

ANEXO
N°45
INFORME
FINAL



Serie Proyectos de Investigación e Innovación
Superintendencia de
Seguridad Social Santiago
- Chile

SISTEMA DE INTELIGENCIA ARTIFICIAL PARA LA DETECCIÓN DE
NEUMOCONIOSIS
INFORME FINAL

Autor:
Carolina Maldonado Ortega

Año publicación 2022



SUPERINTENDENCIA DE SEGURIDAD SOCIAL SUPERINTENDENCE OF SOCIAL SECURITY

La serie Proyectos de Investigación e Innovación corresponde a una línea de publicaciones de la Superintendencia de Seguridad Social, que tiene por objetivo divulgar los trabajos de investigación e innovación en Prevención de Accidentes y Enfermedades del Trabajo financiados por los recursos del Seguro Social de la Ley 16.744.

Los trabajos aquí publicados son los informes finales y están disponibles para su conocimiento y uso. Los contenidos, análisis y conclusiones expresados son de exclusiva responsabilidad de su(s) autor(es), y no reflejan necesariamente la opinión de la Superintendencia de Seguridad Social.

Si requiere de mayor información, sobre el estudio o proyecto escriba a: investigaciones@suseso.cl.

Si desea conocer otras publicaciones, artículos de investigación y proyectos de la Superintendencia de Seguridad Social, visite nuestro sitio web: www.suseso.cl.

The Research and Innovation Projects series corresponds to a line of publications of the Superintendencia of Social Security, which aims to disseminate the research and innovation work in the Prevention of Occupational Accidents and Illnesses financed by the resources of Law Insurance 16,744.

The papers published here are the final reports and are available for your knowledge and use. The content, analysis and conclusions are solely the responsibility of the author (s), and do not necessarily reflect the opinion of the Superintendencia of Social Security.

For further information, please write to: investigaciones@suseso.cl.

For other publications, research papers and projects of the Superintendencia of Social Security, please visit our website: www.suseso.cl.

Superintendencia de
Seguridad Social Huérfanos
1376 Santiago,
Chile.



INFORME FINAL: SISTEMA DE INTELIGENCIA ARTIFICIAL PARA LA DETECCIÓN DE NEUMOCONIOSIS

SUBGERENCIA DE INNOVACIÓN E INVESTIGACIÓN

PREPARADO POR:

Innovadora Principal: Carolina Maldonado.

Subgerencia de Innovación e Investigación.

Innovadora Alterna: Carmen Archila.

Dirección Médica Salud Ocupacional.

CLIENTE INTERNO:

Dirección Médica de Salud Ocupacional

EQUIPO EXTERNO:

Carolina Andrea Arellano Castro.

Eduardo Renán Ruiz Retamal.

GERENCIA CORPORATIVA DE SST

NOVIEMBRE 2022

CONTEXTO

- Marco teórico:

Dentro de las principales funciones de las Mutualidades se encuentra otorgar prestaciones preventivas, que contempla evaluaciones de medicina preventiva en acciones de promoción y prevención de salud como son los exámenes ocupacionales, pre ocupacionales y Programas de vigilancia por exposición a agentes neumoconiogenos, dichas funciones se encuentran definidas en la Ley 16.744 que indica que cada mutualidad es responsable de dichas evaluaciones, definiendo estudios, lugares y protocolos de acción [ERPO1].

En relación a estas prestaciones destacan los casos de exposición a agentes de riesgo, específicamente a polvos neumoconiogenicos. Dentro de las pautas internacionales (OIT) y nacionales para la evaluación se recomienda el uso de una radiografía de tórax como medio para la detección precoz.

Existe una serie de casos a nivel nacional no detectados a tiempo, lo que ha ocasionado que se produzca un deterioro en calidad de vida del trabajador, imposibilidad de trabajar y gastos adicionales a la Mutual de Seguridad. Destaca que entre los años 2003 y 2007 en Chile se detectó una prevalencia de silicosis en 240 pirquineros, de los cuales el 16.7%, fue detectado por radiografía de tórax.

Respecto a lo tecnológico en 1943 se creó el primer modelo matemático simplificado de una neurona biológica. A este trabajo le siguieron varios otros que refinaron el modelo y mostraron que las llamadas neuronas artificiales podían aprender a resolver tareas complejas. Se empujó entonces la idea de conectar pequeñas unidades computacionales simples formando grandes redes de neuronas artificiales sería la clave para construir sistemas computacionales con capacidades similares a las humanas. Las investigaciones de Hinton, Bengio y LeCunn mantuvieron con vida a las redes neuronales artificiales dando paso al área de aprendizaje profundo (Deep learning) y redes convolucionales, que engloba las técnicas de construcción, entrenamiento y aprendizaje de estas redes. Existiendo actualmente y gracias a estas técnicas los sistemas computacionales hoy son capaces de analizar imágenes e indicar la presencia de objetos en ellas. Y es relevante en el contexto de salud por que puede localizar patrones en una radiografía de tórax.

- Problema:

Actualmente la Mutual de Seguridad realiza revisiones preventivas ocupacionales y pre ocupacionales de trabajadores dentro del protocolo de vigilancia de los trabajadores con exposición a sílice. En este protocolo se establece la radiografía de tórax como uno de los exámenes obligatorios que define de acuerdo a sus resultados diferentes conductas a seguir. De acuerdo con la normativa y el protocolo Ministerial PLANESI, la lectura de Radiografía de Tórax para detectar neumoconiosis requiere que estas radiografías además de la técnica específica para la toma, se

requiere por parte de los médicos una capacitación y aprobación como lector OIT. En Chile solo la Universidad de Chile realiza esta capacitación, por poca disponibilidad de cupos al año y con una tasa de aprobación muy baja. Por lo tanto, las habilidades específicas con las que cuentan los médicos generales de los sistemas previsionales y los OAL para detectar neumoconiosis es de baja sensibilidad para estados iniciales de la patología, detectándose muchas veces en estados avanzados, lo que no permite tomar acciones de prevención, como retirar de la exposición al riesgo. Esta misma falta de capacidad para interpretar las radiografías por parte de los médicos además genera otro problema que es la derivación para mayores estudios de trabajadores que no lo requieren, generando un mal uso de recursos por parte de las mutualidades. Y finalmente desde el punto de vista del sistema de salud común, se generan muchos pacientes no sean derivados a tiempo al sistema laboral, por diagnóstico o sospecha tardía.

- **Objetivos:**

Apoyar a los médicos generales en sus revisiones ocupacionales, pre ocupacionales y las realizadas en el marco de los programas de vigilancia por exposición a agentes neumoconiógenos otorgándoles información relevante sobre la radiografía de tórax, con el fin de que este tome una mejor decisión y logre una mejor detección de las enfermedades de origen laboral (mejoramiento del asertividad), logrando con esto:

Detectar en pacientes signos compatibles con neumoconiosis en etapas leves, moderadas y graves con al menos un 70% de sensibilidad, contribuyendo al abordaje y retiro de la exposición.

Descartar pacientes sanos apoyado con la tecnología desarrollada que brinde al menos un 70% de especificidad, lo que logrará disminuir las derivaciones incorrectas de pacientes sanos, logrando una reducción de costos en traslados y uso de equipamiento físico y humano.

- **Solución:**

Creación de Sistema web de Inteligencia Artificial que sea capaz de detectar por medio de una imagen de tórax frontal signos de neumoconiosis.

1 METODOLOGÍA

A continuación, se explica las grandes etapas en la creación de un sistema de inteligencia artificial y visión por computadora bajo la estructura de aprendizaje supervisado.



Figura 3: Metodología en cascada para el desarrollo del sistema.

- Extracción y preparación de datos:

Mutual de Seguridad es una corporación privada sin fines de lucro que otorga cobertura en accidentes laborales, estos manejan un gran número de sistemas informáticos y se destaca el sistema de almacenamiento de radiografías (sistemas de Imagenología), donde la cantidad de radiografías de tórax frontal con signos patológicos de neumoconiosis asciende a 1.787. Por otro lado, la cantidad de imágenes sanas recopiladas fueron un total de 1.650 contenidas todas las imágenes en formato DICOM. Luego de su descarga, se realizó una extracción de la imagen (algoritmo de extracción) donde estas fueron anonimizadas con el fin de resguardar la confidencialidad de la información del paciente.

- Análisis y selección de imágenes:

Una vez extraídas las imágenes son analizadas para clasificar el grado de compromiso pulmonar por neumoconiosis. La escala propuesta fue de tres grados: leve, moderado y severo, ver figura 4.

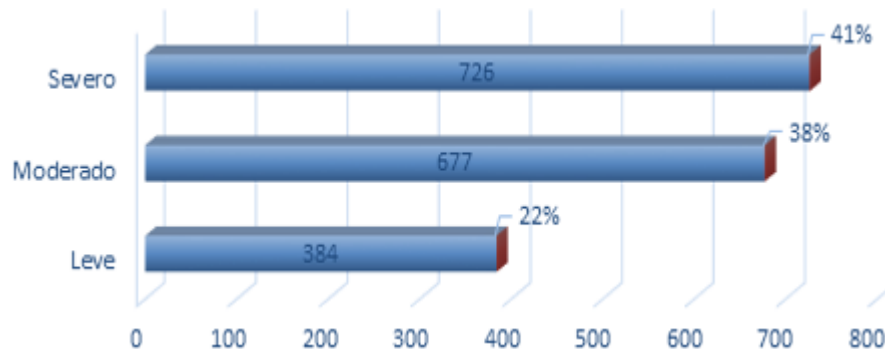


Figura 4: Distribución de Radiografías analizadas según grado de compromiso pulmonar por neumoconiosis obtenidas desde la Mutual de Seguridad.

Donde la clasificación se definió de la siguiente manera:

Leve: Patrones OIT: 1/0, 1/1 y 1/2:

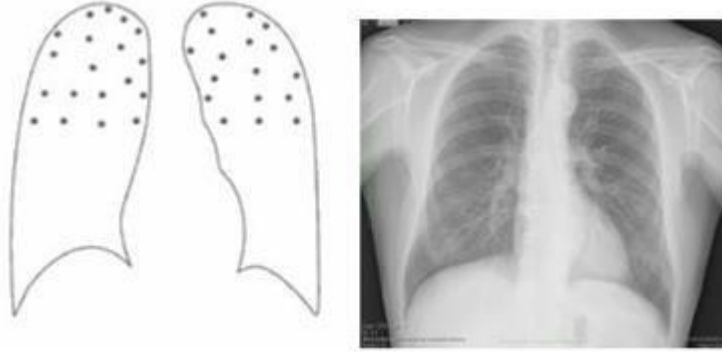


Figura 5: Ejemplo de imagen de neumoconiosis leve.

Moderado: Patrones OIT: 2/1, 2/2 y 2/3

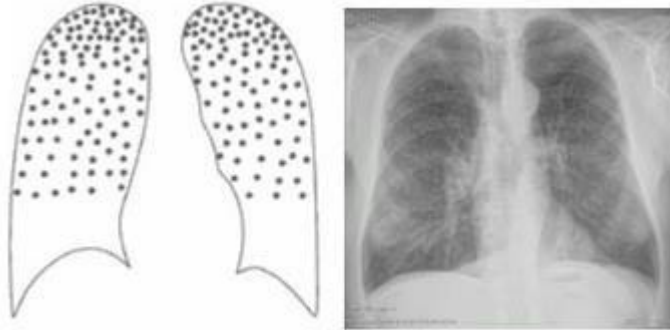


Figura 6: Ejemplo de imagen de neumoconiosis moderada.

Severo Patrones OTI: 3/2, 3/3 y 3/+

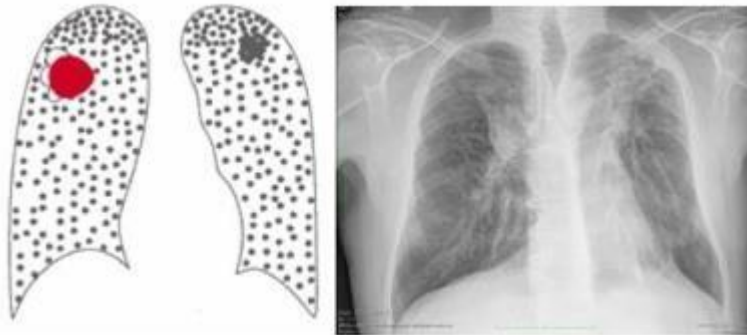


Figura 7: Ejemplo de imagen de neumoconiosis severa.

Luego de analizar las estadísticas de compromiso pulmonar, se observa un punto crítico donde el algoritmo de inteligencia artificial le dificultará diferenciar pacientes con neumoconiosis leve versus pacientes sanos (son muy similares en lectura OIT), por lo que se sugieren el desarrollo de dos modelos de inteligencia artificial uno para descartar pacientes con neumoconiosis leve versus otra alteración llamado "Binary", para que luego actúe un segundo modelo llamado "Multilabel" en caso de presentar

signos de alteración que no corresponda, este segundo modelo pueda discriminar si contiene signos de neumoconiosis o no. Este segundo modelo se espera que sea más sensible al detectar casos leves, moderados y graves de signos de neumoconiosis.

De cara al entrenamiento del modelo de IA siempre se definen al menos tres conjuntos de datos: El conjunto llamado "Training" es donde existen el patrón de manera clara con el fin que el modelo aprenda correctamente y se distinga de otro (edema, consolidación, etc...). Para el conjunto de "Val" es un subconjunto de entrenamiento donde el patrón no es tan claro de modo tal que la maquina se la "exija" entender y distinguir el signo patológico. Entre ambos conjuntos son parte del entrenamiento y ajuste de los pesos del modelo de inteligencia artificial. Finalmente, el conjunto de imágenes de "Test" son imágenes que el modelo entrenado no ha visto ni han sido usado en su entrenamiento, este set es principalmente obtener métricas y medir los valores predictivos.

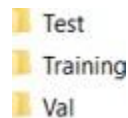


Figura 8: Conjuntos de datos a usar en el entrenamiento del modelo de IA.

Para el modelo de inteligencia artificial que distingue pacientes neumoconiosis leve versus alterados, (llamado "Binary"), la composición es la siguiente:

El set de "training" se compone de 9.479 imágenes de pacientes sin signos de neumoconiosis (OIT como no OIT) y 1230 imágenes de neumoconiosis (sólo OIT).

El set de "Val" se compone de 4.069 imágenes de pacientes sin signos de neumoconiosis (OIT y no OIT) y 406 imágenes de pacientes con signos de neumoconiosis (sólo OIT).

Para el segundo modelo de inteligencia artificial que es altamente sensible distinguiendo patrón de neumoconiosis 0/1 o superior (llamado "Multilabel") la composición es la siguiente:

El set de "training" se compone de 8.000 imágenes con signos de patologías comunes (No OIT) y 1230 imágenes de neumoconiosis (sólo OIT).

El set de "Val" se compone de 2.000 imágenes con signos de patologías comunes (NO OIT) y 406 imágenes de neumoconiosis (sólo OIT).

Para ambos modelos se realizará un solo set de "Test" conteniendo 3026 imágenes de pacientes no neumoconióticos (sanos y pacientes con y patología respiratoria común todos no OIT) y 151 imágenes que contienen patologías de neumoconiótica (21.3% compromiso pulmonar leve, 32.7%

compromiso pulmonar moderado y 46% compromiso pulmonar severo), dando una prevalencia del set de 4.7%.

- Balance de datos:

Es posible apreciar que los conjuntos tanto de "Training" como de "Val" son naturalmente desbalanceados donde la prevalencia de la enfermedad de neumoconiosis bordea el 11%, en los sets declarados anteriormente, esto quiere decir, que hay muchos de un patrón por ejemplo "consolidación" y muy pocos ejemplos de "neumoconiosis". Este problema es habitual en problemas de salud, pero existe una manera de mitigarlos a continuación se explica cómo:

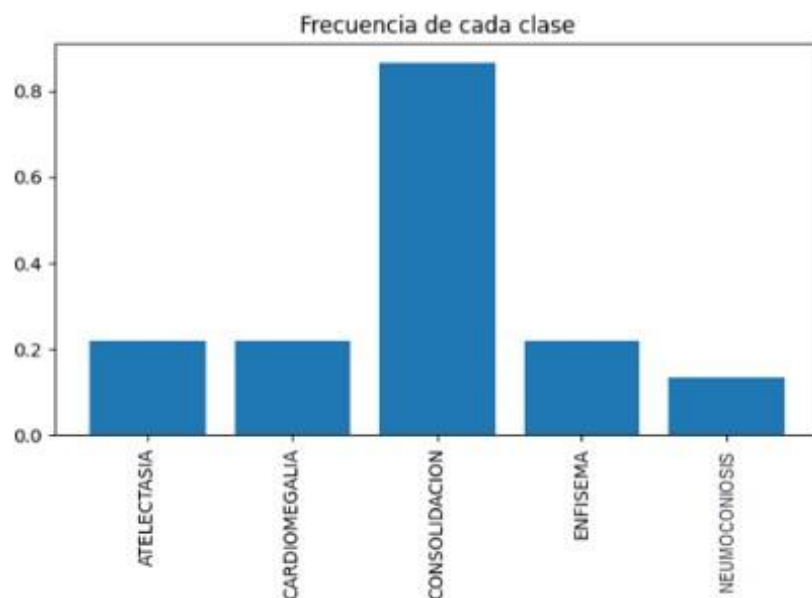


Figura 9: Frecuencias de las principales patologías extraídas de los conjuntos de datos.

Esto significa que cuando se entrenen los modelos el conjunto de patologías con mayor frecuencia contribuye al entrenamiento mucho más porque existe más información de esa característica, mientras que otra patología con menor frecuencia de aparición (positivo) generará un aporte menor en el entrenamiento.

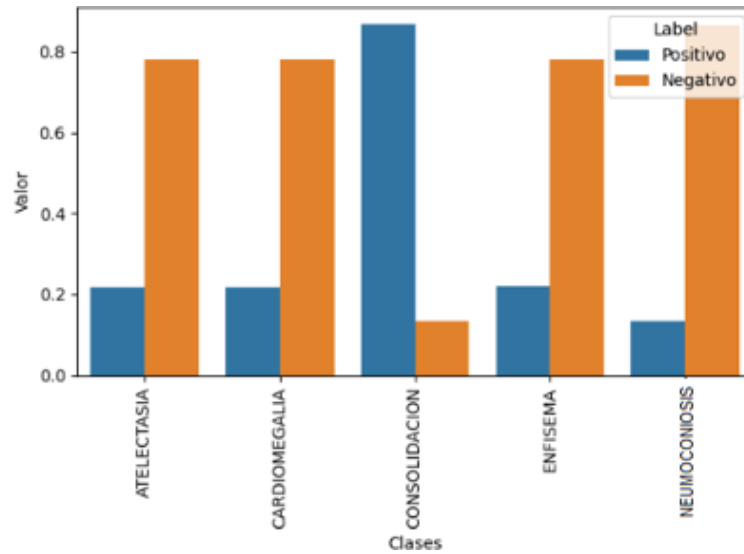


Figura 10: Frecuencias de los casos positivos (existencia de la patología) y casos negativos (ausencia de la patología).

Quando uno mide la cantidad de veces que aparece versus la cantidad de veces que no aparece en el conjunto total (figura 10), se observa claramente que el aporte de la patología "Neumoconi6tica" es menor respecto a los otros signos. Para mejorar esto se realiza un balance entre los casos positivos y negativos entre ellos y por lo tanto ahora todas las clases contribuyen de manera equivalente al entrenamiento ver figura 11.

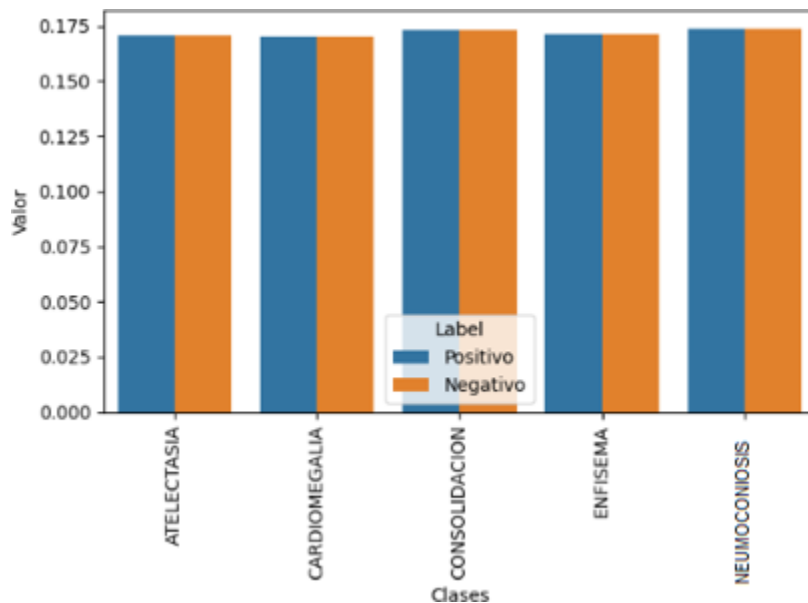


Figura 11: Balance entre las patologías, tanto en aporte para casos positivos, negativos y en el peso relativo entre ellas.

Esta técnica de balance de datos es llamada "Focal Loss" y es ampliamente usada, sirve para indicar por medio de la función "Loss" durante el entrenamiento que los pesos se ajusten dinámicamente mientras se procesan, de manera tal, que el modelo en su conjunto tenga incentivos a aprender los patrones de todas las clases de manera equivalente y no sólo aprender de la clase que tiene más información. Esto supone que modelo entrenado tendrá un mejor comportamiento en la predicción [JMWYZB21].

- Entrenamiento de los modelos de IA:

El entrenamiento de un modelo de inteligencia artificial tiene inicialmente dos conceptos con los que uno puede observar si su entrenamiento es correcto, estos son:

"Loss" o pérdida, se define como la diferencia entre lo que predice la red y el valor real, vale decir, en la medida que la red neuronal "aprenda" el error de "Val" disminuye respecto del error de "Training". El entrenamiento será eficiente mientras no existe un "sobre ajuste", vale decir, hasta que el error "Val" sea menor que el error "Training".

"Accuracy" o exactitud es simplemente una medición que tiene que ver con la cantidad de predicciones correctas versus el total de predicciones. Mientras más alto el número podríamos entender que la red está clasificando con mayor "certeza", considerando que el conjunto de datos se encuentra balanceado, como es nuestro caso.

A continuación, se muestran las gráficas del entrenamiento del modelo "Binary" de inteligencia artificial que diferencia radiografías sin signos patológicos versus radiografías con signos pneumoconióticos:

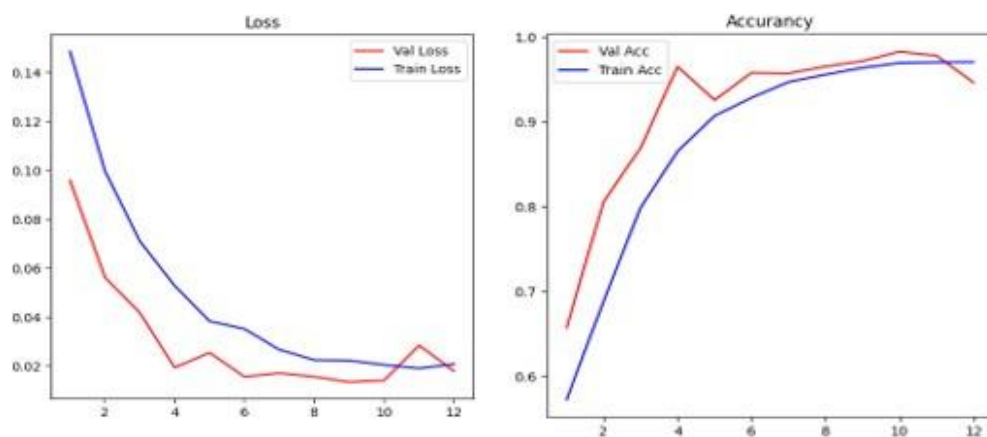


Figura 12: Curvas de entrenamiento de "Loss" y "Accuracy" del Modelo "Binary".

En el caso de "Accuracy", los aciertos de "Training" y "Val", se ven progresivamente aumentando, ajustándose a valores cercanos al 90% en el ciclo 9, donde ambas curvas comienzan a converger.

Para el caso del grafico de "Loss", tanto como "Training" y "Val", tiene una tendencia a ir decreciendo hasta encontrar un valle alrededor del ciclo 9, donde luego comienza a aumentar el error de "Val" sobre "Training" insinuando un "sobre ajuste".

Las gráficas del entrenamiento del modelo de inteligencia artificial llamado "Multilabel" que detecta si existe signos de neumoconiosis 0/1 o superior es como sigue:

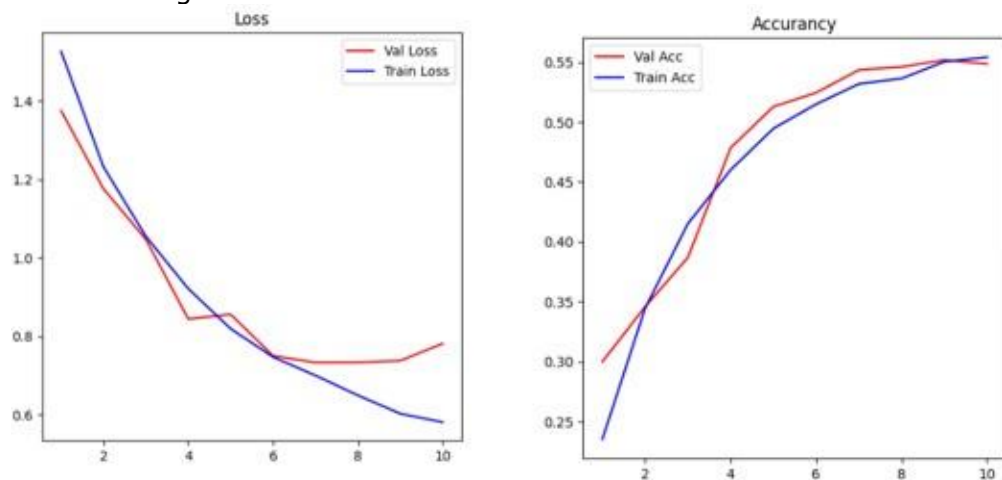


Figura 13: Curvas de entrenamiento de "Loss" y "Accuracy" del Modelo "Multilabel".

En el caso de "Accuracy", los aciertos de "Training" y "Val", se ven progresivamente aumentando, ajustándose a valores cercanos al 55% en el ciclo 7, donde ambas curvas comienzan a converger, no alcanzando valores mayores debido probablemente a falta de información sobre neumoconiosis.

Para el caso del grafico de "Loss", tanto como "Training" y "Val", tienen una tendencia a ir decreciendo hasta encontrar un valle alrededor del ciclo 7, donde luego comienza a aumentar el error de "Val" sobre "Training" insinuando un "sobre ajuste", esto debido a que no es posible disminuir más el error de predicción de "Val".

Finalmente indicar que cada modelo indicado ha sido entrenado bajo la técnica de "Transfer Learning".

- Desarrollo del producto web:

Respecto del desarrollo del producto se indicará de manera gráfica como ambos modelos de inteligencia artificial estarán trabajando bajo una única plataforma web. Iniciando con el envío de la radiografía de tórax frontal desde el hospital hacia la plataforma donde es recibida y procesada para luego ser enviada a los motores de I.A. donde primero actúa el modelo "Binary" descartando si el paciente tiene signos neuromoconióticos o no, para luego volver a ser analizado por el segundo modelo "Multilabel" y en caso de existir alguna alteración está sea correctamente detectada, descartando que se confunda con otra patología respiratoria. Finalmente se envía el resultado de regreso al usuario del hospital.

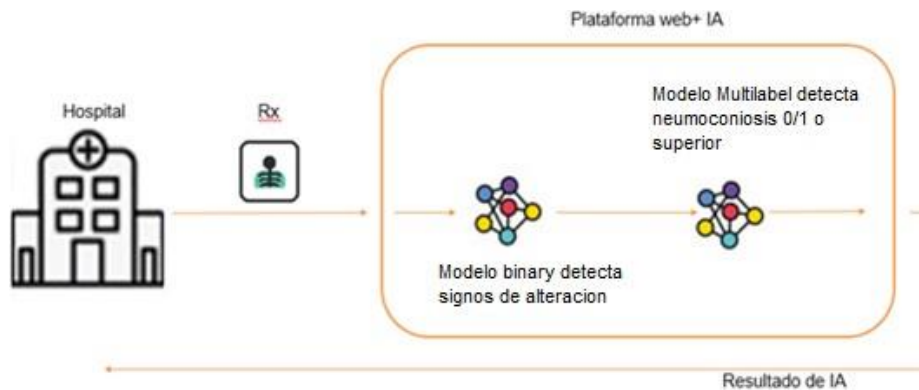


Figura 14: Flujo de información desde el hospital hacia los modelos de inteligencia artificial.

Respecto a atributos del producto este cuenta con:

Seguridad en todos los niveles, cifrado en el envío de datos, bases de datos cifradas y respaldos mensuales.

El sistema es capaz de procesar DICOM y fotografías.

Es un sistema web adaptable a dispositivos como PC de escritorios Tablet y teléfonos celulares inteligentes.

El procesamiento se desarrolla en CPU y tarda segundos en la entrega de resultados.

Se adjuntan diferentes capturas de pantalla del producto web desarrollado ver figura 15.

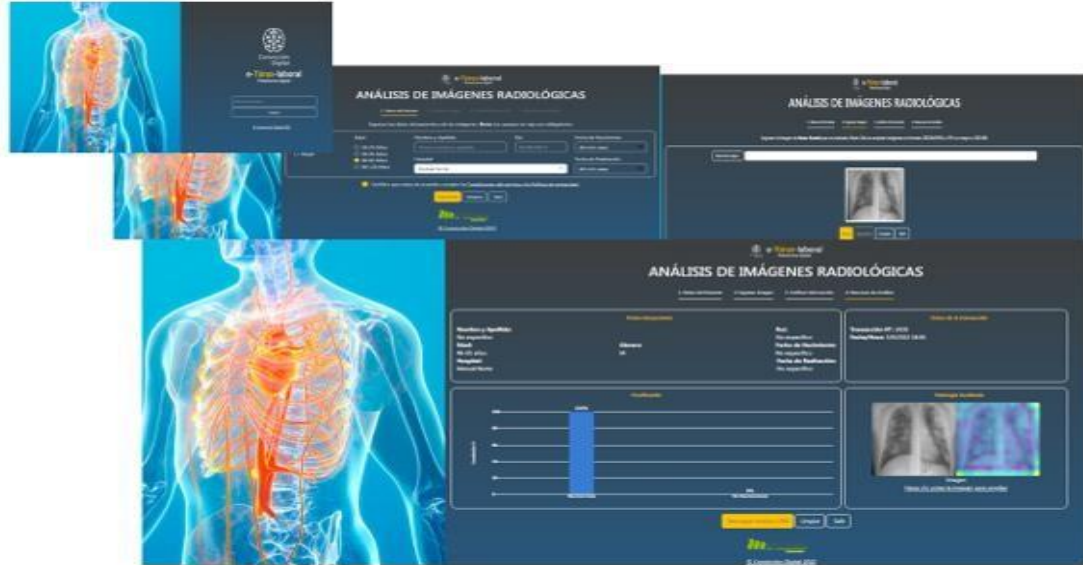


Figura 15: Capturas de proceso de plataforma para procesar datos de imágenes radiológicas.

Es posible acceder a la siguiente URL <https://www.prod-convicciondigital.cl/Laboral/index.html>

El proceso del usuario en la plataforma es como se indica a continuación:

Se accede por un "login" con una clave genérica para cada centro perteneciente al organismo asegurador.

Luego de ingresar al usuario completa un formulario donde completa los datos demográficos del paciente.

En el paso siguiente el usuario adjunta la imagen de la radiografía de tórax frontal o el DICOM del paciente con la imagen y se procede a enviar el archivo para su análisis con I.A.

En un rango de [5-15] segundos se obtiene el resultado del análisis y el usuario puede enviar un "feedback" que tiene relación con almacenar estadísticas correctas o incorrectas sobre las predicciones que realiza la I.A. que serán almacenadas y analizadas con el fin de ir ajustando los modelos en un mejoramiento continuo.

Finalmente, el usuario avanza al último paso donde este contiene un resumen integrando datos demográficos del paciente y resultado del análisis de I.A. junto con la posibilidad de descargar el análisis en un pdf.

RESULTADOS

- Resultados modelo "binary":

Se realizaron pruebas de modelo con un total de 3177 imágenes de pacientes y de ellas 3026 son radiografías de pacientes sin signos de neumoconiosis (no lectura OIT, radiografías con error en toma y con diversas patologías respiratorias de tipo común) y 151 si contienen neumoconiosis (con lectura OIT), al ser procesados por el modelo "Binary" los resultados se grafican en la siguiente matriz de confusión, ver figura 16.

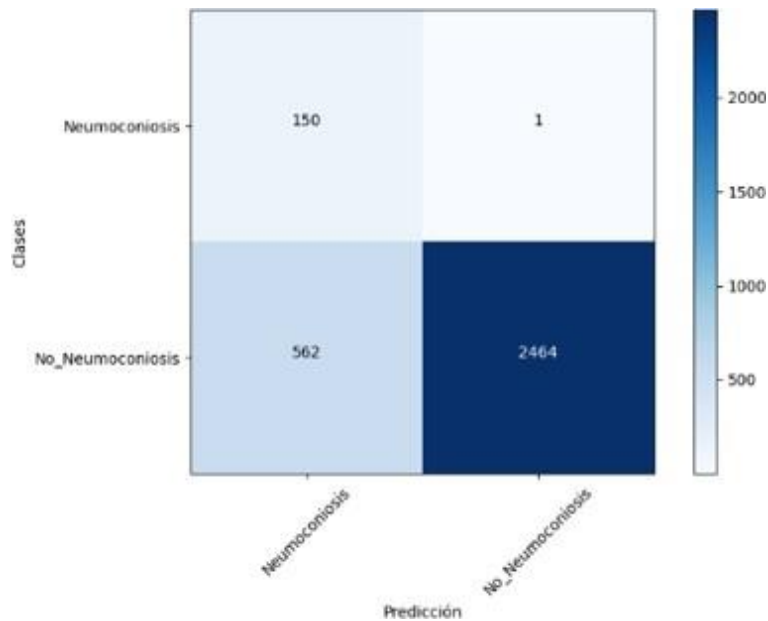


Figura 16: Resultado del modelo "Binary" respecto del set de "Test", graficados los resultados en una Matriz de Confusión.

Obteniendo las siguientes definiciones y resultados:

TP (True positive) =150
 FP (False positive) =562
 TN (True Negative) =2464
 FN (False Negative) =1

Utilizando los valores de la matriz de confusión es posible derivar en indicadores que nos dan una versión más completa del rendimiento del modelo "Binary".

Indicador	Nombre	Formula	Valor	95% de Confianza	
				L. Inferior	L. Superior
Prevalencia		$prev = \frac{TP + FN}{TP + FN + TN + FP}$	0.0475	0.0404	0.0555
Sensibilidad		$se = \frac{TP}{TP + FN}$	0.9934	0.9637	0.9998
Especificidad		$sp = \frac{TN}{TN + FP}$	0.8143	0.8000	0.8280
Falso Positivo rate		$FPR = 1 - sp$	0.1857	0.1720	0.2000
False Negativo rate		$FNR = 1 - se$	0.0066	0.0002	0.0363
Valor Positivo Predictivo		$PPV = \frac{TP}{TP + FP}$	0.2107	0.1813	0.2425
Valor Negativo Predictivo		$NPV = \frac{TN}{TN + FN}$	0.9996	0.9977	1.0000
Positive Likelihood ratio (LR+)		$PLR = \frac{se}{1 - sp}$	5.3500	4.9586	5.7695
Negative Likelihood ratio (LR-)		$NLR = \frac{1 - se}{sp}$	0.0081	0.0012	0.0574
Diagnostico Odds ratio (DOR)		$DOR = \frac{TP + TN}{FP + FN}$	657.6512	91.8407	4709.2989

Figura 17: Indicadores de diagnóstico del modelo "Binary".

Tras evaluar la tabla presentada en la figura 17, las estadísticas indican que el modelo puede discriminar correctamente los pacientes con signos de neumoconiosis de los que no contienen el signo, descartando también sanos debido a:

En términos de Sensibilidad y Especificidad, el modelo entrenado obtuvo un 99% y 81% respectivamente.

Respecto a los valores predictivos positivos y negativos, el valor PPV es de un 21.07% con una prevalencia de un 4.7% del set, mientras que el valor NPV es de un 99.9% indicando que cuando descarta el paciente es muy probable que sea un paciente sin neumoconiosis.

Respecto a los coeficientes de probabilidades o razón de verosimilitud (LR+, LR-), para el caso de LR+ obtuvo un valor de 5.3, resultado que se encuentra en el rango de utilidad clínica "Buena", dado que aún hay un espacio de mejora. Cuando el modelo "Binary" indica que tiene neumoconiosis el paciente entonces debe ser re-evaluado por el modelo "Multilabel" (tal como se diseñó). Por otro lado, el LR- obtuvo un valor menor que 0.1, es decir, un resultado con utilidad clínica "altamente relevante", por lo que cuando la predicción indica que no tiene neumoconiosis

entonces es muy probable que la radiografía enviada del paciente no tenga el signo de neumoconiosis.

Respecto de la curva ROC, que es la forma de medir capacidad discriminativa del diagnóstico (exactitud diagnóstica), el resultado entregado fue de un 90%, siendo lo máximo un 100%, lo que nos indica que encontró un patrón discriminativo respecto a las radiografías con signos de neumoconiosis versus radiografías sin el signo patológico.

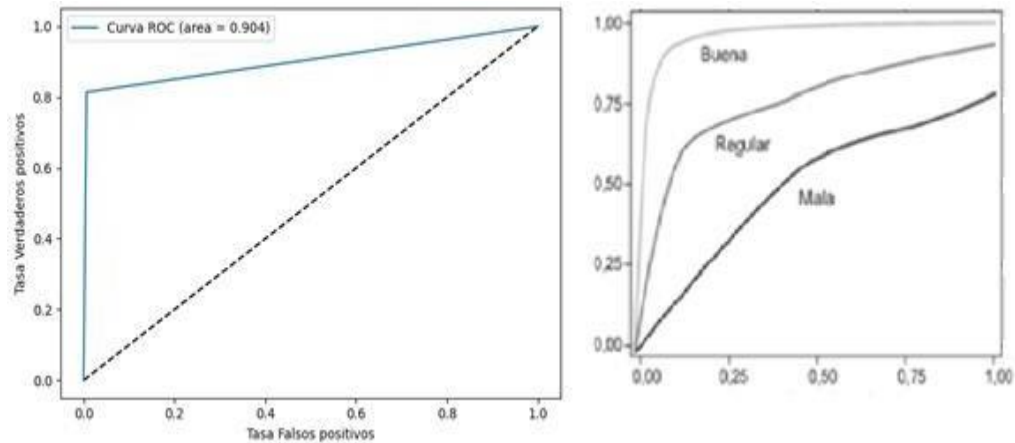


Figura 18: Curva ROC del modelo "Binary" y el indicador AUC (Izquierda). Interpretación de la curva ROC (Derecha).

Al analizar las 562 radiografías donde el modelo "Binary" indicó falsos positivos (FP), se destaca que predice incorrectamente porque:

En hombres sanos con grasa pectoral y/o musculo pectoral denso generando un refuerzo en la imagen, que el sistema clasifica como Neumoconiosis, equivalentes al el 36.2% del total de radiografías de hombres sanos.

En hombres con signos de alguna patología común principalmente consolidaciones, habitualmente acompañada de otros signos patológicos.

En mujeres sanas con mamas densas generando una "sombra" sobre el campo pulmonar clasificando la imagen como neumoconiosis, equivalente al 63.3% de mujeres sanas.

En mujeres con patología respiratoria común y buena toma, equivalente al 13.9% del total de radiografías alteradas.

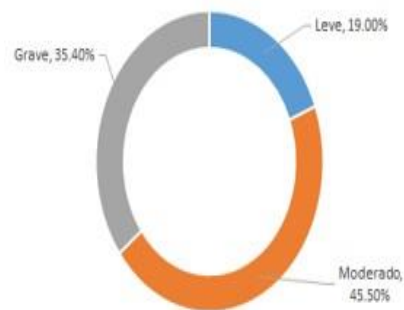
Para el caso de tomas incorrectas correspondiente al 6.6% del total de radiografías falsas positivas, mayoritariamente el modelo falla

donde el campo pulmonar no está completo (principalmente paciente con toma acostado), también en rotaciones generando refuerzos en zonas de las costillas e hilios.

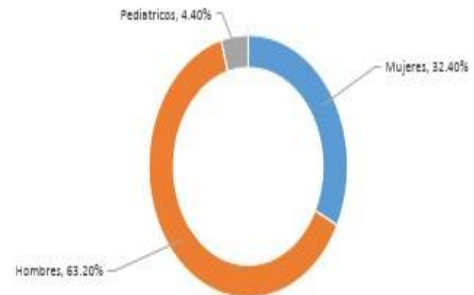
Imágenes pediátricas representan un 4.4% del total de radiografías del ser de "Test" y fueron incluidas a modo de desafiar al sistema, resultando que las incorrectas contienen principalmente signos compatibles con consolidación.

Para el caso de las imágenes con alteraciones, según el grado de compromiso pulmonar la distribución detectada fue 19% leve, 45.5% moderado y un 35.4% compromiso pulmonar grave, donde el hallazgo preponderante es consolidación.

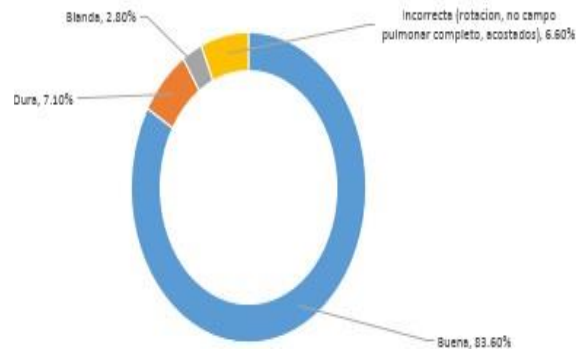
Grado de compromiso pulmonar de RX falsas+



Distribución por genero/edad de RX falsas+



Distribución de calidad de la toma en RX falsas+



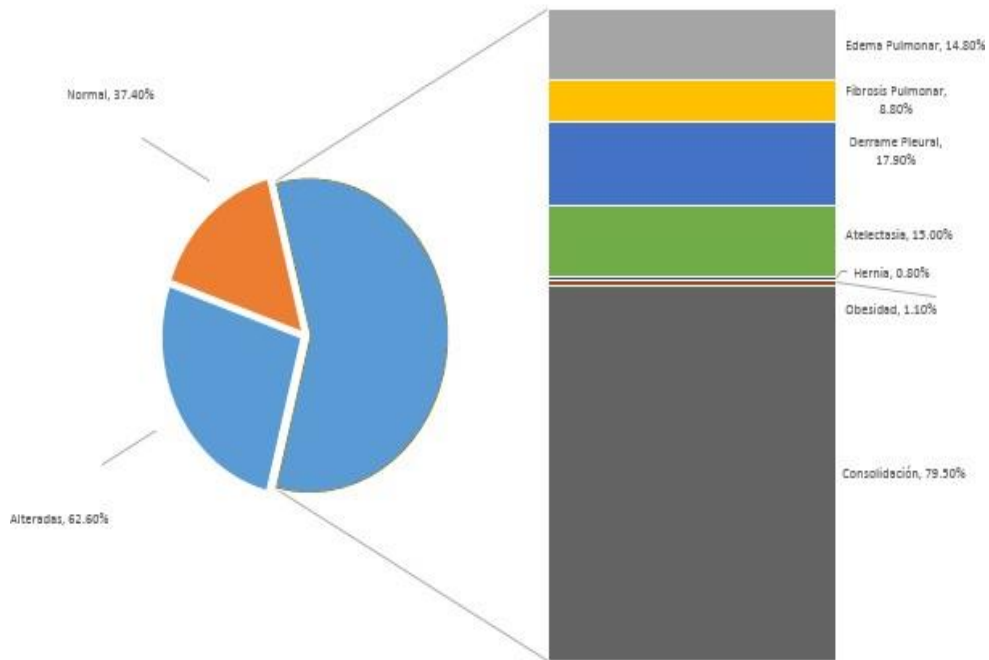


Figura 19: Gráficos de distribución de Genero/edad, clasificación de la calidad de toma, Distribución Hallazgos encontrados y compromiso pulmonar en las de imágenes de tórax que resultaron ser Falsos + de Modelo "Binary".

Como resumen respecto del modelo "Binary" tiene un LR- altamente relevante por cuando indica que una radiografía no contiene signos de neumoconiosis, es muy probable que el paciente no cuente con el signo. Por otro lado, el indicador LR+ nos muestra que puede fallar incrementando la cantidad de falsos positivos porque:

Una mala toma (que contenga rotaciones, campo pulmonar incompleto, etc...) genera ruido en la imagen indicando al modelo que pueda interpretar que existen alteraciones debido a la superposición de elementos y por ellos mal clasificando las imágenes como neumoconiosis.

Falla frecuentemente en radiografías de pacientes sanos (hombres y mujeres) en 37.4% de los casos con mamas y/o pectorales densos que generan un refuerzo, donde el sistema indica que presenta neumoconiosis.

En pacientes con alteraciones, principalmente con consolidaciones.

Según lo anteriormente indicado el modelo "Binary" genera un 17.7% (Falsos positivos/Total del set de "Test") de falsos positivos, clasificando radiografías con signos de neumoconiosis donde no las hay. Por la razón

anteriormente indicada es conveniente volver a analizar si efectivamente existe un patrón compatible con neumoconiosis en la imagen con un segundo modelo "Multilabel".

- Resultados modelo Multilabel:

El modelo "Multilabel" debe ahora procesar unas 712 imágenes (falsos positivos y verdaderos positivos), donde 562 son radiografías con lectura no OIT y 150 son con lectura OIT y signos de neumoconiosis. El resultado de la predicción se muestra en la matriz de confusión de la figura 20.

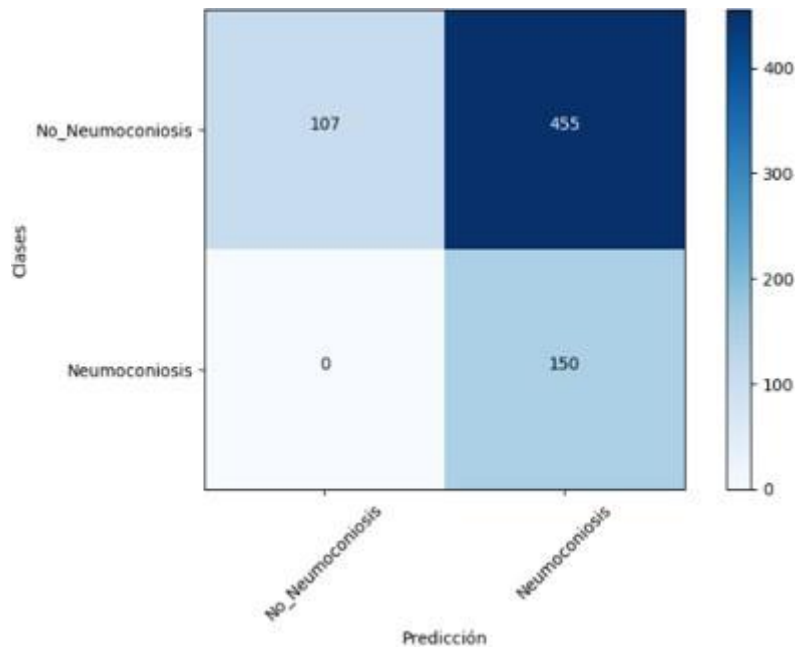


Figura 20: Matriz de confusión de modelo "Multilabel".

Obteniendo las siguientes definiciones y resultados:

TP (True positive) =150

FP (False positive) =455

TN (True Negative)=107

FN (False Negative)=0

Utilizando los valores de la matriz de confusión es posible derivar en indicadores que nos dan una versión más completa del rendimiento del modelo "Multilabel".

Indicador	Nombre	Formula	Valor	95% de Confianza	
				L. Inferior	L. Superior
Prevalencia		$prev = \frac{TP + FN}{TP + FN + TN + FP}$	0.2118	0.1823	0.2436
Sensibilidad		$se = \frac{TP}{TP + FN}$	0.9934	0.9637	0.9998
Especificidad		$sp = \frac{TN}{TN + FP}$	0.1904	0.1587	0.2253
Falso Positivo rate		$FPR = 1 - sp$	0.8096	0.7747	0.8413
False Negativo rate		$FNR = 1 - se$	0.0066	0.0002	0.0363
Valor Positivo Predictivo		$PPV = \frac{TP}{TP + FP}$	0.2474	0.2140	0.2843
Valor Negativo Predictivo		$NPV = \frac{TN}{TN + FN}$	0.9907	0.9495	0.9998
Positive Likelihood ratio (LR+)		$PLR = \frac{se}{1 - sp}$	1.2270	1.1763	1.2798
Negative Likelihood ratio (LR-)		$NLR = \frac{1 - se}{sp}$	0.0035	0.0049	0.2472
Diagnostico Odds ratio (DOR)		$DOR = \frac{TP + TN}{FP + FN}$	35.2747	4.8814	254.9063

Figura 21: Indicadores de diagnóstico del modelo "Multilabel".

Tras evaluar la tabla presentada en la figura 21, las estadísticas indican que el modelo logró detectar pacientes con alteraciones compatibles con neumoconiosis, pero existe una tendencia a entregar resultados de falsos positivos. Por otro lado, cuando indica que no contiene el signo patológico tiende a fallar muy poco.

En términos de Sensibilidad y Especificidad, el modelo entrenado obtuvo un 99% y 20% respectivamente.

Respecto a los valores predictivos positivos y negativos, el valor PPV es de un 24.7% con una prevalencia de un 21% del set, mientras que el valor NPV es de un 99.9% indicando que cuando descarta el paciente es muy probable que sea un paciente sin neumoconiosis, por otro lado el valor bajo de PPV es debido a la cantidad de falsos positivos entregados.

Respecto a los coeficientes de probabilidades o razón de verosimilitud (LR+, LR-), para el caso de LR+ obtuvo un valor de 1.22, muy por debajo de lo esperado, estando en rango de utilidad clínica "Malo", es decir, no es muy confiable a la hora de determinar la presencia de signos de neumoconiosis desde el punto de vista estadístico, sin embargo, hay que destacar que, de 150 radiografías con signo de neumoconiosis clasificó 150. Por otro lado, el indicador LR- obtuvo un valor menor que 0.1 lo que da cuenta que está en un rango de utilidad clínica "Altamente Relevante" para descartar radiografías que no contienen el signo.

Respecto de la curva ROC, que es la forma de medir capacidad discriminativa del diagnóstico (exactitud diagnóstica), el resultado entregado fue de un 87.4%, siendo lo máximo un 100%, lo que nos indica que encontró un patrón discriminativo, pero aún hay trabajo que hacer en mejorar la capacidad predictiva del modelo, ver figura 22.

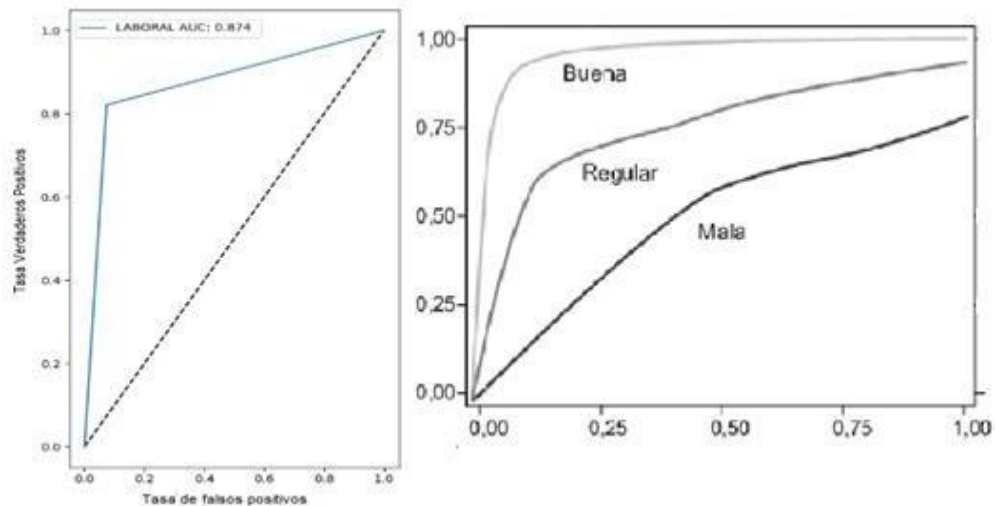


Figura 22: Curva ROC del modelo "Multilabel" y su AUC asociado (Izquierda). Cuadro que explica la interpretación de la curva ROC (Derecha).

Respecto del modelo "Multilabel" al analizar las 455 predicciones incorrectas (falsos positivos FP) uno puede observar que falla en:

El 60% las predicciones incorrectas son al confundir hallazgos de patologías comunes con neumoconiosis, dentro de estas patologías destaca hallazgo de patrón intersticial (compatible con Neumonía) lo que corresponde al 78% del total e radiografías con alteraciones.

El otro 40% son para el caso de Radiografías sin alteraciones (pacientes sanos), dentro de estas destacan casos de mujeres con mamas densas (74% del total de radiografías sanos), lo que produce un refuerzo en la imagen de manera tal que el modelo confunde signos de neumoconiosis con el refuerzo por superposición de elementos.

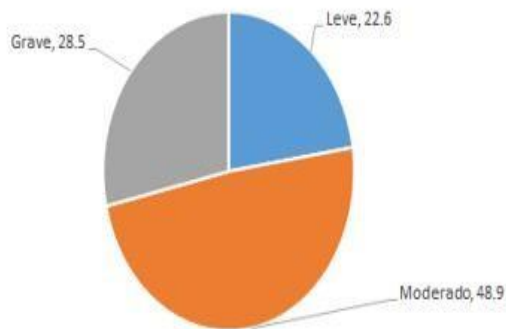
Un elemento importante a destaca en este modelo que, si bien la mayoría de las tomas son correctas, existen tomas incorrectas pero el modelo logra predecir correctamente y descarta acertadamente radiografías que no tienen el signo de neumoconiosis, vale decir, para este modelo la toma no influye manera relevante, correspondiendo al 1.8% de los casos.

Para el caso de pediátricas que fueron incorrectamente clasificadas como neumoconiosis corresponde al 3.1% del total de radiografías

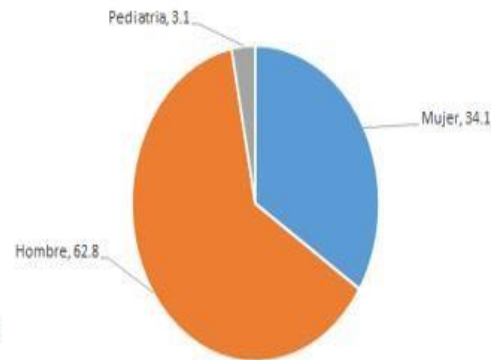
mal clasificadas por el sistema, siendo el principal hallazgo detectado el de consolidación.

Para el caso de las imágenes con alteraciones, según el grado de compromiso pulmonar la distribución detectada fue 22.6% leve, 48.9% moderado y un 28.5% compromiso pulmonar grave, donde el hallazgo preponderante es consolidación.

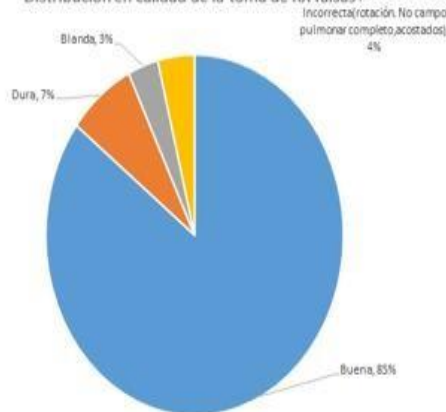
Grados de compromiso pulmonar en RX falsas+



Distribución por genera/edad en RX falsas+



Distribución en calidad de la toma de RX falsas+



Distribución de Hallazgos

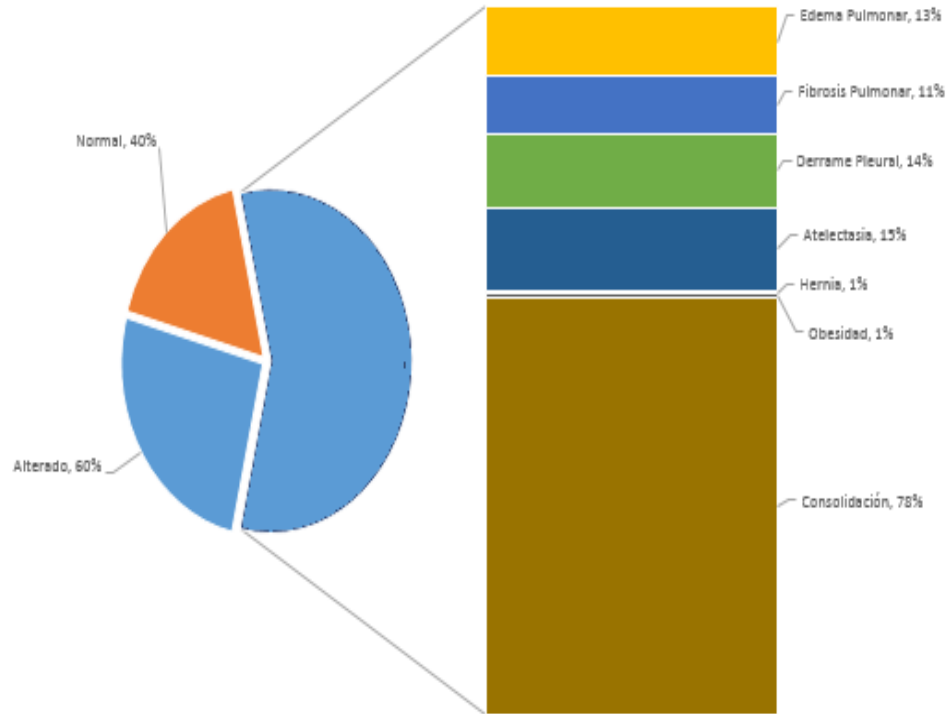


Figura 23: Gráficos de distribución de Genero/edad, clasificación de la lectura, Distribución Hallazgos encontrados y compromiso pulmonar en las de imágenes de tórax, donde el resultado del "Multilabel" fue incorrecto.

Como resumen respecto del modelo "Multilabel" tiene un LR- altamente relevante por cuando indica que una radiografía no contiene signos de neumoconiosis, es muy probable que el paciente no cuente con el signo. Por otro lado, el indicador LR+ nos muestra que puede fallar generando falsos positivos porque:

En el 40% de radiografías sin ningún signo patológico, este grupo de radiografías es de mujeres sanas con mamas densas.

En el 60% de los casos erróneamente clasificados, corresponde a radiografías con alteraciones distintas de neumoconiosis, siendo el hallazgo principal el de signos compatibles con consolidación, específicamente patrón intersticial o secuelas de una consolidación pasada.

Falla en entramado pulmonar aumentado en la mayoría de las veces.

Según lo anteriormente indicado el modelo "Multilabel" genera un 63.9% de falsos positivos (Falsos positivos/Total del set de "Test"), clasificando radiografías con signos de neumoconiosis donde no las hay. Es desde esta perspectiva que se entiende el mal resultado en términos de LR+, y se debe principalmente a la limitada cantidad de imágenes con neumoconiosis que se logró obtener para el entrenamiento del sistema, y que dentro de estas radiografías la cantidad de radiografías con grado de compromiso pulmonar leve es baja dificultando que el sistema "comprenda" todas las variedades en que se presenta el patrón compatible con neumoconiosis.

Figura 24 muestra ejemplo de errores más comunes del sistema "Mutilabel" en paciente sano.

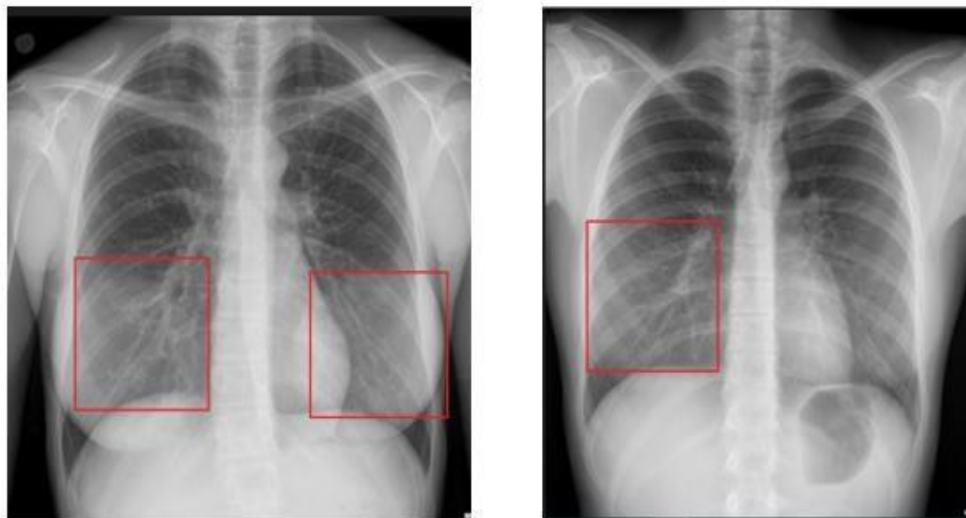


Figura 24: Imágenes de patrón que el modelo "Multilabel" confunde con neumoconiosis. Imagen izquierda mujer sana con mamas densas, imagen derecha hombre sano con patrón intersticial.

Figura 25 muestra ejemplo de errores más comunes del sistema "Mutilabel" en paciente con alteración.

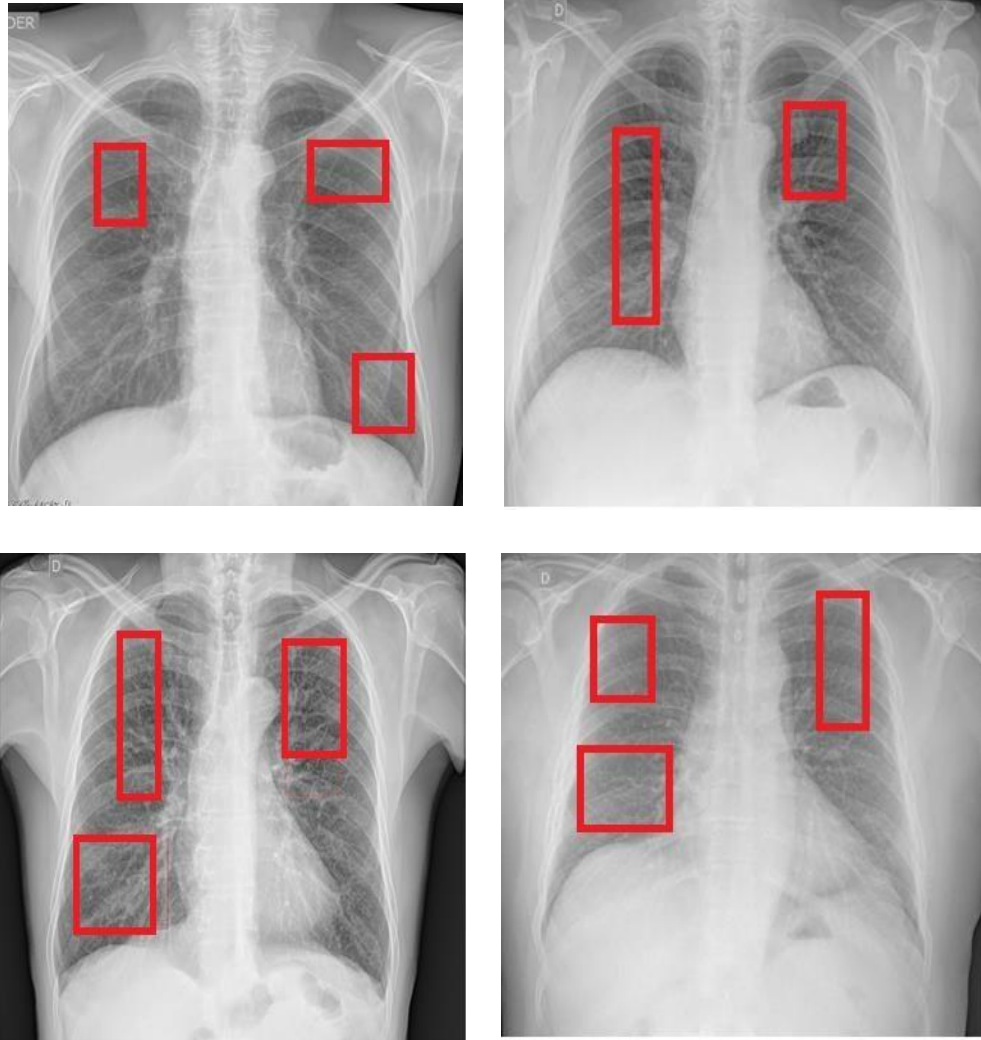


Figura 25: Imágenes de patrón que el modelo "Multilabel" confunde con pneumoconiosis. Imagen con alteración.

- Resultados de ambos modelos:

Al concentrar ambos modelos en un único resultado, la matriz de confusión de confusión resultante se puede apreciar en la figura 26.

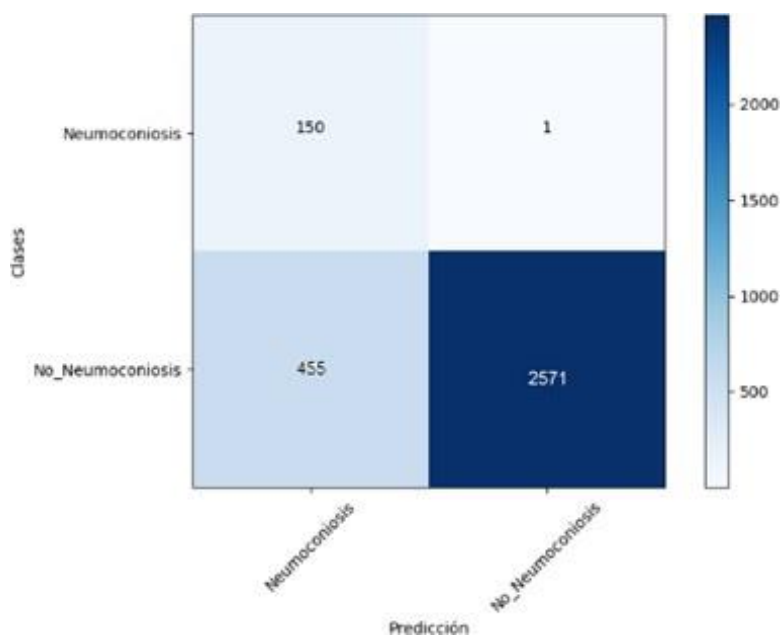


Figura 26: Matriz de confusión del modelo general.

Obteniendo las siguientes definiciones y resultados:

TP (True positive) =150
FP (False positive) =455
TN (True Negative)=2571
FN (False Negative)=1

Utilizando los valores de la matriz de confusión es posible derivar en indicadores que nos dan una versión más completa del rendimiento del modelo "Multilabel".

Indicador	Nombre	Formula	Valor	95% de Confianza	
				L. Inferior	L. Superior
Prevalencia		$prev = \frac{TP + FN}{TP + FN + TN + FP}$	0.0475	0.0404	0.0555
Sensibilidad		$se = \frac{TP}{TP + FN}$	0.9934	0.9637	0.9998
Especificidad		$sp = \frac{TN}{TN + FP}$	0.8496	0.8364	0.8622
Falso Positivo rate		$FPR = 1 - sp$	0.1504	0.1378	0.1636
False Negativo rate		$FNR = 1 - se$	0.0066	0.0002	0.0363
Valor Positivo Predictivo		$PPV = \frac{TP}{TP + FP}$	0.2479	0.2140	0.2843
Valor Negativo Predictivo		$NPV = \frac{TN}{TN + FN}$	0.9996	0.9978	1.0000
Positive Likelihood ratio (LR+)		$PLR = \frac{se}{1 - sp}$	6.6065	6.0640	7.1976
Negative Likelihood ratio (LR-)		$NLR = \frac{1 - se}{sp}$	0.0078	0.0011	0.0555
Diagnostico Odds ratio (DOR)		$DOR = \frac{TP + TN}{FP + FN}$	847.5824	118.3181	6071.7332

Figura 27: Indicadores de diagnóstico del modelo general.

Tras evaluar la tabla presentada en la figura 27, se puede concluir que el modelo en su conjunto logra detectar pacientes alteraciones compatibles con neumoconiosis, pero tienen más aciertos y menos errores al descartar radiografías sin signos patológicos de neumoconiosis, esto se puede evidenciar al mirar los siguientes resultados:

En términos de Sensibilidad y Especificidad, el modelo entrenado obtuvo un 99% y 85% respectivamente, vale decir, para los datos evaluados el modelo detectó pacientes con neumoconiosis en un 99% y descartó pacientes que no tenían el signo en un 85%.

Respecto a los valores predictivos positivos y negativos, el valor PPV es de un 24.7% con una prevalencia de un 4.7% del set, mientras que el valor NPV es de un 99.9% indicando que cuando descarta el paciente es muy probable que sea un paciente sin neumoconiosis.

Respecto a los coeficientes de probabilidades o razón de verosimilitud (LR+, LR-), para el caso de LR+ obtuvo un valor de 6.6, y el LR- un valor menor que 0.1, indicado que para los casos positivos los resultados del sistema tienen una utilidad clínica en grado "Bueno", dado que aún hay espacio de mejora en los resultados importantes se recomienda que estos casos sean evaluados siempre con correlación a historia clínica y derivado para mayores estudios y análisis por parte de equipo certificado por OIT para análisis de imágenes. En el caso de LR- tiene un alto grado de

certeza, encontrándose en un grado de utilidad clínica “Altamente Relevante”, lo que nos indica que puede cambiar conducta clínica con seguridad sin tener que ser derivado para una segunda evaluación. Cuando el sistema indique que la radiografía analizada no cuenta con signos de neumoconiosis.

2 CUMPLIMIENTO DE INDICADORES DE ÉXITO

Los indicadores de éxito referidos para este proyecto fueron varios y en diversos ámbitos tanto técnicos como clínicos, a continuación, se describe los indicadores de cada ámbito:

Ámbito técnico:

Realizar un desarrollo de una plataforma web: Desarrollo cumplido es posible visualizarlo desde el escritorio del médico general, su Tablet y/o smartphone.

Respuesta en segundos (menor que 30 segundos): Desarrollo cumplido, pues a pesar de contener dos modelos de inteligencia artificial para el procesamiento de imágenes, el sistema en conjunto devuelve una respuesta en no más de 15 segundos.

Confidencialidad de la información: Desarrollo cumplido, el sistema cuenta con cifrado en el envío de información por Internet, y también en el almacenamiento de las imágenes e información en su base de datos.

Ámbito Clínicos:

Aumento asertividad: Desarrollo cumplido, pues se traduce en una sensibilidad y especificidad igual o mayor que un 70%, según el estudio realizado el sistema predice en términos de sensibilidad y especificidad valores mayores que el 80%, de esta manera nos acercamos a desarrollos de alto nivel equivalentes a resultados que arrojan otros países del mundo en desarrollos de este tipo de tecnologías.

Indicadores pretest “Altamente Relevante”: Desarrollo cumplido, el LR- tiene un valor menor que 0.1 lo que es estadísticamente “altamente relevante” con lo cual puede confiar en el test para guiar la conducta clínica, no obstante, el LR+ es estadísticamente “Bueno” generando muchos falsos positivos, pero es mejor revisar dos veces el paciente que encontrarnos con neumoconiosis avanzada luego sin ser detectada.

3 CONCLUSIONES Y RECOMENDACIONES

El proyecto desarrollado tiene un impacto relevante para las mutualidades de Chile y sistemas de previsión de salud común, porque está orientado a mejorar la salud de los pacientes desde un punto de vista clínico tecnológico, en donde se ha desarrollado una plataforma web que brinda información relevante al médico sobre la detección de neumoconiosis en flujos ocupacionales, pre ocupacionales y dentro

de programas de vigilancia por exposición a agentes neumoconiogénicos o ante cualquier toma de radiografía de tórax, generando con este proceso una herramienta de apoyo clínico relevante para médicos, con lo cual estos pueden cambiar conducta clínica frente al paciente y con esto generar una revisión más precisa y efectiva, así también se proyecta que podría ayudar a mejorar la salud de la población trabajadora (retirando de exposición de forma precoz e iniciando los tratamientos correspondientes) y ser al mismo tiempo eficiente en el uso de recursos de las Mutualidades (al detectar casos de compromisos leves y realizar mayores estudios solo en los casos pertinentes) y de instituciones de salud común (al asistir a médicos en la detección precoz de casos de modo que sean derivados a la mutualidad correspondiente).

Respecto de los resultados de los modelos de inteligencia artificial propuestos mirados en conjunto, se puede indicar que:

El sistema desarrollado en términos generales presentó resultados de Sensibilidad, Especificidad en un nivel sobre el 80%.

Respecto de los valores PPV y PNV que son valores pos test, donde la prevalencia juega un rol importante, el PPV logró un sólo 24.7% por lo cual se sugiere ante una detección positiva del signo, este debe ser analizado nuevamente con una herramienta más precisa como lo es un scanner y/o por un Radiólogo certificado por OIT, según lo determinen los protocolos de estudios de la Mutualidad. Por otro lado, el valor PNV es de un 99.9% lo que indica que con un alto grado de confianza no encontró el patrón de neumoconiosis en la radiografía y puede ser descartado con confianza sin necesidad de mayor gasto en estudios por parte de las Mutualidades.

Respecto de los indicadores de razón de verosimilitud (LR+, LR-) que son indicadores dicotómicos y en los cuales un LR+ mayor de 10 y un LR- menor que 0.1, indican un cambio importante en la probabilidad pretest, y con ellos determina una alta certeza en cambio de conducta clínica, el estudio indica que el LR+ obtuvo un valor de 6.6, por lo que cuando el modelo predice de manera positiva estos casos se sugieren siempre deben ser corroborados por un especialista de imagenología. Mientras para el caso de LR- este es menor que 0.1 dando a entender que tiene un error muy bajo al descartar un paciente sin el signo, pudiendo el clínico confiar en el resultado.

Todos los resultados son enmarcados en un intervalo de confianza del 95%.

El sistema clasificó radiografías erróneamente en un 14.3% del total (Falsos negativos + falsos positivos/Total de radiografías).

Finalmente mencionar los límites del modelo y recomendaciones para su uso, por ejemplo:

Al predecir que no tienen signos de neumoconiosis, pero el paciente trabaja en un ambiente de exposición revisar con cautela porque puede presentar neumoconiosis leve no detectada (baja cantidad de imágenes disponibles en los entrenamientos).

Al predecir que existe el signo de neuromoconiosis contrastar con clínica e historia del paciente, pues puede ser un falso positivo y mal interpretó una neumonía por efecto de una mala toma o un entramado pulmonar aumentado.

Con los resultados expuestos el sistema logra una alta sensibilidad enviando a lectores OIT cualquier signo de neuromoconiosis, por otro lado, al tener una especificidad relevante es capaz de descartar pacientes con otras patologías incluidas pacientes sanos. Se sugiere dar uso a la plataforma de manera generalizada y evaluar resultados en un ambiente clínico real, el modelo se reentrenará en el momento que se cuente con 1500 radiografías de neuromoconiosis, o unos 6 meses de implementación con el objetivo de mejorar los indicadores del modelo propuesto.

4 REFERENCIAS

- [ERPO1]-Enfermedades Respiratorias Ocupacionales, Mauricio Salinas F., Dr.a José A. Del Solar, Dr.b Revista Médica Clínica Las Condes, Volume 26, Issue 3, May 2015, Pages 357-366.
- [GCIOIT1]-OIT Guía para el uso de la Clasificación Internacional de la OIT de Radiografías de Neuromoconiosis Edición revisada 2011 Ginebra, Oficina Internacional del Trabajo, 2013 Serie Seguridad y Salud en el Trabajo, núm. 22. ISBN 978-92-2-324542-9 (pdf).
- [MP43] W. McCulloch and W. Pitts. A logical calculus of ideas immanent in nervous activity. Bulletin of Mathematical Biophysics, 5:127-147, 1943.
- [K56] S. C. Kleene. Representation of events in nerve nets and finite automata. In Automata Studies, pp. 3-41. Princeton University Press, 1956.
- [R58] F. Rosenblatt. The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain, Psychological Review, vol. 65, 6:386-408, 1958.
- [NYT58] Electronic Brain Teaches Itself. The New York Times, July 13, Page 9, 1958.
- [MP69] M. Minsky, S. A. Papert. Perceptrons, MIT press, 1969.
- [ACM19] ACM Announces 2018 Turing Award Recipients, March 27, 2019, <https://awards.acm.org/about/2018-turing>.
- [RHW86a] D. E. Rumelhart, G. E. Hinton, R. J. Williams. Learning Internal Representations by Error Propagation, Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Vol. 1: Foundations, 318-362, 1986.
- [RHW86b] D. E. Rumelhart, G. E. Hinton, R. J. Williams. Learning representations by back-propagating errors, Nature, 323:533-536, 1986.

-
- [KSH12] A. Krizhevsky, I. Sutskever, G. E. Hinton. ImageNet classification with Deep convolutional neural networks, *Advances in Neural Information Processing Systems*, 1106-1114, 2012.
 - [JMWYZB21] Jiashi Zhao, Mengmeng Li, Weili Shi, Yu Miao, Zhengang Jiang and Bai Ji. A deep learning method for classification of chest X-ray images, 2021. *J. Phys.: Conf. Ser.* 1848 012030.
 - [LeCun89] Y. LeCun. Generalization and Network Design Strategies. Tech Report, U. of Toronto, 1989.
 - [BL88] L. Bottou, Y. LeCun. SN: A simulator for connectionist models. *NeuroNimes*, 1988.
 - [MNIST] Y. LeCun, C. Cortes, C. Burges. MNIST handwritten digit database, <http://yann.lecun.com/exdb/mnist/>.
 - [BSF94] Y. Bengio, P. Simard, P. Frasconi. Learning long-term dependencies with gradient descent is difficult, *IEEE transactions on neural networks* 5 (2), 157-166.
 - [HS97] S. Hochreiter, J. Schmidhuber. Long Short-Term Memory, *Neural computation* 9(8), 1735-1780, 1997.
 - [GBB11] X. Glorot, A. Bordes, Y. Bengio. Deep sparse rectifier neural networks. *AISTATS* 2011.
 - [BChB15] D. Bahdanau, K. Cho, Y. Bengio. Neural Machine Translation by Jointly Learning to Align and Translate. *International Conference on Learning Representations*, 2015.
 - [G14] I. Goodfellow et. al. Generative Adversarial Nets, *Advances in Neural Information Processing Systems*, 2672-2680, 2014.