porcentaje clasificó correctamente el modelo como negativas?

En este contexto, un TNR alto asegura que los casos benignos no sean clasificados erróneamente como malignos, evitando alarmas falsas, estrés en los pacientes y pruebas adicionales innecesarias.

3.5. ARQUITECTURA DEL MODELO DE APRENDIZAJE AUTOMÁTICO PROPUESTO

La arquitectura del modelo de aprendizaje automático propuesto para la clasificación y segmentación de imágenes en el contexto de detección de cáncer oral se basa en una combinación de modelos de clasificación preentrenados (ResNet50, InceptionV3 y Xception) y un modelo de segmentación, *Segment Anything Model* (SAM). Este enfoque híbrido tiene como objetivo no solo clasificar las imágenes en benignas o malignas, sino también proporcionar segmentaciones precisas que puedan ser útiles para una mejor visualización y comprensión de las áreas afectadas en las imágenes. El proceso general involucra varias etapas, que incluyen la recopilación y preprocesamiento de datos, el uso de modelos preentrenados para la clasificación, el ensamblaje de múltiples modelos para obtener predicciones robustas, y el uso de un modelo de segmentación que complementa los resultados de clasificación, mostrando visualmente las regiones relevantes de las imágenes.

3.5.1. PROCESAMIENTO DE DATA

El procesamiento de datos es una etapa crítica en el *pipeline* de aprendizaje automático, ya que garantiza que los datos de entrada estén en el formato y calidad adecuados para ser utilizados por los modelos. En este proyecto, se recopilaron múltiples conjuntos de datos de imágenes de lesiones orales provenientes de diversas fuentes, cada uno con características y formatos específicos.

El procesamiento de datos incluyó las siguientes etapas:

1. **Recopilación y acumulación de datos:** los conjuntos de datos utilizados fueron obtenidos de varias fuentes públicas, cada uno dividido en subconjuntos de entrenamiento, validación y prueba. Para garantizar una mejor generalización del modelo, los datos fueron organizados en carpetas correspondientes a sus clases (“benigno” y “maligno”) y a sus propósitos (entrenamiento, validación y prueba).
2. **Preprocesamiento de imágenes:**
 - **Redimensionamiento:** las imágenes fueron estiradas y ajustadas uniformemente a 224x224 píxeles.
 - **Orientación automática:** todas las imágenes fueron reorientadas automáticamente para asegurar alineación.
 - **Normalización:** los valores de los píxeles fueron escalados en el rango [0, 1] utilizando un factor de normalización de 1/255.
 - **Ajuste de Contraste:** solo aplicado en el *dataset* 3 para mejorar la visibilidad de características.

3. **Aumento de datos:** para mejorar la robustez del modelo y evitar sobreajuste debido a la cantidad limitada de imágenes, se aplicaron diversas técnicas de aumento de datos en tiempo real, tales como:

- **Reescalado (Normalization):** todas las imágenes son reescaladas por un factor de $1/255$ para normalizar los valores de los píxeles en el rango de $[0, 1]$, lo cual es común para mejorar el rendimiento de las redes neuronales.
- **Rotación (Rotation):** las imágenes se rotan de manera aleatoria en un rango de -15° a $+15^\circ$, lo que introduce variaciones en la orientación sin modificar la semántica de la imagen.
- **Desplazamiento Horizontal y Vertical (Width and Height Shift):** las imágenes pueden ser desplazadas horizontalmente y verticalmente hasta un 10% de su ancho y alto, lo que simula una pequeña traslación y mejora la robustez ante variaciones de posición.
- **Zoom:** se permite hacer zoom en las imágenes con un rango de hasta 20%, lo que introduce variaciones en la escala y ayuda al modelo a manejar imágenes más cercanas o alejadas.
- **Flip Horizontal y Vertical:** las imágenes se pueden reflejar horizontal y verticalmente, lo que introduce simetrías y ayuda a capturar variaciones espaciales comunes, especialmente útiles para detectar patrones simétricos en lesiones.
- **Ajuste de Brillo (Brightness Adjustment):** el brillo de las imágenes puede variar entre un 80% y 120% del brillo original, lo que introduce variaciones en la iluminación y mejora la adaptabilidad del modelo a diferentes condiciones de luz.
- **Cizallamiento (Shear):** se aplica una transformación de cizalladura de hasta 0.1, lo que distorsiona las imágenes al inclinarlas ligeramente, ayudando a captar deformaciones que puedan ocurrir en las imágenes reales.

3.5.2. MODELOS DE CLASIFICACIÓN

Al momento de realizar el desarrollo de los modelos, se implementaron y evaluaron varios modelos de clasificación basados en arquitecturas de redes neuronales convolucionales (CNN) preentrenadas. Estos modelos fueron seleccionados y adaptados con modificaciones específicas para abordar el problema de detección temprana de cáncer bucal. Luego de experimentar con diversos modelos, se eligieron los siguientes:

- **Xception - Model 3 (MultiHead Attention)**

El modelo basado en la arquitectura Xception se modificó para incluir componentes avanzados que mejoran la extracción de características y el aprendizaje jerárquico. Entre las modificaciones más destacadas se encuentran:

1. **Capas convolucionales y normalización inicial:** se añadió un bloque convolucional con 128 filtros, seguido de normalización mediante *BatchNormalization* y una operación de *MaxPooling*, para reducir la dimensionalidad mientras se conservan las características importantes.
2. **Atención multi-cabeza:** se incorporó un mecanismo de *MultiHeadAttention*¹⁴ con 4 cabezas, permitiendo que el modelo aprenda representaciones ricas al capturar dependencias globales en los datos.
3. **Capa densa final:** la arquitectura incluye un bloque compuesto de una capa densa con regularización mediante penalización L2, normalización, y activación ReLU, seguido de una capa de predicción con función *sigmoid*.

Este modelo utiliza el optimizador Adam¹⁵ con una tasa de aprendizaje inicial de 5×10^{-5} , lo que asegura una convergencia estable y eficiente.

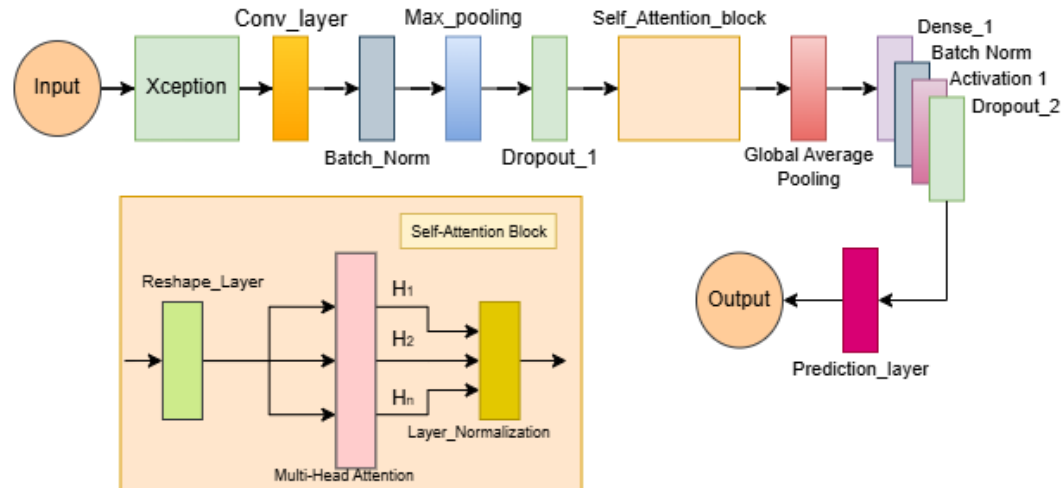
¹⁴La atención multi-cabeza permite que un modelo enfoque en diferentes partes de los datos simultáneamente, mejorando la extracción de patrones complejos.

¹⁵Adam es un algoritmo de optimización utilizado en *Machine Learning* que combina SGD con el manejo adaptativo de tasas de aprendizaje, utilizando los promedios acumulativos de gradientes y sus cuadrados para un ajuste eficiente de los parámetros del modelo.

Estructura:

Figura 8

Diagrama de arquitectura agregada a Xception.



Nota. Fuente: Autoría propia.

■ **ResNet50 - Model 6 (Self-Attention)**

El modelo basado en ResNet50 se diseñó para incorporar capacidades avanzadas de atención y regularización. Las principales características incluyen:

1. **Transformación inicial:** se añadió un bloque convolucional y operaciones de *Dropout* para mitigar el sobreajuste y aumentar la robustez.
2. **Atención tipo Transformer:** el modelo emplea un bloque de *Self-Attention*¹⁶, inspirado en los *Transformers*¹⁷, para mejorar el aprendizaje de relaciones globales entre características.
3. **Capas densas con regularización:** el modelo integra capas densas con activación ReLU y regularización L2, lo que ayuda a reducir la complejidad del modelo sin comprometer el rendimiento.

La optimización se realiza mediante SGD¹⁸ con una tasa de aprendizaje inicial de 5×10^{-4} , *Nesterov* activado y con un momentum de 0,9, lo que mejora la estabilidad y velocidad de convergencia.

¹⁶Es un mecanismo de aprendizaje automático que permite a las redes neuronales evaluar la importancia de los distintos elementos de entrada para realizar predicciones.

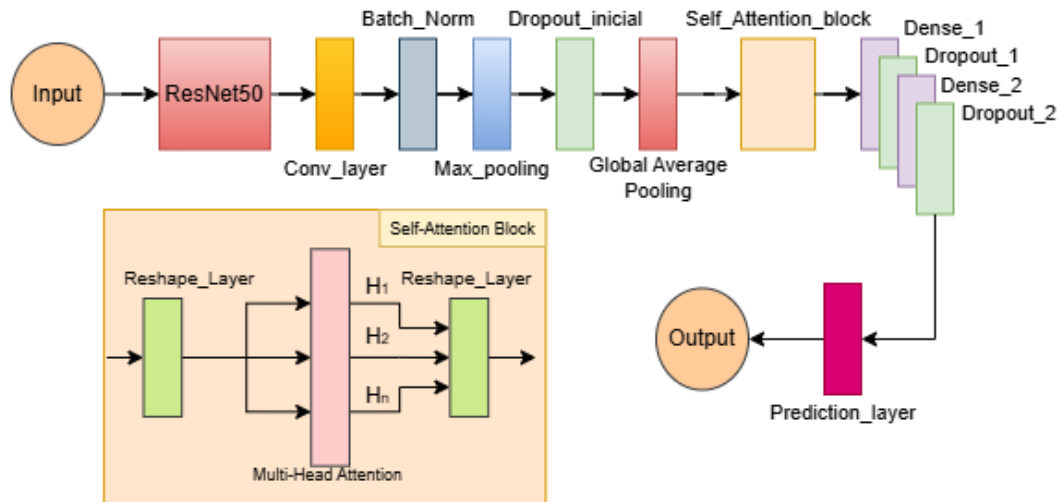
¹⁷Arquitectura de redes neuronales diseñada para procesar secuencias de datos, destacándose en tareas de Procesamiento de Lenguaje Natural y visión por computadora. Utilizan mecanismos de *Self-Attention* y *Feed-forward Layers* para capturar relaciones contextuales en los datos, permitiendo un procesamiento eficiente y paralelo. Fueron introducidos por Vaswani *et al.* (2023) en el paper *Attention Is All You Need*.

¹⁸SGD es un método de optimización iterativo utilizado en el entrenamiento de modelos de *Machine Learning*, ajusta los parámetros del modelo utilizando gradientes calculados sobre pequeñas muestras del *dataset*, propiciando la convergencia eficiente en grandes conjuntos de datos.

■ **Estructura:**

Figura 9

Diagrama de arquitectura agregada a ResNet50.



Nota. Fuente: Autoría propia.

■ **InceptionV3 - Model 4 (Enfoque Mixto)**

El modelo basado en InceptionV3 se diseñó como una arquitectura híbrida que combina las capacidades de InceptionV3 y ResNet50. Sus componentes principales son:

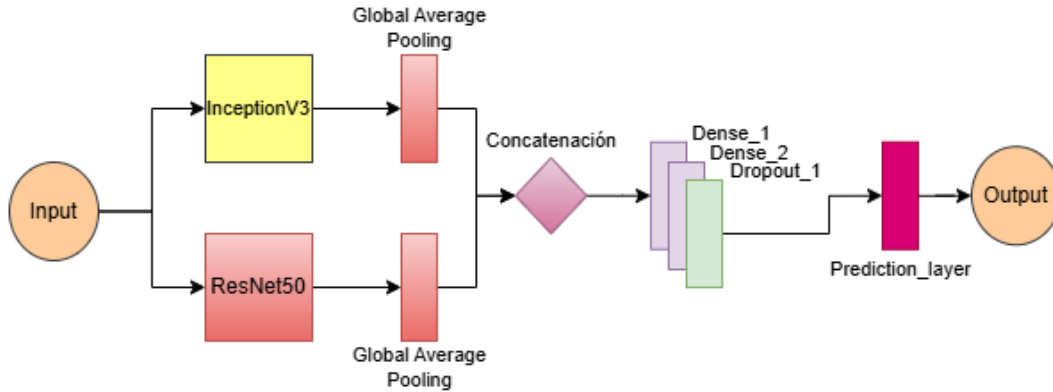
1. **Combinación de arquitecturas:** este modelo procesa la entrada simultáneamente a través de las arquitecturas InceptionV3 y ResNet50, utilizando pesos preentrenados en ImageNet.
2. **Fusión de características:** las características extraídas por ambas arquitecturas se combinan mediante una operación de concatenación, seguida de un bloque de densas con alta capacidad (4096 y 1024 unidades) para capturar interacciones complejas.
3. **Regularización avanzada:** se emplea regularización L2 y *Dropout* en las capas densas para prevenir el sobreajuste.

El modelo utiliza el optimizador Adam con una tasa de aprendizaje inicial de $8,7 \times 10^{-5}$, ajustado específicamente para este diseño híbrido.

Estructura:

Figura 10

Diagrama de arquitectura agregada a InceptionV3.



Nota. Fuente: Autoría propia.

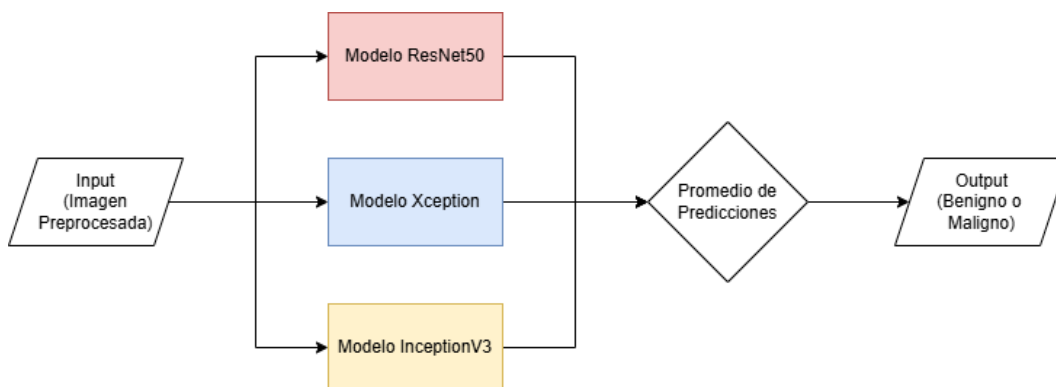
■ **Ensamble**

El ensamble de modelos integra las predicciones de diversas arquitecturas preentrenadas (Xception, ResNet50 e InceptionV3) con el objetivo de mejorar la precisión y robustez en el diagnóstico del cáncer oral. Al promediar los resultados individuales de los modelos, se obtiene una predicción más equilibrada y confiable, mitigando las limitaciones y sesgos inherentes a cada arquitectura por separado.

Estructura

Figura 11

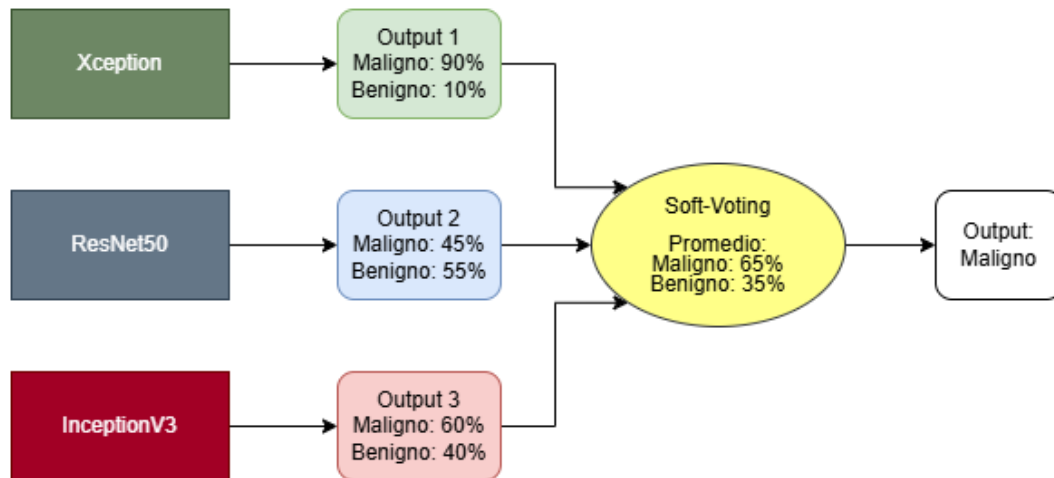
Diagrama general del ensamble.



Nota. Un diagrama para visualizar el ensamble de modelos, Fuente: Autoría propia.

Figura 12

Diagrama general de votación del ensemble.



Nota. Un diagrama para visualizar como votarían los modelos en el ensemble, Fuente: Autoría propia.

3.5.3. ETAPA DE ENTRENAMIENTO

La etapa de entrenamiento de los modelos propuestos se llevó a cabo utilizando una estrategia de *K-Fold Cross-Validation* con 5 pliegues, permitiendo una evaluación más robusta y reduciendo el riesgo de resultados dependientes de una única división del conjunto de datos. Los modelos se entrenaron con un máximo de 100 épocas, aunque en la práctica el entrenamiento solía concluir alrededor de las épocas 60 y 70 gracias a la implementación de *Early Stopping* y *ReduceLROnPlateau* como *callbacks*, configurados de la siguiente manera:

1. Early Stopping:

- Monitor: *val_loss* (pérdida en el conjunto de validación).
- Paciencia: 7 épocas sin mejora.
- Restauración de los mejores pesos: Activada.

2. ReduceLROnPlateau:

- a) Monitor: *val_loss*.
- b) Factor: Reducción de la tasa de aprendizaje en un factor de 0,5.
- c) Paciencia: 3 épocas sin mejora.
- d) Tasa de aprendizaje mínima: 1×10^{-7}

La función de pérdida utilizada en los tres modelos fue *Binary Cross-Entropy*¹⁹, optimizando la clasificación binaria entre lesiones benignas y malignas. El tamaño de lote (*batch size*) se fijó en 32 para equilibrar el uso eficiente de los recursos computacionales y la estabilidad del entrenamiento.

El criterio de selección del mejor modelo fue la menor pérdida en el conjunto de validación (*val_loss*), dado que se observó que este criterio producía mejores resultados generales en comparación con el uso de métricas como la *accuracy* o el F1-Score, cuyos valores eran levemente inferiores.

El entrenamiento se realizó en una GPU L4 de Google Colab, con acceso a 22,5 GB de memoria GPU, 53,0 GB de RAM, y 235,7 GB de espacio en disco, utilizando el backend de *Google Compute Engine* en un entorno Python 3. Estas características permitieron un entrenamiento eficiente y estable de los modelos, asegurando tiempos de entrenamiento razonables, con una duración promedio de 2 a 3 horas por modelo.

En general, las curvas de pérdida y precisión mostraron una convergencia estable alrededor de las épocas 60-70, lo que refleja un entrenamiento controlado y sin signos de sobreajuste significativo. La configuración mencionada permitió seleccionar modelos que alcanzaron un buen balance entre sensibilidad y especificidad, optimizando su desempeño en el conjunto de prueba.

El post-procesado se realizó para convertir las probabilidades generadas por los modelos en predicciones finales que pudieran ser evaluadas y analizadas. La salida de cada modelo consiste en una probabilidad generada por la última capa (*sigmoid*), representando la confianza del modelo en que una muestra pertenece a la clase maligna. Estas probabilidades fueron transformadas en etiquetas binarias (benigno/maligno) utilizando un umbral (*threshold*) optimizado para maximizar el desempeño del modelo.

¹⁹Función de pérdida utilizada en problemas de clasificación binaria, que mide la diferencia entre las etiquetas reales (0 o 1) y las probabilidades predichas por el modelo.

El *threshold* se definió en base al conjunto de validación, utilizando el F1-score como criterio principal para optimizar el equilibrio entre *precision* y *recall*. Los valores seleccionados fueron:

- Xception: umbral óptimo de 0,44.

- ResNet50: umbral óptimo de 0,41.

- InceptionV3: umbral óptimo de 0,32.

- Ensamble de modelos: umbral óptimo de 0,46

Estos valores garantizan un balance entre evitar falsos negativos (*recall* alto) y minimizar falsos positivos (*precision* alto), lo cual es crítico en el contexto médico para la detección de cáncer oral.

Además, se generaron matrices de confusión para analizar el comportamiento del modelo en términos de verdaderos positivos (TP), falsos positivos (FP), verdaderos negativos (TN) y falsos negativos (FN). Estas matrices proporcionaron una visión clara de los errores del modelo y permitieron ajustar los umbrales de forma informada.

3.6. DESCRIPCIÓN DE DATOS

A continuación, se presentan las transformaciones y técnicas de aumento de datos aplicadas por los autores de cada *dataset*. Adicionalmente, se detallan las modificaciones realizadas en este trabajo, cuyo objetivo principal fue la estandarización de las dimensiones de las imágenes para unificar todas las fuentes de datos. Esto permitió garantizar la coherencia en el proceso de entrenamiento, facilitar la integración de múltiples conjuntos de datos y optimizar el uso de recursos computacionales sin comprometer la calidad del aprendizaje del modelo.

1. *Dataset* 1 (Project, 2023)

- **Preprocesamiento:**

- a) Auto-orientación: aplicada para asegurar que todas las imágenes estén correctamente orientadas.
- b) Redimensionamiento: originalmente, las imágenes se estiraron para ajustar a un tamaño de 640x640 píxeles, pero para efectos de costo computacional y estandarización de dimensiones, fue escalada a 224x224 píxeles.
- c) Modificación de Clases: originalmente habían 4 clases que fueron remapeadas; “Benign” y “Benign Lesion” fueron reclasificadas como “benigno” y “precancerous” como “maligno”, existía la etiqueta de “null”, pero solo contenía 1 elemento que no fue tomado en cuenta.

- **Aumentos de Datos:** no se aplicaron aumentos en este conjunto de datos.

- **División del Conjunto de Datos:**

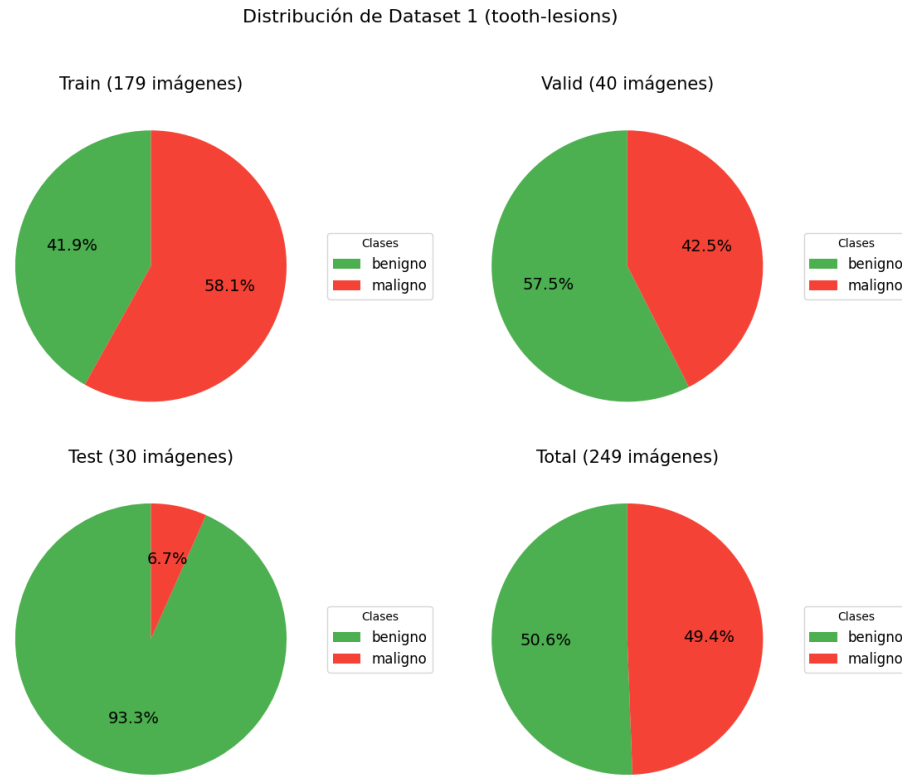
- Entrenamiento: 215 imágenes
- Validación: 50 imágenes
- Prueba: 34 imágenes

- **Observaciones:** este conjunto de datos se enfoca en lesiones dentales, proporcionando imágenes de alta resolución que capturan diversas afecciones en los dientes. La ausencia de aumentos de datos sugiere que se trabajó con las imágenes originales para preservar la integridad de las características presentes.

■ **Proporciones:**

Figura 13

Gráficos de Proporciones (Train, Valid, Test, Total)



Nota. Fuente: Autoría propia.

■ **Muestra:**

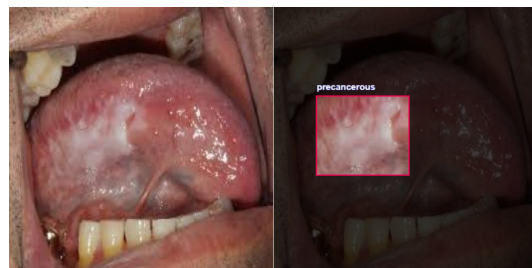
Figura 14

Comparación entre ejemplos de lesiones benignas y malignas.

(a) Ejemplo Benigno



(b) Ejemplo Maligno



Nota. Obtenido de tooth lesions Dataset, Fuente: Project (2023)

2. Dataset 2 (Framejframej_1, 2023)

■ **Preprocesamiento:**

- a) Auto-orientación: aplicada.
- b) Redimensionamiento: originalmente estirado a 640x640 píxeles, y fue reajustado a 224x224 píxeles.
- c) Modificación de clases: originalmente existían 3 clases, “common-oral-lesion” fue reclasificada como “benigno”, “precancerous-oral-lesion” fue reclasificada como “maligno” y “null” que no poseía elementos.

■ **Aumentos de Datos:**

- a) **Outputs por ejemplo de entrenamiento:** 3;
- b) **Flip Horizontal:** aplicado para duplicar las imágenes reflejándolas horizontalmente.
- c) **Crop (Recorte):** zoom mínimo del 0% y máximo del 40%, permitiendo variaciones en el enfoque de las imágenes.
- d) **Rotación:** entre -15° y $+15^\circ$, introduciendo variaciones angulares.
- e) **Desenfocado:** hasta 1.5 píxeles, simulando condiciones de imagen menos nítidas.

■ **División del Conjunto de Datos:**

- Entrenamiento: 518 imágenes
- Validación: 49 imágenes
- Prueba: 22 imágenes

- ### ■ **Observaciones:** Este conjunto de datos se centra en lesiones orales generales. Los aumentos aplicados incrementan la diversidad del conjunto de entrenamiento, ayudando al modelo a generalizar mejor al introducir variaciones en orientación, escala y nitidez.

■ **Proporciones:**

Figura 15

Gráficos de Proporciones (Train, Valid, Test, Total)



Nota. Fuente: Autoría propia.

■ **Muestra:**

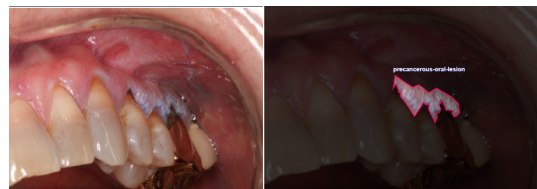
Figura 16

Comparación entre ejemplos de lesiones benignas y malignas.

(a) Ejemplo Benigno



(b) Ejemplo Maligno



Nota. Obtenido de Oral lesion Dataset, Fuente: Framejframej_1 (2023)

3. Dataset 3 (Kean, 2024)

■ **Preprocesamiento:**

- a) Auto-orientación: aplicada.
- b) Ajuste automático de contraste: utilizando ecualización de histograma para mejorar la visibilidad de las características en las imágenes.
- c) Redimensionamiento: originalmente, las dimensiones de las imágenes eran variadas, pero fueron escaladas a 224x224 para mantener la consistencia y costo computacional asociado.
- d) Modificación de clases: originalmente, poseía 3 clases, "Normal_mouth" y "Aphthous"²⁰ fueron reclasificadas como "benigno" y "oral_cancer" fue reclasificada como "maligno".

■ **Aumentos de Datos:**

- a) Salidas por ejemplo de entrenamiento: 2
- b) Rotación: entre -7° y +7°.
- c) Brillo: variaciones entre -10 % y +10 %, simulando diferentes condiciones de iluminación.

■ **División del Conjunto de Datos:**

- Entrenamiento: 260 imágenes
- Validación: 42 imágenes
- Prueba: 42 imágenes

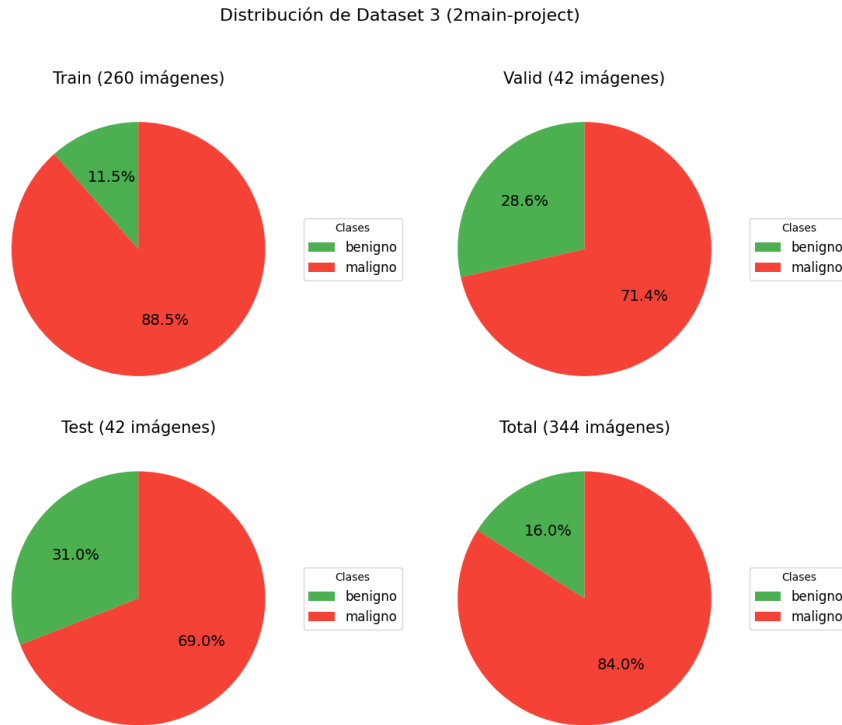
- #### ■ **Observaciones:** este conjunto incorpora un ajuste de contraste para resaltar detalles cruciales en las imágenes de lesiones orales. Los aumentos de rotación y brillo amplían la variedad de condiciones visuales, fortaleciendo la robustez del modelo ante cambios en iluminación y orientación.

²⁰Las lesiones aftosas no son cancerígenas y suelen ser causadas por un crecimiento excesivo del hongo cándida o por una deficiencia de vitamina B. (DeMarco, 2022)

■ **Proporciones:**

Figura 17

Gráficos de Proporciones (Train, Valid, Test, Total)



Nota. Fuente: Autoría propia.

■ **Muestra:**

Figura 18

Comparación entre ejemplos de lesiones benignas y malignas.

(a) Ejemplo Benigno



(b) Ejemplo Maligno



Nota. Obtenido de 2main project Dataset, Fuente: Kean (2024)

4. Dataset 4 (Bits, 2023)

- **Preprocesamiento:**

- a) Auto-orientación: aplicada.
- b) Redimensionamiento: originalmente estirado a 640x640 píxeles, pero para efectos de costo computacional y estandarización de dimensiones, fue escalada a 224x224 píxeles, igual que los demás datasets estirados.
- c) Modificación de clases: anteriormente, se habían remapeado 5 clases, no obstante, para efectos de consistencia, las clases “cancer”, “oral-cancer” y “oral cancer” fueron reclasificadas como “maligno”, las clases “non-cancer” y “n-cancer” fueron reclasificadas como “benigno” y la clase “null” fue ignorada, puesto que no poseía elementos.

- **Aumentos de Datos:** no se aplicaron aumentos.

- **División del Conjunto de Datos:**

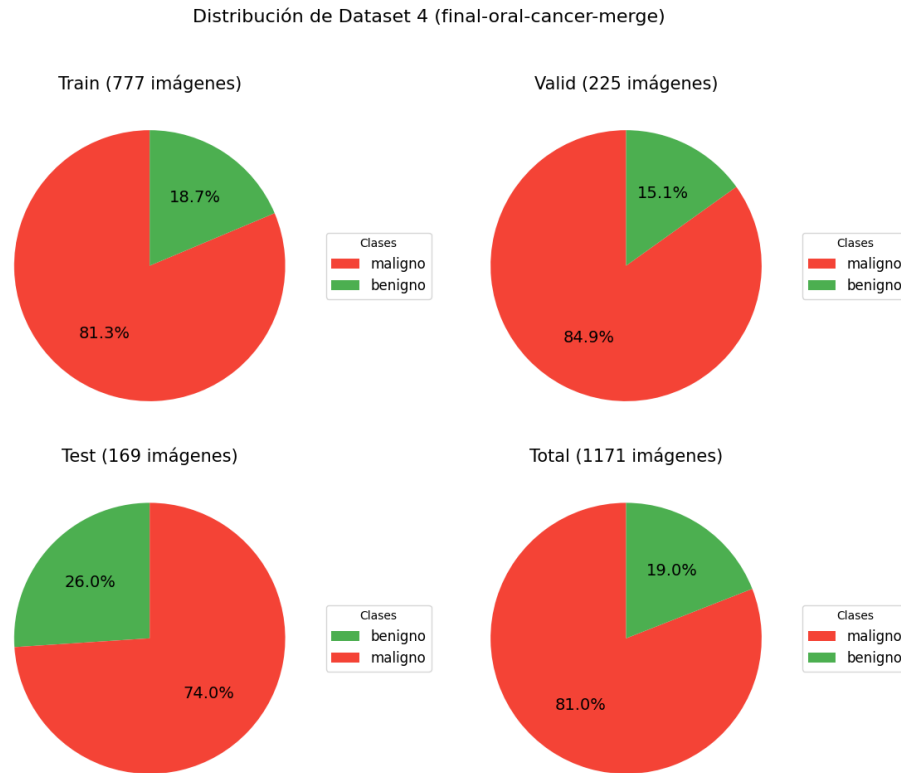
- Entrenamiento: 778 imágenes
- Validación: 225 imágenes
- Prueba: 169 imágenes

- **Observaciones:** este conjunto está orientado al cáncer oral, con un considerable número de imágenes que proporcionan una base sólida para el entrenamiento. La modificación de clases indica una reestructuración de las categorías para mejorar la clasificación. La ausencia de aumentos sugiere confianza en la variabilidad intrínseca del conjunto.

■ **Proporciones:**

Figura 19

Gráficos de Proporciones (Train, Valid, Test, Total)



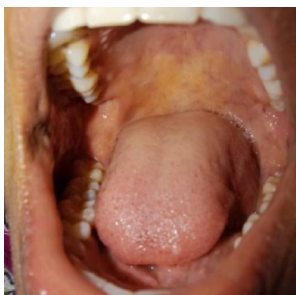
Nota. Fuente: Autoría propia.

■ **Muestra:**

Figura 20

Comparación entre ejemplos de lesiones benignas y malignas.

(a) Ejemplo Benigno



(b) Ejemplo Maligno



Nota. Obtenido de *final oral cancer merge Dataset*, Fuente: Bits (2023)

5. Dataset 5 (Cancer Research, 2024)

■ **Preprocesamiento:**

- a) Auto-orientación: aplicada.
- b) Redimensionamiento: originalmente estirado a 640x640 píxeles, no obstante, fue escalado a 224x224 para mantener la consistencia y disminuir el costo computacional.
- c) Modificación de clases: esta ya poseía 2 clases, "NONCANCER" fue renombrada como "benigno" y "CANCER" fue renombrada como "maligno".

■ **Aumentos de Datos:**

- a) Salidas por ejemplo de entrenamiento: 3
- b) *Flip* Horizontal: para aumentar la diversidad espacial.
- c) *Crop* (Recorte): zoom mínimo del 0% y máximo del 20%.

■ **División del Conjunto de Datos:**

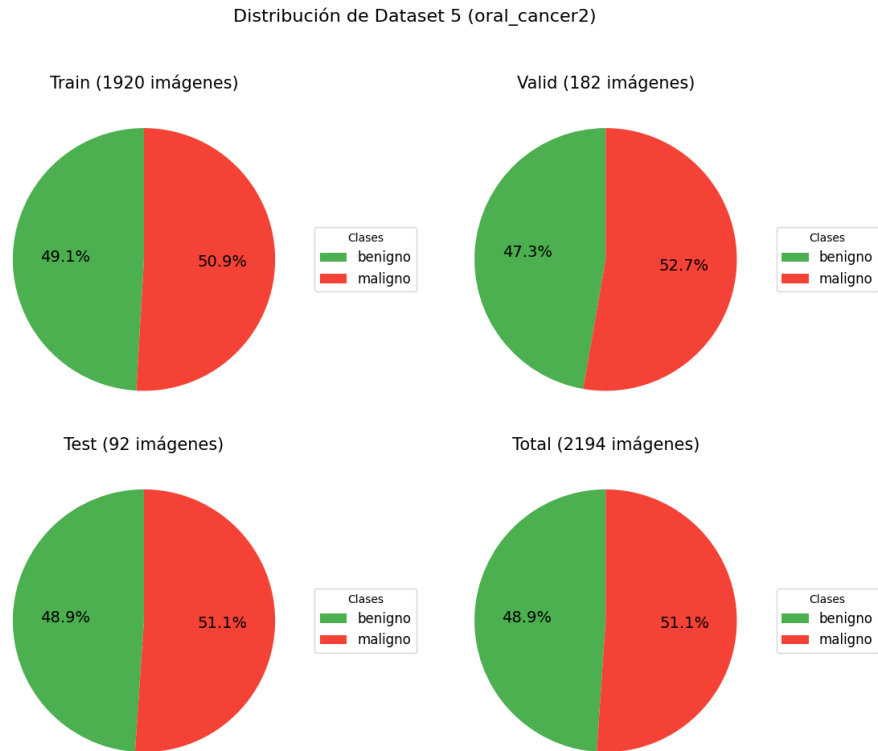
- Entrenamiento: 1920 imágenes
- Validación: 182 imágenes
- Prueba: 92 imágenes

- ### ■ **Observaciones:** este es el conjunto de datos más extenso utilizado, enfocado en cáncer oral. Los aumentos aplicados contribuyen a una mayor diversidad en las imágenes de entrenamiento, crucial para capturar variaciones anatómicas y de presentación de la enfermedad.

■ **Proporciones:**

Figura 21

Gráficos de Proporciones (Train, Valid, Test, Total)



Nota. Fuente: Autoría propia.

■ **Muestra:**

Figura 22

Comparación entre ejemplos de lesiones benignas y malignas.

(a) Ejemplo Benigno



(b) Ejemplo Maligno



Nota. Obtenido de OralCancer2 Dataset, Fuente: Cancer Research (2024)

6. Dataset 6 (Haridas, 2024)

■ **Preprocesamiento:**

- a) Auto-orientación: aplicada.
- b) Redimensionamiento: originalmente estirado a 640x640 píxeles, no obstante, fue escalado a 224x224 para mantener la consistencia y disminuir el costo computacional.
- c) Modificación de clases: esta ya poseía 2 clases, "NONCANCER" fue renombrada como "benigno" y "CANCER" fue renombrada como "maligno".

■ **Aumentos de Datos:**

- a) Salidas por ejemplo de entrenamiento: 3
- b) *Flip* Horizontal: para aumentar la diversidad espacial.

■ **División del Conjunto de Datos:**

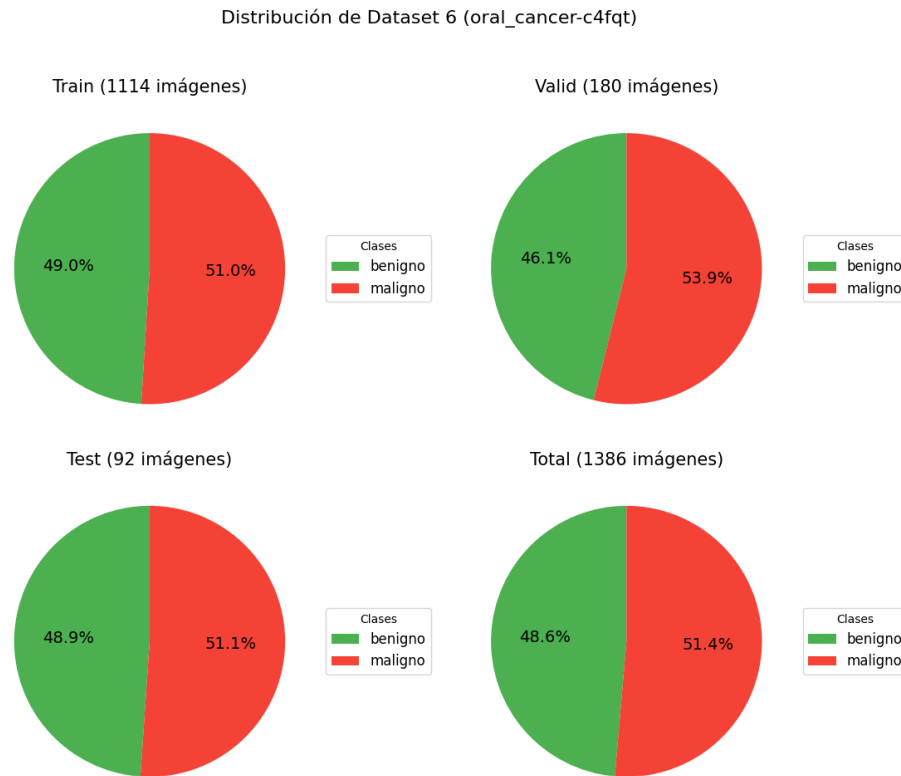
- Entrenamiento: 1114 imágenes
- Validación: 180 imágenes
- Prueba: 92 imágenes

- **Observaciones:** el conjunto de datos incluye una variedad de ejemplos de imágenes de la lengua, abarcando tanto patologías cancerosas como benignas, además de muestras de pacientes sanos. También, se presentan capturas de otras áreas de la cavidad bucal, como encías y paladar (tanto duro como blando), lo que otorga un gran valor al dataset debido a su diversidad anatómica y una resolución adecuada para tareas de análisis médico.

■ **Proporciones:**

Figura 23

Gráficos de Proporciones (Train, Valid, Test, Total)



Nota. Fuente: Autoría propia.

■ **Muestra:**

Figura 24

Comparación entre ejemplos de lesiones benignas y malignas.

(a) Ejemplo Benigno



(b) Ejemplo Maligno



Nota. Obtenido de ORAL CANCER Dataset, Fuente: Haridas (2024)

7. Dataset 7 (UNAB, 2023)

■ **Preprocesamiento:**

- a) Redimensionamiento: originalmente estirado a 640x640 píxeles, no obstante, fue escalado a 224x224 para mantener la consistencia y disminuir el costo computacional.
- b) Auto-orientación: aplicada.
- c) Modificación de clases: las clases “Abnormal”, “oral cancer” y “Null” fueron re-clasificadas como “maligno”, mientras que las clases restantes como “Abnormal Normal”, “detect-lesions”, “non-cancer”, “Normal”, “Normal Null” fueron reclasificadas como “benigno”.

■ **Aumentos de Datos:** no se aplicaron aumentos.

■ **División del Conjunto de Datos:**

- Entrenamiento: 2190 imágenes
- Validación: 286 imágenes
- Prueba: 201 imágenes

■ **Observaciones:** esta posee la clase ‘maligno’ levemente desbalanceada respecto a la clase ‘benigno’ por alrededor de una diferencia de 400 imágenes; esto se compensa levemente por la mitigación en el balance de los otros datasets y, además, esto también podría resultar beneficioso, puesto que muchas veces se omite la captura de imágenes que detallen a un paciente sano.

■ **Proporciones:**

Figura 25

Gráficos de Proporciones (Train, Valid, Test, Total)



Nota. Fuente: Autoría propia.

■ **Muestra:**

Figura 26

Comparación entre ejemplos de lesiones benignas y malignas.

(a) Ejemplo Benigno



(b) Ejemplo Maligno



Nota. Obtenidos de ORALCANCERUNAB Dataset, Fuente: UNAB (2023)

8. Dataset 8 (Yuk, 2024)

■ Preprocesamiento:

- a) Auto-orientación: aplicada.
- b) Redimensionamiento: originalmente estirado a 640x640 píxeles, no obstante, fue escalado a 224x224 para mantener la consistencia y disminuir el costo computacional.
- c) Modificación de clases: esta ya poseía 2 clases, "NONCANCER" fue renombrada como "benigno" y "CANCER" fue renombrada como "maligno".

■ Aumentos de Datos:

- a) Salidas por ejemplo de entrenamiento: 3
- b) *Flip* Horizontal: para aumentar la diversidad espacial.
- c) *Flip* Vertical: mismo propósito que el flip horizontal, aumentar variabilidad.
- d) Rotación: en sentido horario y antihorario.
- e) Escala Grises: aplicación a 15% de los datos por parte del autor, aumenta levemente el ruido y disminuye la dependencia del color y promueve el aprendizaje de formas.

■ División del Conjunto de Datos:

- Entrenamiento: 1048 imágenes
- Validación: 91 imágenes
- Prueba: 45 imágenes

- ### ■ Observaciones:
- este *dataset* presenta un desbalance de clases a favor de la etiqueta "benigno", lo que podría facilitar que los modelos desarrollen una mayor capacidad para reconocer casos sin patologías cancerosas. Además, incluye múltiples ejemplos de imágenes de la lengua y, como aspecto adicional, contiene datos destinados al entrenamiento en segmentación. No obstante, estas anotaciones de segmentación requieren una revisión por parte de un especialista para garantizar su calidad. De igual manera, es necesario recopilar más datos con las segmentaciones correspondientes para asegurar un entrenamiento robusto y efectivo que pueda complementar la clasificación.

■ **Proporciones:**

Figura 27
Gráficos de Proporciones (Train, Valid, Test, Total)



Nota. Fuente: Autoría propia.

■ **Muestra:**

Figura 28
Comparación entre ejemplos de lesiones benignas y malignas.

(a) Ejemplo Benigno



(b) Ejemplo Maligno



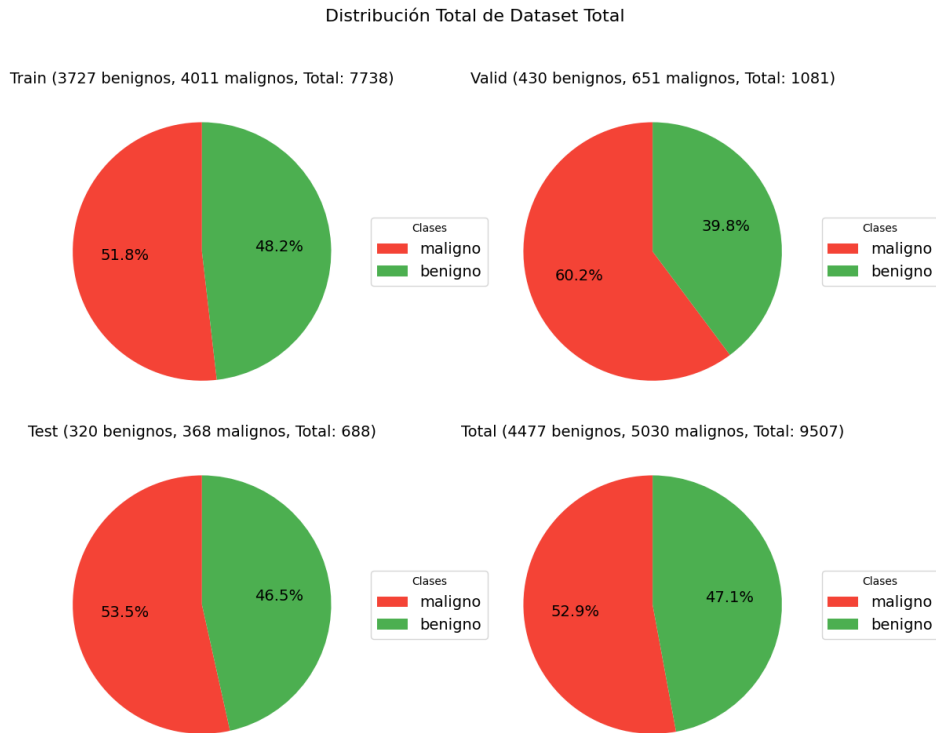
Nota. Obtenidos de *CekLidah Dataset*, Fuente: Yuk (2024)

9. Dataset Total

El conjunto de datos final se compone de la integración de los ocho *datasets* previamente mencionados. A continuación, se presentan las proporciones del *dataset* consolidado, detallando la distribución de cada clase, así como la segmentación en los subconjuntos de entrenamiento, validación y prueba.

Figura 29

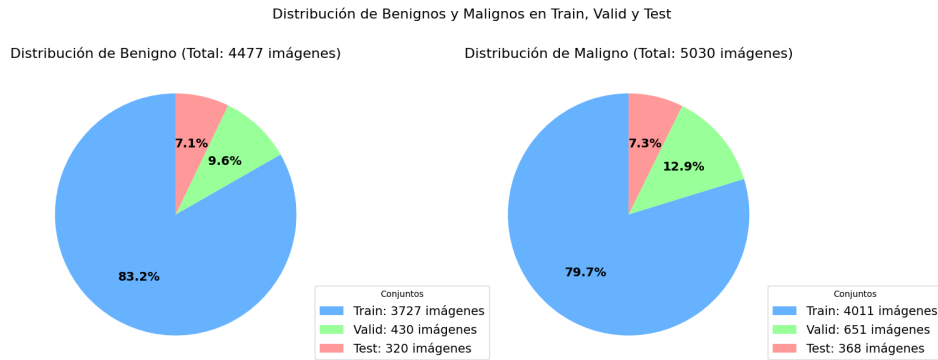
Distribución total en el dataset.



Nota. Los gráficos representan la proporción de las clases en el conjunto total y los subconjuntos *Train*, *Valid* y *Test* Fuente: Autoría propia.

Figura 30

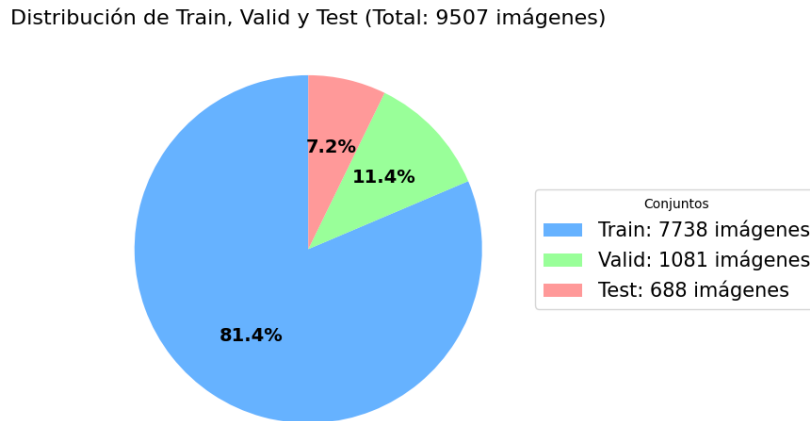
Distribución de clases en cada subconjunto del dataset.



Nota. Indica cuánto es el porcentaje de benignos y malignos destinados a cada subconjunto, Fuente: Autoría propia.

Figura 31

Distribución de Train, Valid y Test.



Nota. Muestra el porcentaje que representa cada subconjunto en relación al dataset total, Fuente: Autoría propia.

CAPÍTULO 4

VALIDACIÓN DE LA SOLUCIÓN

4.1. RESULTADOS Y MÉTRICAS DEL MODELO PROPUESTO

A través de numerosos experimentos donde se probaron diversos modelos pre-entrenados, junto con capas adicionales para aumentar la capacidad del modelo de extraer patrones de la data, se obtuvieron 3 modelos cuyas métricas fueron destacables en la tarea que se aborda en este trabajo, se realizaron modificaciones personalizadas a tres arquitecturas de redes neuronales profundas pre-entrenadas: Xception, ResNet50 e InceptionV3. Estas arquitecturas se seleccionaron por su eficacia en tareas de clasificación de imágenes y su capacidad para extraer características significativas a partir de los datos de entrada. Cada modelo base fue extendido con capas adicionales diseñadas para optimizar su desempeño en una tarea específica de clasificación binaria.

4.1.1. ARQUITECTURAS Y MODIFICACIONES

Las modificaciones a los modelos pre-entrenados consistieron en la adición de capas personalizadas que complementan las representaciones aprendidas por cada modelo base. A continuación, se presentan las métricas y matrices de confusión de cada una, finalizando con el análisis del rendimiento del ensamble de estos modelos:²¹²²

²¹Aquellas configuraciones con * significa que son aquellas elegidas para el resultado final que se verán a continuación.

²²Tal como mencionan Smith (2018), Laarhoven (2017) y Hoffer *et al.* (2018), al implementar *Dropout* y *BatchNormalization*, se compensa de cierta manera la necesidad de implementar *Weight Decay*, esto explica la ausencia de su implementación, aunque no se descarta su uso en valores pequeños para trabajos futuros que profundicen la aplicación de estos modelos.

1. Xception

El modelo Xception, conocido por su capacidad de aprender representaciones espaciales complejas mediante convoluciones separables en profundidad, fue adaptado con capas adicionales de atención y regularización para mejorar su capacidad de generalización en la tarea específica de clasificación binaria.

a) Métricas

Tabla 2

Métricas del modelo Xception en conjunto Test.

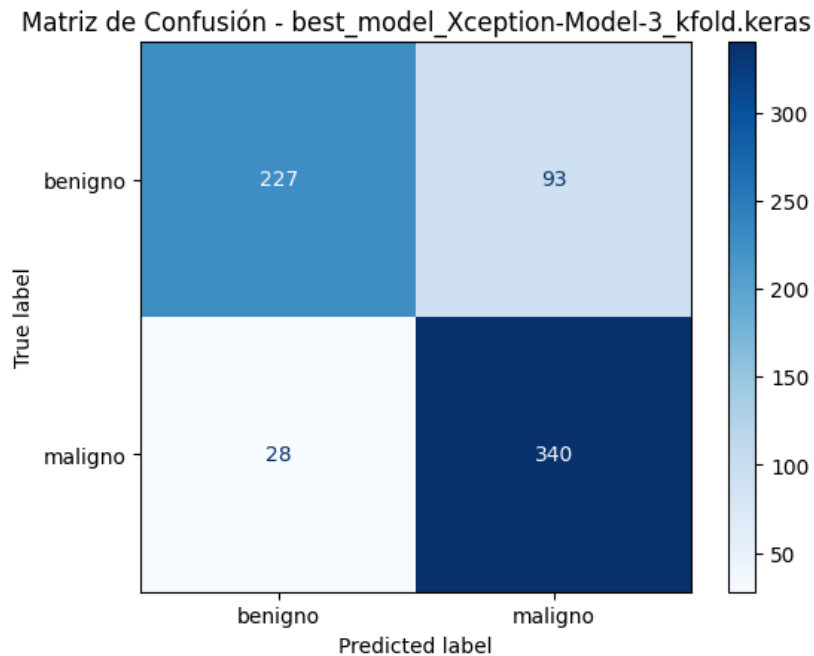
Métrica	Valor
Threshold óptimo	0,44
Accuracy	0,8241
Precision	0,7852
Recall	0,9239
F1-Score	0,8489
ROC-AUC	0,8566

b) Matriz de Confusión

La matriz de confusión correspondiente evaluada en el conjunto Test se presenta en la Figura 32, mostrando un desempeño sólido en ambas clases.

Figura 32

Matriz de confusión del modelo Xception.



Nota. Fuente: Autoría propia.

c) **Observaciones**

El modelo mostró un excelente balance entre *precision* y *recall*, logrando un F1-Score de 0,8489. Sin embargo, se observa una ligera tendencia a clasificar falsos negativos, lo cual podría ser mitigado con una regularización más óptima o más (y/o mejores) datos.

Fortalezas:

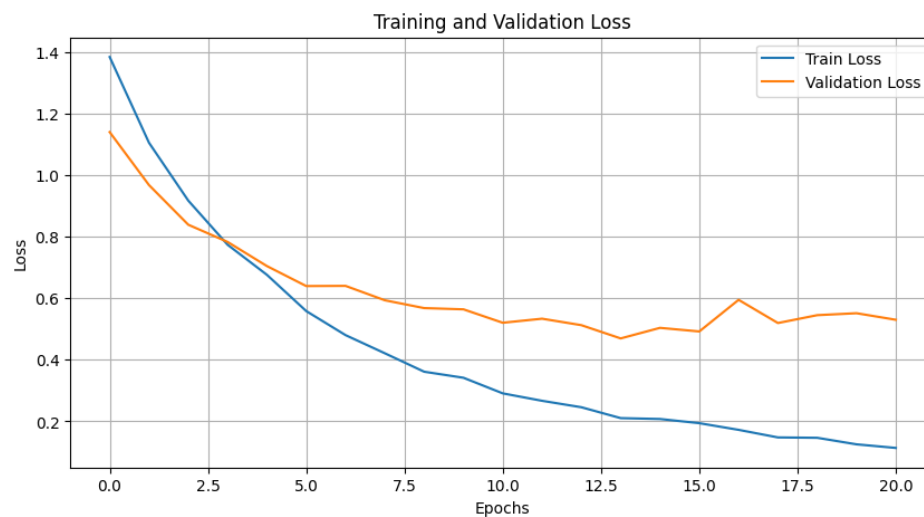
El modelo tiene un *recall* alto (0,924), lo que significa que detecta la mayoría de los casos “malignos”. Tiene un buen balance entre *precision* y *recall*, con un F1-Score de 0,849.

Debilidades:

La tasa de falsos positivos²³ (0,290) es significativa, lo que podría ser problemático si los casos “malignos” llevan a acciones invasivas. La especificidad (0,710) indica que el modelo podría mejorar en la clasificación de casos “benignos”.

Figura 33

Gráfico de Loss del modelo Xception.

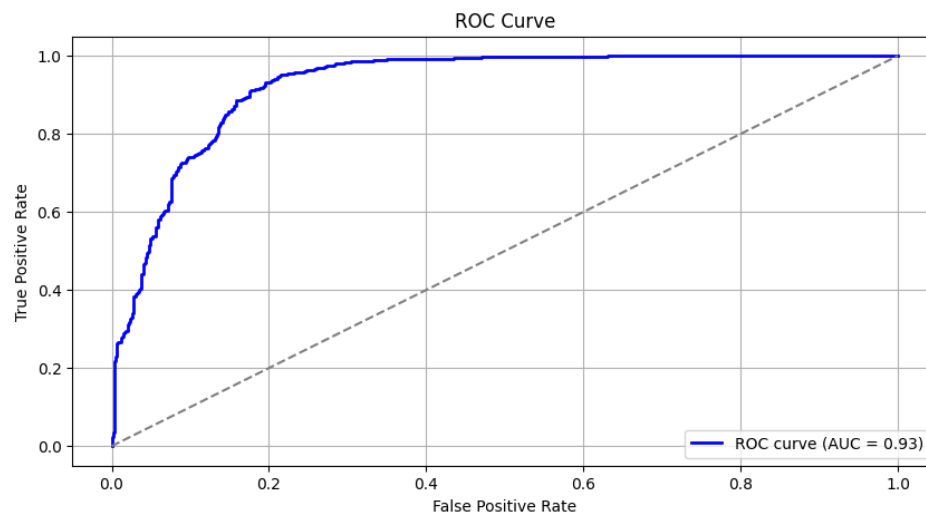


Nota. Fuente: Autoría propia.

²³Próximamente se denotará como FPR (*False Positive Rate*). Esta mide el porcentaje que el modelo clasificó incorrectamente como positivas.

Figura 34

Gráfico de ROC-AUC de Validación del modelo Xception.



Nota. Fuente: Autoría propia.

d) **Hiperparámetros:**

Tabla 3

Hiperparámetros de la Arquitectura Adicional a Xception.

Hiperparámetro	Valor
Epochs	100
Batch Size	{8; 16; 32* ; 64}
Dimensión Conv_Layer	{64; 128* ; 256; 512; 1024; 2048}
Funcion Conv_Layer	{ ReLU* ; Swish}
Kernel Size Conv_Layer	{{(1 x 1); (2 x 2); (3 x 3)* }
Kernel Size Max_Pooling	{{(1 x 1); (2 x 2)* ; (3 x 3)}
Dropout_1	{0,1; 0,2; 0,3; 0,4* ; 0,5; 0,6}
Número Cabezales	{2; 4* ; 6; 8}
Dimensión de Key	{64; 128 }
Nº Unidades Dense_1	64
Regularización_dense_1	{ 1e-2* ; 1e-3; 1e-4}
Función Activación_1	{ ReLU* ; Swish}
Dropout_2	{0,1; 0,2; 0,3; 0,4; 0,5* ; 0,6}
Regularización_output	{ 1e-2* , 1e-3, 1e-4}
Función Activación Output_Layer	{ Sigmoide* ; Tanh Centrada; Softmax}
Optimizador	{SGD, Adam* }
Learning Rate (Max)	{1e-4; 9e-5; 8,7e-5; 8e-5; 7,5e-5; 7e-5; 6e-5; 5e-5* ; 4e-5; 3e-5; 2e-5; 1e-5; 1e-6; 1e-7; 1e-8; 1e-9}
Weight Decay	None (Default)
ReduceLRonPlateau	
Learning Rate (Min)	1e-7
Paciencia	3
Factor	0,5
EarlyStopping	
Paciencia	7

2. ResNet50

El modelo ResNet50 fue elegido por su capacidad para mitigar el problema de la degradación en arquitecturas profundas. Se añadieron capas adicionales que integraron atención tipo *Transformer* para capturar relaciones globales entre las características aprendidas.

a) Tabla Métricas

Tabla 4

Métricas del modelo ResNet50 en conjunto Test.

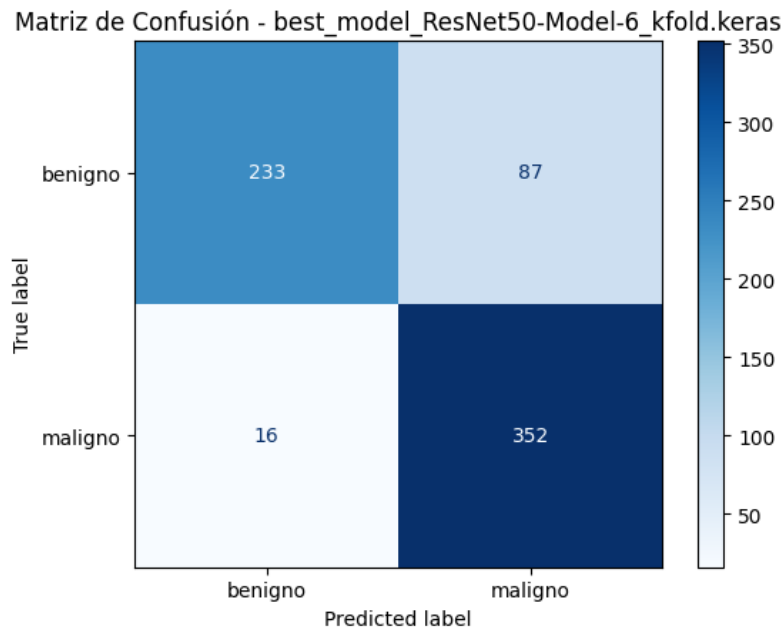
Métrica	Valor
Threshold óptimo	0,41
Accuracy	0,8503
Precision	0,8018
Recall	0,9565
F1-Score	0,8724
ROC-AUC	0,8597

b) Matriz de Confusión

La matriz de confusión correspondiente evaluada en el conjunto Test (Figura 35) muestra un excelente desempeño, con pocas instancias malignas clasificadas incorrectamente.

Figura 35

Matriz de confusión del modelo ResNet50.



Nota. Fuente: Autoría propia.

c) **Observaciones**

El modelo ResNet50 destacó particularmente en su capacidad para capturar patrones generales y específicos, logrando el mejor balance entre *recall* (0,9565) y F1-Score (0,8724).

Fortalezas:

Alto *recall* (0,956): El modelo detecta la mayoría de los casos “malignos”. Esto es crucial si los casos “malignos” representan una condición crítica donde se debe priorizar minimizar los falsos negativos. Buen balance: con un F1-Score de 0,872, el modelo muestra un equilibrio sólido entre precisión y sensibilidad. Alto *accuracy* (0,850): Un porcentaje elevado de predicciones correctas en general.

Debilidades:

Moderada tasa de falsos positivos (FPR ~ 27,2 %): un número considerable de casos “benignos” son clasificados como “malignos”, lo que podría generar alarmas innecesarias o tratamientos invasivos. Especificidad relativamente baja (0,728): Aunque el modelo es excelente en detectar casos “malignos”, su rendimiento es más limitado en identificar correctamente los casos “benignos”.

Figura 36

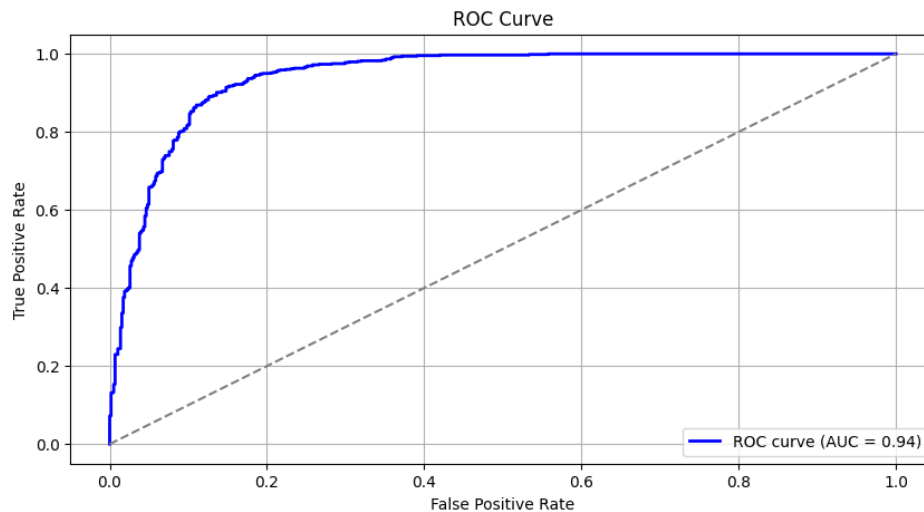
Gráfico de Loss del modelo ResNet50.



Nota. Fuente: Autoría propia.

Figura 37

Gráfico de ROC-AUC de Validación del modelo ResNet50.



Nota. Fuente: Autoría propia.

Preferencia por Recall:

Considerando que el problema exige el enfoque en la detección temprana de patologías orales cancerígenas, este modelo es altamente adecuado debido a su alta sensibilidad.

Mitigar falsos positivos:

Algo importante a mencionar (a pesar de su importante capacidad para identificar malignidad), es que aplicaciones donde los falsos positivos puedan generar un costo elevado (por ejemplo, pruebas adicionales y/o ansiedad en los pacientes); se podría ajustar el umbral de decisión o explorar técnicas adicionales para mejorar la especificidad.

d) **Hiperparámetros:**

Tabla 5

Hiperparámetros de la Arquitectura Adicional a ResNet50.

Hiperparámetro	Valor
Epochs	100
Batch Size	{8; 16; 32* ; 64}
Dimensión Conv_Layer	{64; 128* ; 256; 512; 1024; 2048}
Funcion Conv_Layer	{ ReLU* ; Swish}
Kernel Size Conv_Layer	{{(1 x 1); (2 x 2); (3 x 3)* }
Kernel Size Max_Pooling	{{(1 x 1); (2 x 2)* ; (3 x 3)}
Dropout_inicial	{0,1; 0,2; 0,3; 0,4* ; 0,5; 0,6}
Nº Cabezales Multi-Head	{2; 4* ; 8}
Nº Unidades Dense_1	{8; 16; 32; 64; 96; 128* ; 256}
Regularización Dense_1	{ 1e-2* ; 1e-3; 1e-4}
Función Activación Dense_1	{ ReLU* ; Swish}
Dropout_1	{0,1; 0,2; 0,3; 0,4* ; 0,5; 0,6}
Nº Unidades Dense_2	{8; 16; 32; 64* ; 96; 128; 256}
Regularización Dense_2	{ 1e-2* ; 1e-3; 1e-4}
Función Activación Dense_2	{ ReLU* ; Swish}
Dropout_2	{0,1; 0,2; 0,3; 0,4; 0,5* ; 0,6}
Regularización Output_Layer	{ 1e-2* ; 1e-3; 1e-4}
Función Activación Output_Layer	{ Sigmoide* ; Tanh Centrada; Softmax}
Optimizador	{ SGD* , Adam}
Learning Rate (Max)	{1e-3; 7,5e-4; 5e-4* ; 2,5e-4; 1e-4; 7,5e-5; 5e-5; 3e-5; 2,5e-5}
Momentum	0,9
Nesterov	{ True* , False}
Weight Decay	None (Default)
ReduceLRonPlateau	
Learning Rate (Min)	1e-7
Paciencia	3
Factor	0,5
EarlyStopping	
Paciencia	7

3. InceptionV3

El modelo InceptionV3, caracterizado por su diseño eficiente para procesar datos en múltiples escalas, fue combinado con ResNet50 para crear un modelo híbrido. Esta combinación permitió captar tanto características globales como específicas.

a) Tabla Métricas

Tabla 6

Métricas del modelo InceptionV3 en conjunto Test

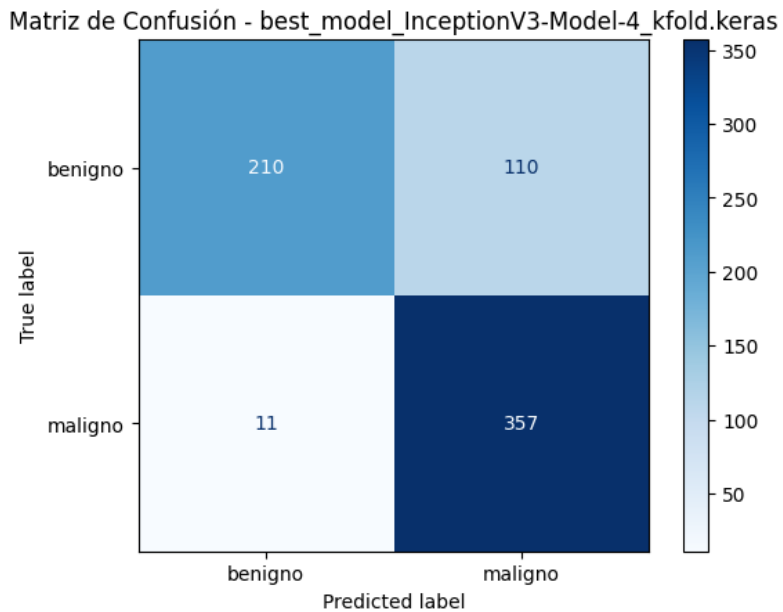
Métrica	Valor
Threshold óptimo	0,32
Accuracy	0,8241
Precision	0,7645
Recall	0,9701
F1-Score	0,8551
ROC-AUC	0,8673

b) Matriz de Confusión

La matriz de confusión correspondiente evaluada en el conjunto Test se muestra en la Figura 41, destacando el excelente desempeño en la clasificación de instancias positivas.

Figura 38

Matriz de confusión del modelo InceptionV3.



Nota. Fuente: Autoría propia.

c) **Observaciones**

El modelo basado en InceptionV3 muestra un alto rendimiento en la detección de casos malignos, con un *recall* de 0,970, lo que indica que identifica correctamente casi todos los casos positivos, minimizando falsos negativos. Su F1-Score de 0,855 y *accuracy* de 0,824 reflejan un balance adecuado entre precisión y sensibilidad. Sin embargo, su baja especificidad (0,656) y una tasa de falsos positivos de 34,4% pueden derivar en un número considerable de diagnósticos erróneos de lesiones benignas como malignas, lo que podría causar procedimientos innecesarios.

Fortalezas:

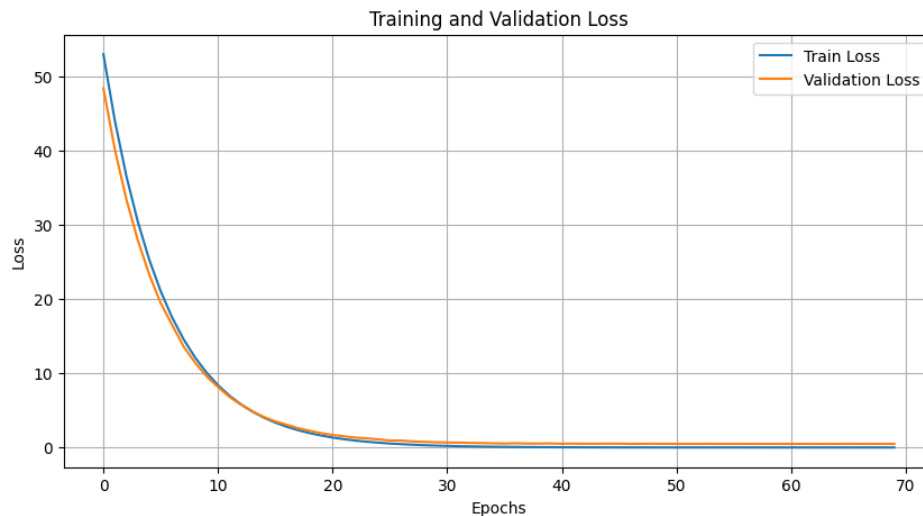
Alto *recall* (0,970): Este modelo es excelente para detectar casos “malignos”, perdiendo solo el 3,0% de ellos. Es ideal en situaciones donde minimizar los falsos negativos es prioritario. F1-Score alto (0,855): esto demuestra un equilibrio razonable entre la precisión y el *recall*, lo que lo convierte en un modelo sólido para la clasificación binaria. Buen *accuracy* (0,824): el modelo logra clasificar correctamente la mayoría de los ejemplos.

Debilidades:

Baja especificidad (0,656): La capacidad del modelo para clasificar correctamente los casos “benignos” es limitada. Esto puede ser problemático si los falsos positivos generan costos significativos. Moderada tasa de falsos positivos (FPR ~ 34,4%): un tercio de los casos “benignos” son mal clasificados como “malignos”, lo que podría resultar en evaluaciones o intervenciones adicionales innecesarias.

Figura 39

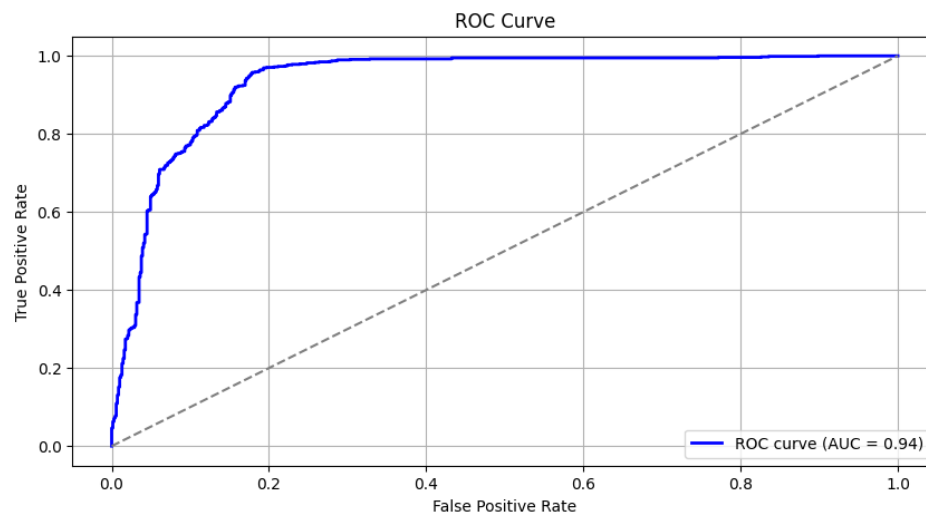
Gráfico de Loss del modelo InceptionV3.



Nota. Fuente: Autoría propia.

Figura 40

Gráfico de ROC-AUC de Validación del modelo InceptionV3.



Nota. Fuente: Autoría propia.

d) **Hiperparámetros:**

Tabla 7

Hiperparámetros de la Arquitectura Adicional a InceptionV3

Hiperparámetro	Valor
Epochs	100
Batch Size	{8; 16; 32* ; 64}
Dense_1	{8; 16; 32; 64; 96; 128; 256; 512; 1024; 2048; 4096* }
Regularización Dense_1	{ 1e-2* ; 1e-3; 1e-4}
Función Activación Dense_1	{ ReLU* ; Swish}
Dense_2	{8; 16; 32; 64; 96; 128; 256; 512; 1024* ; 2048; 4096}
Regularización Dense_2	{ 1e-2* ; 1e-3; 1e-4}
Función Activación Dense_2	{ ReLU* ; Swish}
Dropout_1	{0,1; 0,2; 0,3; 0,4* ; 0,5; 0,6}
Regularización Output_Layer	{1e-2; 1e-3; 1e-4* }
Función Activación Output_Layer	{ Sigmoide* ; Tanh Centrada; Softmax}
Optimizador	{SGD, Adam* }
Learning Rate (Max)	{1e-4; 9e-5; 8,7e-5* ; 8e-5; 7,5e-5; 7e-5; 6e-5; 5e-5; 4e-5; 3e-5; 2e-5; 1e-5; 1e-6; 1e-7; 1e-8; 1e-9}
Weight Decay	None (Default)
ReduceLRonPlateau	
Learning Rate (Min)	1e-7
Paciencia	3
Factor	0,5
EarlyStopping	
Paciencia	7

4. Ensamble

a) Tabla Métricas

Tabla 8

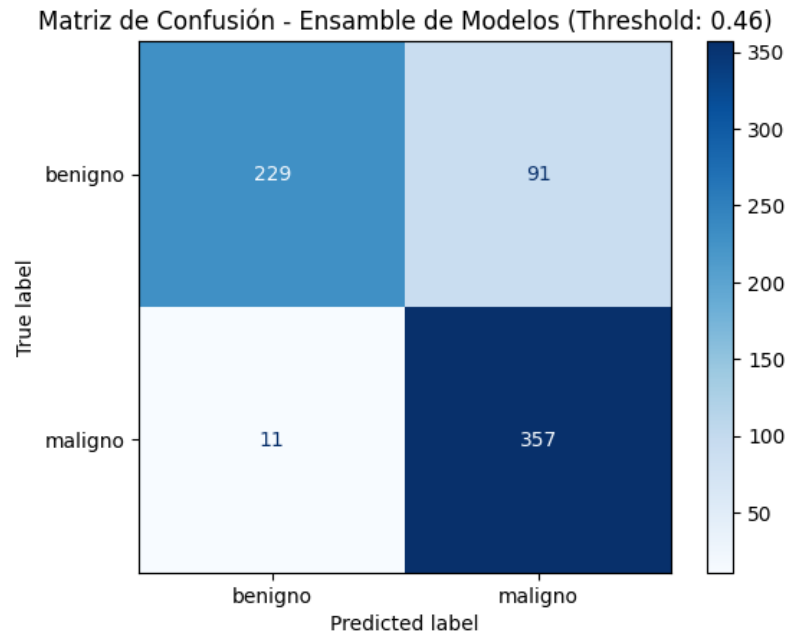
Métricas del ensamble en conjunto Test

Métrica	Valor
Threshold óptimo	0,46
Accuracy	0,8575
Precision	0,8068
Recall	0,9646
F1-Score	0,8787
ROC-AUC	0,8842

b) Matriz de Confusión

Figura 41

Matriz de confusión del ensamble.



Nota. Fuente: Autoría propia.

Tabla 9

Tabla Resumen de comparación de métricas entre modelos individuales y ensamble.

Métrica	Xception	ResNet50	InceptionV3	Ensamble
Threshold óptimo	0,44	0,41	0,32	0,46
Accuracy	0,8241	0,8503	0,8241	0,8575
Precision	0,7852	0,8018	0,7645	0,8068
Recall	0,9239	0,9565	0,9701	0,9646
F1-Score	0,8489	0,8724	0,8551	0,8787
ROC-AUC	0,8566	0,8597	0,8673	0,8842

c) Observaciones

Si bien InceptionV3 presenta el mayor *recall* (0,9701), lo que indica una alta sensibilidad para detectar casos malignos, su correspondiente *precision* es la más baja (0,7645), lo que sugiere una mayor tasa de falsos positivos. Por otro lado, ResNet50 logra un mejor balance entre *precision* (0,8018) y *recall* (0,9565), resultando en un F1-Score competitivo (0,8724). Xception, aunque con un rendimiento sólido, queda por debajo de estos modelos en algunas métricas clave. Finalmente, el ensamble demuestra ser la mejor estrategia al mejorar la estabilidad de la predicción, alcanzando el mayor *precision* (0,8068), un *recall* elevado (0,9646), el mejor F1-Score (0,8787) y el mayor AUC-ROC (0,8842), lo que evidencia su capacidad para reducir errores individuales y mejorar la fiabilidad del diagnóstico.

Fortalezas:

El ensamble de modelos presenta limitaciones como un *precision* de 0,8068, reflejando una cantidad significativa de falsos positivos (91 “benignos” clasificados como “malignos”), lo que podría generar evaluaciones innecesarias y mayores costos. Además, su implementación implica un alto costo computacional y mayor tiempo de inferencia. No obstante, mantiene un equilibrio entre *precision* y *recall* con un F1-score de 0,8787 y un AUC-ROC de 0,8842, lo que indica una alta capacidad discriminativa. Además, al combinar múltiples modelos, se mitigan sesgos individuales, logrando un sistema más robusto y confiable en el ámbito clínico.

Debilidades:

A pesar de sus ventajas, el ensamble de modelos presenta limitaciones, como una precisión de 0,8068, lo que refleja una cantidad significativa de falsos positivos (91 “benignos” clasificados como “malignos”). Esto podría generar evaluaciones médicas innecesarias y aumentar los costos de diagnóstico. Además, su implementación exige un alto costo computacional, prolongando el tiempo de inferencia y requiriendo mayores recursos.

4.1.2. EJEMPLOS PREDECIDOS

En esta sección se presentan dos ejemplos representativos de cada etiqueta (benigno y maligno) para ilustrar un esbozo del sistema propuesto. Cada ejemplo incluye la etiqueta real de la lesión, las etiquetas predecidas por los tres modelos (Xception, ResNet50 e InceptionV3) y el ensamble, lo que permite evaluar la concordancia entre las predicciones y la realidad.

Figura 42

Imágenes con etiqueta "benigno"

(a) Ejemplo Benigno 1

Predicciones
Real: Benigno
Xception: 0.01 (Benigno)
ResNet: 0.00 (Benigno)
Inception: 0.01 (Benigno)
Ensamble: 0.01 (Benigno)



(b) Ejemplo Benigno 2

Predicciones
Real: Benigno
Xception: 0.02 (Benigno)
ResNet: 0.00 (Benigno)
Inception: 0.11 (Benigno)
Ensamble: 0.04 (Benigno)



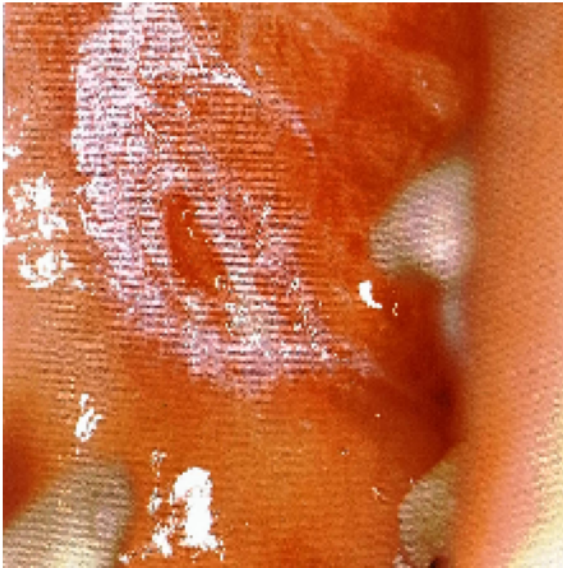
Nota. Obtenido de *Datasets* anteriores y predicción de autoría propia.

Figura 43

Imágenes con etiqueta "maligno"

(a) Ejemplo Maligno 1

Predicciones
 Real: Maligno
 Xception: 0.98 (Maligno)
 ResNet: 1.00 (Maligno)
 Inception: 1.00 (Maligno)
 Ensamble: 0.99 (Maligno)



(b) Ejemplo Maligno 2

Predicciones
 Real: Maligno
 Xception: 0.91 (Maligno)
 ResNet: 1.00 (Maligno)
 Inception: 0.92 (Maligno)
 Ensamble: 0.94 (Maligno)



Nota. Obtenido de *datasets* anteriores y predicción de autoría propia.

4.2. COMPARACIÓN DE DESEMPEÑO

4.2.1. MÉTODOS TRADICIONALES - CLÍNICOS

Ahora bien, tomando en cuenta lo señalado por DÍAZ-CABALLERO *et al.* (2021), quienes analizan y cuantifican la efectividad de los métodos clásicos para la detección del cáncer oral, estos han sido evaluados y sus métricas se presentan de la siguiente manera:

Tabla 10

Métricas de los Métodos Tradicionales para la Detección de Lesiones Orales Potencialmente Malignas.

Método	Sensibilidad (%)	Especificidad (%)
ViziLite	81,31	25,40
Microlux DL	82,63	69,52
Azul de Toluidina	82,07	66,27
Citología Exfoliativa ²⁴	94,60	99,50

Contrastamos con las métricas obtenidas de los modelos construidos:

Tabla 11

Métricas de los Modelos Propuestos.

Modelo	Sensibilidad (%)	Especificidad (%)
Xception	92,39	70,95
ResNet50	95,65	72,80
InceptionV3	97,01	65,63
Ensamble	98,09	79,52

1. Contexto del Uso de los Métodos

■ Métodos Tradicionales:

Accesibilidad: métodos como el ViziLite y el Microlux DL son ampliamente accesibles y económicos, pero tienen limitaciones respecto a la sensibilidad y especificidad.

Invasividad: la citología exfoliativa requiere la recolección de muestras, un procedimiento que puede ser más invasivo y altamente dependiente de la habilidad del profesional que realiza la toma. Además, su efectividad está limitada por la calidad de la muestra obtenida, la técnica empleada y la infraestructura disponible para su procesamiento y análisis (Acha *et al.*, 2005). Estas limitaciones pueden llevar a resultados inconsistentes, como falsos negativos, particularmente si las células recolectadas no son representativas de la lesión analizada.

- **Modelos Propuestos: Escalabilidad:** los modelos pueden procesar imágenes con herramientas digitales estándar, lo que los hace fácilmente escalables en entornos con acceso a imágenes médicas (integración a sistemas médicos internos y bases de datos).

Automatización: se ofrece la ventaja de diagnósticos automatizados, reduciendo el tiempo de análisis y provee herramientas para los profesionales que realizan el diagnóstico de estas patologías.

²⁴Este procedimiento consiste en el raspado del área a analizar y el tejido recolectado se estudia sobre una placa de vidrio. Posteriormente se utiliza un tinte en la muestra para poder ver las células con claridad. Si alguna de las células tiene una apariencia anormal, se puede hacer una biopsia del área. (Society, 2021)

2. Fortalezas y Limitaciones Específicas

■ Métodos Tradicionales:

Fortalezas:

la citología exfoliativa tiene alta especificidad (99,50 %), lo que reduce los falsos positivos.

Métodos como el Microlux DL son portátiles y relativamente económicos.

Limitaciones:

ViziLite tiene especificidad extremadamente baja (25,40 %), lo que genera muchos falsos positivos, haciendo su aplicación clínica limitada y poco confiable.

Todo proceso de tinción dependerá en gran medida de la minuciosidad de que se realicen todos los pasos adecuadamente, el tiempo adecuado de tinción de los colorantes y tratamiento adecuado de estos (sin olvidar mencionar que muchos de estos colorantes son extremadamente volátiles) (Sanchís Bielsa *et al.*, 1990). La citología exfoliativa presenta limitaciones en la identificación de lesiones sub-mucosas o en estadios iniciales, lo que puede resultar en diagnósticos incompletos o falsos negativos.

■ Modelos Propuestos:

Fortalezas:

Sensibilidad superior en todos los modelos (hasta 98,09 % en el ensamble), asegurando una alta detección de casos malignos y una especificidad comparable a los métodos tradicionales más confiables, como el Microlux DL.

Limitaciones:

Aunque automatizados, los modelos requieren imágenes de alta calidad y datos diversos para garantizar su robustez en distintos escenarios clínicos. Además, se requiere una infraestructura digital robusta (*hardware* y *software*), lo que podría dificultar su aplicación en áreas con recursos limitados.

3. Impacto en el Paciente

■ Métodos Tradicionales:

Efectos positivos: métodos accesibles como Azul de Toluidina permiten identificar rápidamente áreas sospechosas.

Efectos negativos: métodos con baja especificidad, como ViziLite, pueden generar falsas alarmas, causando ansiedad innecesaria en los pacientes y aumentando los costos de tiempo y económicos en pruebas adicionales.

■ Modelos Propuestos:

Efectos positivos: alta sensibilidad reduce la posibilidad de pasar por alto lesiones malignas, crucial para detección temprana, un tratamiento a tiempo y adecuado y un incremento de las probabilidades de supervivencia del paciente. Además, el enfoque no invasivo minimiza molestias físicas y psicológicas para el paciente.

Efectos negativos: los falsos positivos debido a especificidades más bajas (ejemplo: InceptionV3 con 65,63 %) podrían llevar a pruebas innecesarias y estrés en el paciente.

4. Comparación Económica y Técnica:

Tabla 12

Comparación Económica y Técnica entre Métodos Tradicionales y Modelos Basados en Aprendizaje Profundo.

Aspecto	Métodos Tradicionales	Modelos Propuestos
Costo de Implementación	Generalmente bajo. Métodos como Azul de Toluidina y Microlux DL requieren insumos básicos y son económicamente accesibles.	Requieren inversión inicial en infraestructura (hardware y software), pero los costos marginales de diagnóstico son bajos a gran escala.
Costo por Uso	Insumos como tintes y luces específicas deben reponerse periódicamente.	Una vez entrenados, el costo por diagnóstico es insignificante en comparación.
Dependencia de Operadores	Requieren personal capacitado para aplicar los métodos y analizar los resultados (e.g., citología exfoliativa).	Diagnósticos automatizados reducen la dependencia de operadores humanos, minimizando la variabilidad entre usuarios.
Infraestructura Necesaria	Equipos básicos como luces de fluorescencia (Microlux DL, ViziLite) y reactivos químicos.	Computadoras con capacidad para manejar modelos preentrenados e imágenes médicas digitales.
Mantenimiento	Requieren reemplazo periódico de insumos físicos y calibración básica del equipo.	Actualizaciones de software y mantenimiento del hardware para garantizar precisión y robustez.

Con el fin de mantener una rigurosidad en este trabajo, es necesario añadir la comparación con otros proyectos similares que aborden el tema (ya sea con distintos modelos, enfoque de los datos, volumen de datos, etc.).

4.2.2. MÉTODOS BASADOS EN MACHINE LEARNING

Para evaluar el desempeño del modelo propuesto en la detección de cáncer oral, se realizó una comparación con estudios previos que han abordado este problema utilizando técnicas de *Machine Learning*. En particular, se contrastaron los resultados con los obtenidos en los trabajos de Aubreville *et al.* (2017), Shamim *et al.* (2019) y Koriakina *et al.* (2022). Estos estudios fueron seleccionados por su relevancia en el área, su aplicación de CNNs y el uso de diferentes tipos de datos para el diagnóstico del cáncer oral.

A continuación, se presenta un análisis comparativo detallado con cada uno de estos enfoques, destacando diferencias en términos de metodología, tipo de imágenes utilizadas, volumen de datos y métricas de desempeño.

1. Comparación con Aubreville *et al.* (2017)

- a) **Diferencias en datos:** mientras que Aubreville *et al.* (2017) utilizaron imágenes CLE, el ensamble trabaja con imágenes fotográficas RGB de diversas áreas de la cavidad bucal, lo que permite una mayor aplicabilidad en entornos clínicos convencionales.
- b) **Diferencias en modelos:** se usó una arquitectura más moderna basada en un ensamble de modelos (Xception, ResNet50 e InceptionV3), en contraste con la red LeNet-5 empleada en el paper.
- c) **Métricas comparativas:**
Aubreville *et al.* (2017): Accuracy = 88,3 %, Sensibilidad = 86,6 %, Especificidad = 90,0 %.
Modelo propuesto: Accuracy = 85,75 %, Sensibilidad = 96,46 %, Especificidad = 80,68 %.
- d) **Observaciones:** si bien la precisión del modelo de Aubreville *et al.* (2017) es ligeramente mayor, el modelo propuesto logra una mayor sensibilidad, lo que es clave en un problema de detección temprana de cáncer oral.

Tabla 13
Comparación entre tus modelos y Aubreville et al. (2017)

Aspecto	Tus Modelos	Aubreville et al. (2017)
Arquitectura Base	Xception, ResNet50, InceptionV3	CNN basado en LeNet-5
Capas Adicionales	Self-Attention, Dense Layers	Fusión de patches (80x80 px)
Imágenes Usadas	Fotográficas (RGB, Clínicas)	Confocal Laser Endomicroscopy
Dataset Tamaño	9507 imágenes	7894 imágenes
Métricas Clave	Xception: Acc = 82,41 %, Sens = 92,39 %, Espec = 78,52 % ResNet50: Acc = 84,45 %, Sens = 95,38 %, Espec = 79,59 % InceptionV3: Acc = 82,41 %, Sens = 97,01 %, Espec = 76,45 % Ensamble: Acc = 85,75 %, Sens = 96,46 %, Espec = 80,68 %	Sens = 86,6 %, Espec = 90 % Acc = 88,3 %
Imágenes Preprocesadas	Rescalado a 224x224 px	Recorte por patches
Principal Limitación	Validación externa necesaria	Variabilidad en calidad

2. Comparación con Shamim et al. (2019)

- a) **Diferencias en datos:** El modelo de Shamim et al. (2019) se limitó a imágenes de la lengua, mientras que el ensamble propuesto abarca una diversidad de imágenes de lengua, mejillas, paladar y labios, lo que mejora la generalización.
- b) **Diferencias en modelos:** Aunque ambos estudios utilizan ResNet50, este trabajo emplea un ensamble de modelos que incorpora Xception e InceptionV3 junto con técnicas avanzadas como *self-attention*.
- c) **Métricas comparativas:**
 Shamim et al. (2019): Accuracy = 98 %, Sensibilidad = 89 %, Especificidad = 97 %.
 Modelo propuesto: Accuracy = 85,75 %, Sensibilidad = 96,46 %, Especificidad = 80,68 %.
- d) **Observaciones:** Aunque el modelo de Shamim et al. (2019) muestra una mayor precisión, su bajo número de imágenes hace que sus resultados sean menos generalizables. En contraste, nuestro modelo ha sido entrenado con 9507 imágenes, lo que lo hace más robusto en distintos escenarios clínicos.

Tabla 14
 Comparación entre tus modelos y Shamim et al. (2019)

Aspecto	Tus Modelos	Shamim et al. (2019)
Arquitectura Base	Xception, ResNet50, InceptionV3	VGG19 y ResNet50
Capas Adicionales	Self-Attention, Dense Layers	Ajuste final en Dense
Imágenes Usadas	Fotográficas (RGB, Clínicas, de diversas áreas de la cavidad bucal como lengua, mejillas, paladar, labios, etc.)	Fotográficas de lengua
Dataset Tamaño	9507 imágenes	200-300 imágenes
Métricas Clave	Xception: Acc = 82,41 %, Sens = 92,39 %, Espec = 78,52 % ResNet50: Acc = 84,45 %, Sens = 95,38 %, Espec = 79,59 % InceptionV3: Acc = 82,41 %, Sens = 97,01 %, Espec = 76,45 % Ensamble: Acc = 85,75 %, Sens = 96,46 %, Espec = 80,68 %	Acc = 98 %, Sens = 89 %, Espec = 97 %
Imágenes Preprocesadas	Rescalado a 224x224 px	Recorte manual
Principal Limitación	Validación externa necesaria	Pequeño dataset

3. Comparación con Koriakina *et al.* (2022)

- a) **Diferencias en datos:** Koriakina *et al.* (2022) trabajaron con muestras celulares (citologías), mientras que este trabajo se centra en imágenes fotográficas clínicas, lo que hace que los enfoques sean complementarios en lugar de comparables directamente.
- b) **Diferencias en modelos:** Se usó ResNet18, en contraste al ensamble CNN y uso de *self-attention*.
- c) **Métricas comparativas:**
Koriakina *et al.* (2022): Accuracy a nivel celular = 91.4 % con ResNet18. Modelo propuesto: Accuracy = 85,75 %, F1-score = 0,878; AUC-ROC = 0,87.
- d) **Observaciones:** Aunque el enfoque de Koriakina *et al.* (2022) ofrece una alta precisión en citologías, su aplicabilidad a imágenes clínicas es limitada. Además, el uso de datos sintéticos podría afectar su capacidad de generalización en entornos reales.

Tabla 15

Comparación entre tus modelos y Koriakina *et al.* (2022)

Aspecto	Tus Modelos	Koriakina <i>et al.</i> (2022)
Arquitectura Base	Xception, ResNet50, InceptionV3	Transfer Learning CNN
Capas Adicionales	Self-Attention, Dense Layers	<i>fine-tuning</i> en capa densa
Imágenes Usadas	Fotográficas (RGB, Clínicas)	Histología oral (PAP smear) ²⁵
Dataset Tamaño	9507 imágenes	No especificado
Métricas Clave	Xception: Acc = 82,41 %, ROC-AUC = 85,66 % ResNet50: Acc = 84,45 %, ROC-AUC = 88,38 % InceptionV3: Acc = 82,41 %, ROC-AUC = 86,72 % Ensamble: Acc = 85,75 %, ROC-AUC = 88,42 %	Acc = 91,4 %, ROC-AUC = 0,87
Imágenes Preprocesadas	Rescalado a 224x224 px	Down-sampling
Principal Limitación	Validación externa necesaria	Difícil interpretabilidad

CAPÍTULO 5

CONCLUSIONES

El presente trabajo demuestra que el aprendizaje automático puede ser una herramienta valiosa para asistir en la detección temprana del cáncer oral, abordando una problemática de alta relevancia en el ámbito médico. El estudio inició con una breve explicación de algunos conceptos importantes para el consiguiente análisis del estado del arte respecto a técnicas de aprendizaje automático aplicadas a la detección de cáncer oral, tal como se planteaba en el Objetivo Específico 1, lo que permitió comprender los avances realizados en el campo y las limitaciones presentes en métodos tradicionales y en enfoques basados en *Machine Learning*, lo que permitió construir la base de la investigación realizada en el presente trabajo.

A partir de este análisis, se identificaron aspectos clave del problema, como las señales indicadoras de cáncer en lesiones orales y los desafíos asociados a la correcta clasificación de imágenes médicas, siendo esto lo mencionado en el Objetivo Específico 2. Estos hallazgos sirvieron como base para la recopilación y preprocesamiento de un conjunto de datos compuesto por 9507 imágenes fotográficas clínicas de diversas áreas de la cavidad bucal (lengua, mejillas, paladar, labios, entre otras), siendo esto lo correspondiente a realizar conforme al Objetivo Específico 3. Este *dataset*, cuidadosamente curado y preprocesado, aseguró la calidad necesaria para entrenar modelos confiables.

Posteriormente, se diseñaron y desarrollaron modelos de redes neuronales convolucionales optimizados, basados en arquitecturas preentrenadas como Xception, ResNet50 e InceptionV3, cuyo factor destacable respecto a otros trabajos es la integración de arquitecturas avanzadas con mecanismos de *self-attention* y capas densas, con el objetivo de mejorar la captura de características, tal como se formuló realizar conforme al Objetivo Específico 4. Junto con ello, se implementó un enfoque de ensamble para combinar las predicciones de los diferentes modelos, logrando métricas destacables como un F1-score de 0,878 y un AUC-ROC de 0,87, lo que permitió lograr métricas competitivas que superan a métodos tradicionales y otros trabajos del estado del arte, demostrando un equilibrio adecuado entre precisión y sensibilidad, lo que optimiza el diagnóstico temprano y reduce el tiempo necesario para identificar casos críticos, tal como se mencionaba en el Objetivo Específico 5.

²⁵El PAP smear oral es un método que analiza muestras teñidas de tejido oral al microscopio para detectar anomalías celulares.

Sin embargo, el desarrollo de proyectos como este enfrenta diversas limitaciones inherentes a la naturaleza de los datos médicos, desafíos técnicos, complejidades legales y consideraciones éticas. Por ejemplo, uno de los principales retos fue la confidencialidad de los datos de los pacientes, la cual se manejó mediante el uso de *datasets* recopilados de fuentes éticamente aprobadas, con consentimiento voluntario y en colaboración con instituciones médicas. Aunque estas medidas cumplen con los estándares legales, sigue siendo fundamental implementar sistemas más avanzados de anonimización y protección de datos, especialmente en futuros trabajos que impliquen la recolección de nuevas muestras.

Además, se destacaron limitaciones relacionadas con la calidad y el volumen de las imágenes disponibles. Aunque se dispuso de un dataset amplio, las imágenes presentaban variabilidad en resolución y características, lo que podría afectar la capacidad de los modelos para generalizar en entornos clínicos reales. Este problema se agrava por la falta de un grupo de control de pacientes sanos, lo que puede introducir sesgos significativos en la clasificación al asociar cualquier tipo de lesión con malignidad. Por ello, es fundamental ampliar la diversidad de los datos recopilados para garantizar que los modelos sean robustos y equitativos.

Asimismo, un aspecto crucial que debe ser considerado en futuras aplicaciones es la interpretabilidad de los modelos desarrollados. La capacidad de explicar de manera clara y comprensible cómo se toman las decisiones es esencial para que los profesionales médicos puedan confiar en estas herramientas y utilizarlas eficazmente en la práctica clínica.

Se debe tomar en cuenta que cualquier avance en este campo debe estar guiado por principios éticos sólidos. El uso de inteligencia artificial en la detección del cáncer oral debe priorizar la transparencia, minimizar sesgos perjudiciales y enfocarse en mejorar la calidad de vida del paciente, siempre respetando la confidencialidad y los derechos de las personas involucradas. Esto permitirá que las herramientas desarrolladas no solo sean efectivas, sino también aceptables y responsables en su implementación práctica.

De cara al futuro, estas complicaciones deberán abordarse en trabajos futuros, especialmente en la obtención de datos más diversos y de alta calidad, la integración de los modelos en sistemas clínicos existentes, y el reentrenamiento continuo con datos nuevos para garantizar la vigencia y precisión de los modelos, como también se podría explorar la integración de datos multimodales, como información genómica o clínica, para mejorar la capacidad del modelo de identificar subtipos de cáncer oral. Además, se podría evaluar el uso de la segmentación automática para delimitar áreas de interés en las imágenes, incrementando la precisión y facilitando la interpretación por parte de médicos. Sin embargo, el presente trabajo sienta un marco metodológico sólido para futuras investigaciones, demostrando que es posible desarrollar herramientas tecnológicas basadas en *Machine Learning* que no solo apoyen a los profesionales médicos en el diagnóstico temprano del cáncer oral, sino que además favorezcan el bienestar y la calidad de vida de los pacientes.

REFERENCIAS BIBLIOGRÁFICAS

Acha, A., Ruesga, M. T., Rodríguez, M. J., Martínez de Pancorbo, M. A., y Aguirre, J. M. (2005). Aplicaciones de la citología oral por raspado (exfoliativa) en el cáncer y precáncer oral. *Medicina Oral, Patología Oral y Cirugía Bucal (Ed. impresa)*, 10:95 – 102.

Antonio D. et al., Laura A., C. R. (2019). Análisis de los métodos de detección de lesiones orales potencialmente malignas - revisión sistemática tipo umbrella. *FACULTAD DE ODONTOLOGÍA*.

Aubreville, M., Knipfer, C., Oetter, N., Jaremenko, C., Rodner, E., Denzler, J., Bohr, C., Neumann, H., Stelzle, F., y Maier, A. (2017). Automatic classification of cancerous tissue in laserendomicroscopy images of the oral cavity using deep learning. *Scientific Reports*, 7(1).

AWS (2024). ¿Qué es una red neuronal? - Explicación de las redes neuronales artificiales - AWS.

Baeza, A. (2021). Académico U. de Chile investiga nuevos tratamientos para el cáncer oral - Universidad de Chile 2021.

Bits (2023). final oral cancer merge dataset. <https://universe.roboflow.com/bits-xn4un/final-oral-cancer-merge>.

Cancer Research (2024). Oralcancer2 dataset. https://universe.roboflow.com/cancerresearch/oral_cancer2.

Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions.

Colegio Dentistas, C. (2020). En el Día Latinoamericano de la Lucha contra el Cáncer Bucal recordamos que es clave la prevención y detección temprana.

Cooperativa.cl (2020). Minsal informó preocupantes cifras de incidencia de cáncer bucal. *cooperativa.cl*.

DeMarco, C. (2022). Canker sore vs. oral cancer: How can you tell the difference?

DÍAZ-CABALLERO, A., ANAYA-GUZMÁN, L., y ROJANO-BUELVAS, C. (2021). Analysis of the detection methods for potentially malignant oral lesions. An umbrella systematic review. *Acta Medica Colombiana*, 46:27 – 33.

Escribano-Bermejo M. et al, B.-M. A. (2009). Leucoplasia oral: Conceptos actuales. *SciELO Analytics*, 25(2).

Estrada Pereira, G. A., Márquez Filiú, M., y González Heredia, E. (2012). Incidencia de la eritroleucoplasia en la cavidad bucal. *SciELO Analytics*, 16(11).

Fernandez, P. (2022). Las alarmantes cifras del cáncer en Chile: Hay 45 mil nuevos casos cada año. *24horas.cl*.

Framejframej_1 (2023). Oral Lesion Dataset. <https://universe.roboflow.com/framejframej-1-krvqf/oral-lesion-s57sl>.

Ghosh, M. y Arunachalam, T. (2021). *Introduction to Artificial Intelligence*, pp. 23-44.

Gobierno de Chile (2023). Política nacional de inteligencia artificial.

Gu, W., Bai, S., y Kong, L. (2022). A review on 2d instance segmentation based on deep neural networks. *Image and Vision Computing*, 120:104401.

Haridas (2024). Oral_cancer dataset. https://universe.roboflow.com/haridas/oral_cancer-c4fqt.

Hoffer, E., Banner, R., Golan, I., y Soudry, D. (2018). Norm matters: efficient and accurate normalization schemes in deep networks. NIPS'18, p. 2164-2174, Red Hook, NY, USA. Curran Associates Inc.

Huber, M. A. (2018). Adjunctive diagnostic techniques for oral and oropharyngeal cancer discovery. *Dental Clinics of North America*, 62(1):59-75. Oral Cancer.

Kean, U. K. (2024). 2main project dataset. <https://universe.roboflow.com/khon-kean-university-xqvix/2main-project>.

Koriakina, N., Sladoje, N., Bašić, V., y Lindblad, J. (2022). Oral cancer detection and interpretation: Deep multiple instance learning versus conventional deep single instance learning.

Laarhoven, T. V. (2017). L2 regularization versus batch and weight normalization.

Lucena, C. J. y Porras D., H. (2006). Sistemas de apoyo a las decisiones: génesis y evolución. *REVISTA DE LA FACULTAD DE INGENIERÍAS FISI-COME-CÁNICA*5, 5(1):11-24.

López-Jornet, P., Saura-Inglés, A., y Cozar-Fernández, A. (1999). Estudio de las lesiones precancerosas de la mucosa bucal en el paciente geriátrico. *reGG - Elsevier*, 34(3).

Mahjoubi, M. A., Hamida, S., Siani, L. E., Cherradi, B., Abbassi, A. E., y Raihani, A. (2023). Deep learning for cerebral hemorrhage detection and classification in head ct scans using cnn. *2023 3rd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)*, pp. 1-8.

Maraboli-Contreras, S., Adorno-Farias, D., Maturana-Ramirez, A., Rojas-Alcayaga, G., Fuentes-Alburquenque, M., y Espinoza-Santander, I. (2018). Sobrevida de carcinoma oral de células escamosas: reporte de la universidad de chile. *Revista clínica de periodoncia, implantología y rehabilitación oral*, 11(3):147-151.

Martínez-Sanguinetti, M. A., Parra-Soto, S., Petermann-Rocha, F., Leiva-Ordeñez, A. M., Troncoso-Pantoja, C., Ulloa, N., Diaz-Martínez, X., y Celis-Morales, C. (2020). Cáncer en Chile y en el mundo: una mirada actual y su futuro escenario epidemiológico. *Rev. méd. Chile*, 148(10).

Mateo-Sidrón, A. M. C. y Somacarrera-Pérez, M. L. (2015). Cáncer oral: Genética, prevención, diagnóstico y tratamiento. revisión de la literatura. *SciELO Analytics*, 31(4).

Navarro, B. y Guzmán, M. (2024). Ganglios linfáticos cervicales.

Project, G. (2023). tooth lesions dataset. <https://universe.roboflow.com/graduation-project-pafhg/tooth-lesions>.

Quiza, R. y Davim, J. P. (2011). Computational methods and optimization. *Machining of hard materials*, pp. 177-208.

Rangel-Rivera, J. C., Lauzardo-García del Prado, G., Quintana-Castillo, M., Gutiérrez-Hernández, M. E., y Gutiérrez-Hernández, N. (2009). Necesidad de crear programas de promoción y prevención en el adulto mayor. *Revista Cubana de Estomatología*, 46:0 - 0.

Russell, S. y Norvig, P. (2010). *Artificial Intelligence: A Modern Approach*. Prentice Hall, 3 edición.

Sanchís Bielsa, J. M., Silvestre Donat, F. J., y Peñarrocha Diago, M. (1990). La citología exfoliativa oral: un método diagnóstico en estomatología. *Oris*, 40(3):71-82.

Santelices, M. J., Cárcamo, M., Brenner, C., y Montes, R. (2016). Cáncer oral en Chile. revisión de la literatura. *Rev. méd. Chile*, 144(6).

Shahinfar, S., Meek, P., y Falzon, G. (2020). "how many images do i need?" understanding how sample size per class affects deep learning model performance metrics for balanced designs in autonomous wildlife monitoring. *Ecological Informatics*, 57:101085.

Shamim, M. Z. M., Sadatullah, S., Shiblee, M., Usman, M., y Ali, S. K. (2019). Automated detection of oral pre-cancerous tongue lesions using deep learning for early diagnosis of oral cavity cancer (accepted - oxford university press: Computer journal).

Simonyan, K. y Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition.

Smith, L. N. (2018). A disciplined approach to neural network hyper-parameters: Part 1 - learning rate, batch size, momentum, and weight decay.

Society, A. C. (2021). Pruebas para los cánceres de cavidad oral (boca) y de orofaringe (garganta).

Stillfried, A., Rocha, A., Colella, G., y Escobar, E. (2016). Cáncer Oral y Dentistas: Conocimientos, Actitudes y Prácticas en Chile. *International journal of odontostomatology*, 10:521 - 529.

Suárez-Obando, F. y Ordóñez Vásquez, A. (2012). Aspectos éticos de la informática médica: principios de uso y usuario apropiado de sistemas computacionales en la atención clínica. *SciELO Analytics*, 18(2):199-208.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., y Wojna, Z. (2015). Rethinking the inception architecture for computer vision.

UNAB, C. O. (2023). Oralcancerunab dataset. <https://universe.roboflow.com/cancer-oral-unab/oralcancerunab>.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., y Polosukhin, I. (2023). Attention is all you need.

Warin, K., Limprasert, W., Suebnukarn, S., Jinaporntham, S., y Jantana, P. (2021). Automatic classification and detection of oral cancer in photographic images using deep learning algorithms. *Journal of Oral Pathology & Medicine*, 50(9):911–918.

Yang E et al., . (2018). Noninvasive diagnostic adjuncts for the evaluation of potentially premalignant oral epithelial lesions: current limitations and future directions. *Oral Surg Oral Med Oral Pathol Oral Radiol.*, 125(6):670–681.

Yuk, C. (2024). Ceklidah dataset. <https://universe.roboflow.com/ceklidah-yuk/ceklidah>.

Úrquiza, M. P. (2010). Endomicroscopia confocal, ¿debemos cambiar nuestra aproximación al diagnóstico endoscópico de las lesiones mucosas del tracto gastrointestinal? *Elsevier*, 33:267–270.