

Diagnóstico Automático de Tuberculosis en Radiografías de Tórax aplicando Aprendizaje Automático

Guanca, J. Daniel Antonio¹; Izetta Riera, C. Javier¹; Pérez Otero, Nilda M.¹; Verazay, Abigail N.¹

(1) Facultad de Ingeniería, Universidad Nacional de Jujuy.

dany12rp13@gmail.com; javierizetta@fi.unju.edu.ar; nilperez@fi.unju.edu.ar;

abigailverazay@hotmail.com

RESUMEN

En los últimos años, ha habido un creciente interés en el uso de técnicas de aprendizaje automático en el diagnóstico asistido por computadora, logrando resultados satisfactorios en la detección de algunas patologías y enfermedades como fracturas, fisuras, tendinopatías, cáncer y tuberculosis, entre otras. En particular, el diagnóstico automático de la tuberculosis a través de radiografías de tórax es de vital importancia, ya que permitiría la detección de esta enfermedad en una etapa temprana, aumentando las posibilidades de recuperación del paciente. El presente trabajo consiste en el análisis, diseño y prueba de un modelo que, mediante el uso de técnicas de aprendizaje automático, permite el diagnóstico de tuberculosis pulmonar a partir de imágenes radiográficas de tórax de pacientes de ambos sexos y diferentes edades.

ABSTRACT

In recent years, there has been a growing interest in using machine learning techniques in computer-aided diagnosis, achieving satisfactory results in detecting some pathologies and diseases such as fractures, fissures, tendinopathies, cancer and tuberculosis, among others. Particularly, the automatic diagnosis of tuberculosis through chest radiographs is of vital importance as it would allow the detection of this disease at an early stage, increasing the patient's chances of recovery. The present work consists of the analysis, design and testing of a model that, using machine learning techniques, allows for the diagnosis of pulmonary tuberculosis from chest radiographic images of patients of both sexes and different ages.

Palabras claves: Aprendizaje Automático - Diagnóstico Asistido por Computadora - Tuberculosis

Keywords: Machine Learning - Computer-aided Diagnosis- Tuberculosis

1. INTRODUCCIÓN

El continuo avance tecnológico permitió la evolución en el desarrollo de áreas como la visión artificial, antes limitada por las bajas prestaciones de hardware de la época. Hoy en día, dentro del campo de la visión artificial, las técnicas de aprendizaje automático están siendo aplicadas exitosamente para dar solución a diversas problemáticas (Bradski, 1998, Freeman *et al.*, 1996, Houet *et al.*, 2017).

Entre las técnicas de Aprendizaje Automático (Machine Learning, ML) se destacan las Máquinas de Soporte Vectorial (SVM, por sus

siglas en inglés), RandomForest (RF) y métodos de aprendizaje profundo como las Redes Neuronales Convolucionales (CNN, por sus siglas en inglés).

Los métodos SVM y RF son técnicas de aprendizaje supervisado que, mediante el entrenamiento a partir de un conjunto de datos con sus respectivas etiquetas, permiten detectar o predecir las etiquetas de los datos de entrada que no pertenecen al conjunto previo de entrenamiento, en otras palabras, son potentes clasificadores. El buen rendimiento que presentan estos clasificadores los convierte en una buena opción para la resolución de problemas de visión

artificial. Es decir, utilizar imágenes en el conjunto de datos de entrenamiento permite que estas técnicas resuelvan una amplia variedad de problemas de clasificación en la visión artificial (Calandria *et al.*, 2007; Bosch *et al.*, 2007). Por otro lado, las CNN son utilizadas exitosamente en problemas con imágenes ya que permiten realizar la extracción de las características de una imagen en forma automática. Esta extracción mejora la detección de objetos dentro de una imagen y su posterior clasificación a mayor velocidad y menor costo que métodos manuales. Es por esta razón que estas técnicas son muy utilizadas en imágenes médicas tales como el diagnóstico de enfermedades mediante tomografías, resonancias y radiografías, ya que brindan la posibilidad de detectar los patrones más importantes de enfermedades y lesiones que se manifiestan en las mismas permitiendo realizar un diagnóstico de estas patologías de forma automática con una precisión aceptable.

En los últimos años se ha profundizado el interés de utilizar técnicas de ML en los diagnósticos asistidos por computadora (CAD) logrando resultados satisfactorios en la detección de algunas patologías y enfermedades como fracturas, fisuras, tendinopatías, cáncer, Tuberculosis, entre otras. Particularmente el diagnóstico automático de tuberculosis mediante radiografías de tórax es de vital importancia ya que permitiría detectar esta enfermedad en una etapa temprana aumentando las posibilidades de recuperación del paciente.

El presente trabajo consiste en el análisis, diseño, y prueba de un modelo que, mediante el uso de técnicas de Aprendizaje Automatizado, permita el diagnóstico de la enfermedad tuberculosis pulmonar a partir de imágenes radiográficas de tórax de pacientes de ambos sexos y distintas edades.

1.1 Antecedentes

Existen varias aplicaciones de técnicas de aprendizaje automático en el diagnóstico automático de Tuberculosis en radiografías de tórax, a continuación, se mencionan las que se tomaron como guía y ayuda para el desarrollo del presente trabajo.

En el artículo de Stirenko *et al.* (2018), los autores proponen utilizar técnicas de segmentación y aumentación con y sin pérdida de datos cuando se

cuenta con conjuntos de datos pequeños (menos de 1000 imágenes). Con las técnicas propuestas los autores logran mejorar la eficiencia de la CNN. El caso de estudio de este trabajo fue el diagnóstico de Tuberculosis en radiografías médicas de tórax.

La propuesta de Lakhani y Sundara (2017) es evaluar el rendimiento de las CNN en la detección de Tuberculosis en radiografías de tórax. Para la evaluación, comparan el rendimiento de un modelo que fue entrenado previamente con aproximadamente un millón de imágenes diversas obtenidas de la base de datos ImageNet, contra el rendimiento de un modelo sin previo entrenamiento. Ambos modelos utilizan una arquitectura de dos redes muy conocidas: AlexNet y GoogleNet. Además, utilizan métodos de aumentación de datos llegando a la conclusión que el modelo preentrenado y con aumentación de datos produce un mejor rendimiento que el modelo no entrenado previamente.

Evalgelista y Guedes (2018) proponen entrenar y probar nueve arquitecturas CNN diferentes y dos ensembles de ellas para evaluar cual logra un mejor rendimiento en la detección de tuberculosis. Para entrenar las arquitecturas se usaron tres *datasets* diferentes y se destaca que no se utilizó ningún método de aumentación de datos. De las nueve arquitecturas CNN entrenadas se seleccionaron las tres que dieron mejores resultados en rendimiento y se construyó un ensemble con ellas. Luego, otro ensemble fue construido utilizando las nueve arquitecturas CNN entrenadas, incluidas las tres arquitecturas mencionadas previamente. Al comparar ambos ensembles se demuestra que el ensemble formado por las tres arquitecturas CNN de mejor rendimiento superó al ensemble formado por las nueve arquitecturas CNN propuestas.

En (Heo *et al.*, 2019) se demuestra que el uso de variables demográficas tales como género, edad, peso y altura mejoran significativamente el rendimiento de las CNN en la detección de tuberculosis. Para demostrarlo comparan el rendimiento de dos CNN, una que fue entrenada solo con las radiografías, y otra que fue entrenada con radiografías y con las variables demográficas correspondientes al paciente. Se llegó a la conclusión que a medida que se agregan variables demográficas el rendimiento de la red aumenta, para este estudio el uso de las variables

demográficas tales como peso, edad, género y altura, resultaron en un aumento del 2,88% en la precisión de la red.

En Dasanayaka et al. (2021) se presenta una red convolucional profunda, que consta de tres estructuras diferentes para optimizar, generar, segmentar y clasificar imágenes de radiografías de tórax. Se aplican aumentación de datos y ajuste de parámetros utilizando un algoritmo genético para mejorar los resultados diagnósticos. Con este método se logra una precisión del 97.1%.

Van Ginneken *et al.*, (2013) proponen utilizar un conjunto de subprocesos tales como normalización y segmentación de las imágenes, análisis de simetría, textura y forma del pulmón. Cada proceso genera un resultado independiente, los cuales son usados como características de entradas de un clasificador RandomForest, que fue entrenado con un conjunto de mil imágenes radiográficas normales y anormales. El clasificador genera como salida un puntaje global que indica que tan anormal es la imagen, mayor puntaje de la imagen indica mayor probabilidad de la presencia de tuberculosis en el paciente. Al analizar los resultados se llegó a la conclusión de que el rendimiento del sistema conformado por los subprocesos y el clasificador es cercano al de un radiólogo.

Fatiet. *al.* (2022) centran su trabajo en aplicar dos técnicas de inteligencia artificial, CNN y ANN. Proponen dos enfoques diferentes para diagnosticar la tuberculosis a partir de dos conjuntos de datos con radiografías de tórax. El primer enfoque combina dos modelos de CNN. Utilizan el algoritmo de análisis de componentes principales (PCA) para reducir la dimensionalidad. El segundo enfoque aplica redes neuronales artificiales (ANN) basadas en características fusionadas extraídas aplicando diferentes algoritmos.

2 DIAGNÓSTICO DE TUBERCULOSIS EN RADIOGRAFIAS DE TÓRAX

La tuberculosis es una enfermedad bacteriana contagiosa vía aérea que afecta principalmente a los pulmones. Es una de las enfermedades que más muertes causa, estando dentro de las 10 primeras causas de muerte en el mundo, superando al virus de inmunodeficiencia humana

(VIH o HIV, por sus siglas en inglés) y a la malaria (Evalgelista y Guedes, 2018).

El uso de radiografías de tórax para la detección temprana de tuberculosis es el primer método de aproximación diagnóstica, es realizado por radiólogos y consiste en analizar manualmente las radiografías de tórax del paciente en busca de la presencia de patrones como adenopatías, patrones miliare y derrames basales para el caso de tuberculosis primaria que se manifiesta en niños y adultos con VIH. Los patrones correspondientes a tuberculosis postprimaria que se pueden detectar en radiografías son condensación y/o excavación de lóbulos, como así también la reducción de aquellos, esta forma de la enfermedad es más común en adultos (Miranda et al. 2004). Un ejemplo de esto se puede apreciar en la Figura 1 donde se muestran las radiografías de dos pacientes. La radiografía de la Figura 1 (a) no presenta signos de tuberculosis, mientras que la Figura 1 (b) presenta signos de tuberculosis postprimaria (Lopes y Valiati, 2017).

Una ventaja del uso de imágenes radiográficas frente a otros métodos de detección que no las utilizan es que constituyen una de las vías para la detección de esta enfermedad en etapas tempranas. El diagnóstico temprano es crucial para controlar la enfermedad evitando su propagación, aumentando las posibilidades de recuperación del paciente y mejorando su calidad de vida. Además, el diagnóstico temprano es esencial para interrumpir la transmisión de la tuberculosis (Vallejo *et al.*, 2015). Se estima que por cada año sin tratamiento que pasa un paciente con esta enfermedad puede llegar a infectar entre 10 y 15 personas más (Cáceres-Manrique y Orozco-Vargas, 2008). Otra ventaja es que el diagnóstico de tuberculosis mediante radiografías de tórax es recomendable debido a su fácil acceso y bajo costo comparado con otros métodos de detección. Por ejemplo, algunos métodos de detección no basados en radiografías implican el diagnóstico microbiológico en el cual buscan la presencia de bacilos acidorresistentes (BAAR) en un frotis de esputo. Debido a que algunos bacilos acidorresistentes pueden no ser de tuberculosis la simple presencia de ellos no confirma la enfermedad. Por esta razón, es necesario realizar un cultivo adicional en busca de bacilos de tuberculosis, pero las sensibilidades de estos métodos son reducidas en pacientes con VIH por lo que se recomienda el diagnóstico de esta

enfermedad mediante imágenes radiográficas (Rohilla y Hooda, 2017).



(a)



(b)

Figura 1:(a)Radiografía sin signos de tuberculosis. (b) Radiografía que presenta signos de tuberculosis postprimaria.

3 DIAGNÓSTICO DE TUBERCULOSIS ASISTIDO POR COMPUTADORA

A pesar de las ventajas previamente mencionadas del diagnóstico de tuberculosis mediante radiografías, existe un gran obstáculo y es que el análisis de la radiografía necesita ser realizado manualmente por radiólogos entrenados. Esto genera, por un lado, que la precisión de este método dependa de la habilidad y experiencia en la interpretación del radiólogo (Rohilla y Hooda, 2017) y, por otro, la disponibilidad de estos radiólogos supone un gran problema en países de medios y bajos recursos en donde hay escasez de radiólogos para esta tarea (Liu *et al.*, 2017).

Una solución a este obstáculo es el uso de sistemas de diagnóstico asistidos por computadoras (CADs, por sus siglas en inglés) que permiten automatizar todo el proceso, lo que disminuye drásticamente los costos y el tiempo, acelerando el proceso de detección de casos activos. Además, éstos logran resultados

comparables y, en algunos casos, superiores al de los *datasets* formados por expertos en el diagnóstico de tuberculosis (Heo *et al.*, 2019).

Desde el punto de vista de la visión artificial, el CADs de tuberculosis se puede dividir en cuatro tareas: el preprocesamiento de las imágenes radiográficas, la segmentación de los pulmones, la extracción de características y el reconocimiento o detección de la enfermedad (Figura2).

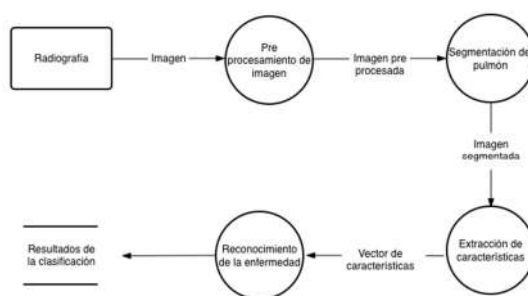


Figura 2: Esquema del proceso de diagnóstico asistido por computadora de tuberculosis.

En el preprocesamiento de las imágenes radiográficas se busca mejorar la calidad de las radiografías con el objetivo de resaltar sus características para así aumentar el rendimiento de las siguientes tareas como la extracción de características. Para este preprocesamiento se pueden aplicar métodos de corrección de contraste y brillo, reducción de ruido, filtros, entre otros. Detalles de estos métodos se pueden encontrar en los trabajos (Sherrier y Johnson, 1987; Kwan y Kwan, 2011; Savitha y Naveen, 2016).

La segmentación de las imágenes consiste en separar la región de interés de la imagen, pulmones en este caso, para así solo utilizar esta región. Esto mejora el rendimiento de los métodos de extracción de características, ya que no introducirá ruido de otras regiones no importantes para el diagnóstico de la enfermedad. Para el realizar el proceso de segmentación generalmente se utilizan máscaras, estas máscaras imitan el contorno de los pulmones y se superponen sobre la radiografía correspondiente (Stirenko *et al.*, 2018). Luego se procede a eliminar las regiones que se encuentran fuera del contorno de la máscara, quedando como resultado

únicamente los pulmones a analizar, como se muestra en la Figura 3.

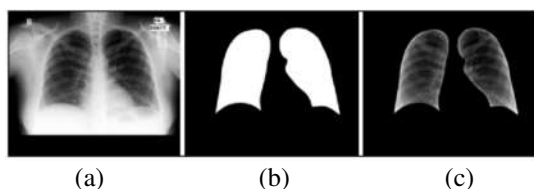


Figura 3: Ejemplos de una radiografía de tórax (a), la máscara correspondiente al pulmón (b), y el resultado de la imagen segmentada (c) (Stirenko *et al.*, 2018).

Con las imágenes preprocesadas y segmentadas se procede a realizar las tareas de extracción de características tanto globales como locales de la imagen segmentada. Para realizar esta tarea existen dos formas, una mediante métodos de análisis y otra utilizando las ventajas de las redes convolucionales.

En la literatura existen una gran variedad de métodos de análisis para la extracción de características en el diagnóstico de tuberculosis asistido por computadora, la más simple consiste en usar como características los valores de los píxeles de las imágenes originales. Otras alternativas son: el análisis del receso costodiafragmático, el análisis de textura, el análisis de la forma del pulmón y el análisis de la simetría (Fatiet. *al.* 2022; van Ginneken *et al.*, 2013; Cao *et al.*, 2016).

La extracción de características también puede realizarse utilizando redes convolucionales. Mediante éstas, las características de la imagen son automáticamente aprendidas por la red a través de las capas convolucionales, es decir que la extracción de características es realizada automáticamente (Rohilla y Hooda, 2017).

Finalmente, el reconocimiento de esta enfermedad se puede tomar como un problema de clasificación, en el que a partir de las características extraídas previamente se clasifica el resultado como positivo o negativo en la presencia de signos de tuberculosis. Para realizar esta tarea se pueden utilizar técnicas de aprendizaje automático como las SVM, RF y técnicas de aprendizaje profundo como las CNN. Estas técnicas son las más utilizadas debido al alto nivel de precisión y efectividad que logran en este tipo de tareas (Lakhani y Sundaram, 2017).

4 DISEÑO DEL MODELO

4.1 Preparación de los datos de entrada del modelo

Para la construcción del modelo se utilizó un conjunto de imágenes radiográficas de tórax con y sin evidencia de tuberculosis obtenidas del post “*PulmonaryChest X-RayAbnormalities*” de Kaggle(Mader, 2017).

El conjunto de datos consta de 662 imágenes radiográficas de tórax de personas de ambos sexos y diversas edades. De las 662 imágenes, 336 de ellas corresponden a casos con tuberculosis y las restantes (326) no presentan signos de tuberculosis. Todas las imágenes del conjunto de datos se encuentran en escala de grises, y las dimensiones de cada una se encuentran entre el rango de 1130×948 píxeles y 3001×3001 píxeles. Para determinar un subconjunto de datos que permita un correcto entrenamiento de los modelos de clasificación se debe realizar la segmentación del área de los pulmones en las imágenes, de esta manera se elimina la información irrelevante de los píxeles fuera de esta área. Antes de la segmentación se realizó un análisis y selección manual del conjunto de imágenes originales, es decir, se eliminaron aquellas imágenes en las que no era posible segmentar el área de los pulmones debido a que en las mismas los pulmones no estaban bien definidos (un ejemplo se puede observar en la Figura 4). La segmentación se realizó por medio del conjunto de máscaras brindadas en el conjunto de datos. Esta tarea se describe en apartado 4.3. Al terminar la selección de las imágenes se obtuvo un conjunto de datos de 564 imágenes para ser utilizadas en el entrenamiento de los clasificadores.



Figura 4: Radiografía de tórax del conjunto de datos original que no se incluyó en el entrenamiento.

4.2 Construcción del modelo

Para la construcción del modelo propuesto en este trabajo, las tareas fueron divididas en dos etapas claramente definidas. La primera etapa, llamada etapa de entrenamiento consiste en el entrenamiento de los clasificadores para poder distinguir entre una radiografía de tórax con tuberculosis de una sin tuberculosis y clasificarlas correctamente. Para realizar la primera etapa, fue necesario preparar el conjunto de datos sobre el cual se realiza el entrenamiento para obtener un mejor aprendizaje de los clasificadores. En este sentido se aplicaron técnicas bien conocidas del área de procesamiento de imágenes digitales como la segmentación y la reducción de las dimensiones de las imágenes. La segunda etapa de la construcción del modelo, llamada etapa de prueba consiste en la evaluación del desempeño de los clasificadores con un conjunto de radiografías nuevas no utilizadas durante la etapa de entrenamiento. Es importante aclarar que en cada etapa se utilizó un conjunto diferente de imágenes, para ello se aplicó la técnica *hold-out* (Weiss y Kulikowski, 1991) al conjunto de imágenes originales. De esta manera se obtuvo un subconjunto de datos correspondiente a la etapa de entrenamiento, que contiene el 80% de las imágenes, y otro subconjunto para la etapa de prueba del clasificador, que contiene el 20% restante de las imágenes. Se decidió por esta técnica por su facilidad de ejecución y por ser una de las técnicas más aplicadas.

En la Figura 5 es posible observar con más detalle los pasos básicos para la construcción del modelo propuesto. Cabe aclarar que estas etapas se corresponden a las etapas del esquema del proceso de diagnóstico de tuberculosis presentado en la Figura 2.

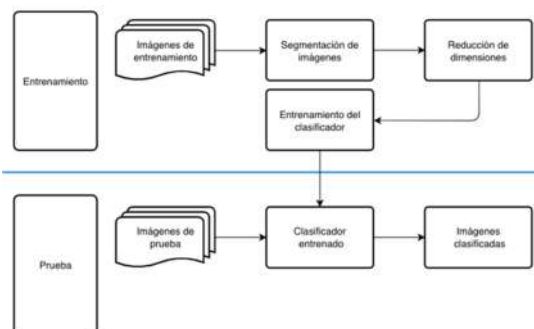


Figura 5: Esquema para la construcción del modelo.

4.3 Segmentación de las imágenes

El proceso de segmentación se realizó a partir de una colección de máscaras del conjunto de datos. Cada radiografía fue emparejada con su correspondiente máscara y procesada individualmente. El proceso consistió en superponer manualmente la máscara sobre la radiografía correspondiente, luego se asignó un valor de cero(0) a los píxeles comprendidos fuera del área de los pulmones y se conservó los valores de los píxeles dentro del área. En la Figura 3 se muestra una radiografía con su correspondiente máscara y el resultado final de aplicar el proceso de segmentación. El objetivo de aplicar este proceso es mejorar el aprendizaje de los clasificadores al eliminar las regiones que no son de importancia para el diagnóstico de tuberculosis.

4.4 Reducción de dimensiones

Las dimensiones del conjunto de imágenes segmentadas resultantes de la tarea anterior varían entre el rango de 1130×948 píxeles y 3001×3001 píxeles. Esto implica contar con un gran poder de cómputo para la siguiente tarea, el entrenamiento. Es por este motivo que se decidió reducir las dimensiones, tarea que se debe hacer tratando de conservar la mayor información posible. Cuando se aplican transformaciones a las imágenes, como la reducción de sus dimensiones, se presentan cambios de posición de los píxeles respecto a su posición inicial en el plano y por lo tanto, se requieren métodos de interpolación numérica para los valores que se asignarán a las nuevas posiciones. Al interpolar, se estima el valor más apropiado para cada uno de los nuevos píxeles (Melo, 2009). Existen diversas técnicas de interpolación entre las que se encuentra: vecino más cercano (Jiang y Wang, 2015), interpolación bilineal (Quezada Moncayo y Mera Collantes, 2009) e interpolación bicúbica (Acea Mena, 2012), entre otras. Esta última es la que se decidió aplicar en este trabajo. La interpolación bicúbica se puede lograr utilizando los polinomios de Lagrange, *splines* cúbicos o el algoritmo de convolución cúbica. Con interpolación bicúbica el valor del píxel de salida es un promedio ponderado del valor de los 16 píxeles en la vecindad 4×4 más cercana, más detalles de esta técnica se pueden encontrar en (Acea Mena, 2012). Se utilizó el método de interpolación bicúbica porque este genera una superficie

interpolada más suave que los demás métodos, manteniendo el detalle de una imagen remuestrada. Este método tampoco sufre el problema del efecto de salto y proporciona el menor emborronamiento.

Al finalizar el proceso de reducción, el conjunto de datos quedó conformado por 564 imágenes donde la dimensión de cada una de ellas es de 96×99 píxeles.

4.5 Entrenamiento y prueba del clasificador

Para el entrenamiento del modelo se utilizaron tres métodos para la clasificación: SVM, RF y CNN. De esta manera se obtienen tres variantes del modelo. El objetivo de entrenar los clasificadores es encontrar los valores más adecuados para cada método en los distintos subconjuntos de entrenamiento, esto consiste en la variación de ciertos parámetros. Los parámetros a variar están definidos por el método de clasificación utilizado. Para SVM, los parámetros a variar son la función de *kernel* y el costo. En RF el parámetro a variar es el número de árboles usados. Y, por último, para CNN el número de neuronas por capa, el número de capas ocultas, las repeticiones y el *learningrate*.

Para una estimación más precisa del desempeño, el entrenamiento y la prueba de SVM y RF se repiten 30 veces y se calcula su error promedio. Mientras que, para CNN, el entrenamiento y prueba se repiten 5 veces debido al costo computacional que supone cada entrenamiento, luego se calcula el error promedio como en SVM y RF.

Al finalizar el entrenamiento se selecciona la variante con el clasificador cuyo error promedio de clasificación fue el más bajo. Esta variante formará parte de la versión final del modelo propuesto.

5 PRUEBA Y ANÁLISIS DE RESULTADOS

5.1 Configuración experimental

El desarrollo de las experimentaciones se realizó mediante los pasos ilustrados en el diagrama de la Figura 6. Este proceso implica aplicar la técnica *Hold-outs* sobre el conjunto de datos, en un número de iteraciones representado por “X”. El valor de “X” corresponde a 30 (treinta) iteraciones para las

configuraciones de las variantes con SVM y RF. Mientras que para las configuraciones con CNN el valor de “X” fue de 5 (cinco). El objetivo de esto fue calcular el error promedio de clasificación por cada configuración que se evaluó de manera de obtener una estimación más certera del mismo.

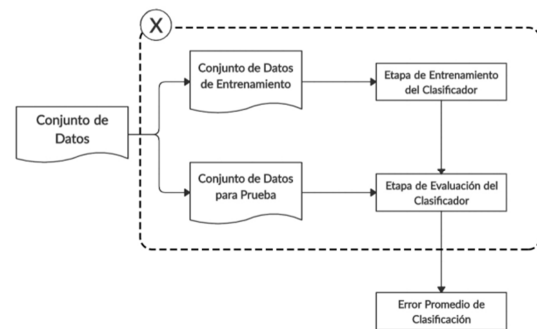


Figura 6: Diagrama de configuración experimental.

5.2 Ajustes de parámetros para SVM

Los parámetros ajustados en SVM fueron el *kernel* y costo. Se utilizaron dos configuraciones de *kernel* distintas, radial y lineal. Además, para cada configuración de *kernel* se utilizaron distintos valores para el parámetro costo siendo estos 1, 10 y 100.

En la Tabla 1 se expone un resumen de los errores promedio de clasificación obtenidos a partir de 30 corridas de las experimentaciones realizadas sobre el conjunto de datos de prueba con cada configuración.

Tabla 1. Resultados de SVM con *kernel* lineal, radial y distintos valores de costo.

| <i>Kernel</i> | Costo | Error promedio de clasificación |
|---------------|-------|---------------------------------|
| Lineal | 1 | 0.3142424 |
| | 10 | 0.3142424 |
| | 100 | 0.3142424 |
| Radial | 1 | 0.2818182 |
| | 10 | 0.2472727 |
| | 100 | 0.2839394 |

Se puede observar en la Tabla 1, que el valor de costo en la configuración con *kernel* lineal no influye en el error promedio de clasificación alcanzado. Mientras que para la configuración con *kernel* radial, el error promedio de

clasificación varía con respecto al valor de costo, obteniendo la menor tasa de error entre las configuraciones evaluadas con un valor de costo igual a 10. Si el costo es pequeño, los márgenes serán estrechos y muy pocas observaciones podrán estar en el lado incorrecto del mismo, esto equivale a un modelo bastante bien ajustado a los datos. A medida que se incrementa el costo, mayor es la tolerancia a las clasificaciones incorrectas sobre el margen, con lo que el margen será más ancho y habrá más vectores soporte, esto equivale a un modelo más flexible y evitamos el sobreajuste. Por otro lado, podemos observar que se alcanzó tasas de errores más bajas con un *kernel* radial. Esto puede deberse a que la no linealidad de los datos es alta, por lo que al usar un *kernel* radial posibilita separar los datos de forma más eficiente lo cual beneficia el desempeño del modelo.

5.3 Ajustes de parámetros para RF

Uno de los parámetros principales del método RF es el número de árboles de decisión que se utilizarán dentro del ensamble de clasificación. El número de árboles de decisión va a depender de la complejidad del problema. Generalmente, a mayor número de árboles se obtienen mejores resultados, pero requiere mayor costo computacional y mayor tiempo de entrenamiento. El ajuste de este parámetro se realiza mediante la experimentación.

En la Tabla 2 se exponen los resultados de las experiencias realizadas con RF con una configuración de 1, 10, 100, 1.000 y 10.000 árboles de decisión. Se muestra el error promedio de clasificación al realizar 30 corridas por cada configuración sobre el conjunto de datos de prueba.

Tabla 2. Resultados de RF con 1, 10, 100, 1.000 y 10.000 árboles de decisión.

| Número de árboles | Error promedio de clasificación |
|-------------------|---------------------------------|
| 1 | 0,3469697 |
| 10 | 0,2895758 |
| 100 | 0,2306515 |
| 1.000 | 0,2197497 |
| 10.000 | 0,2160022 |

Se puede observar en la Tabla 2 que el error promedio de clasificación disminuye a medida

que el número de árboles de decisión aumenta. Este comportamiento se debe a que RF utiliza diferentes conjuntos de ejemplos y atributos para la construcción de cada árbol de decisión. Entonces, cada conjunto de árboles aprende una parte diferente del problema por lo que, al aumentar la cantidad de árboles, estos pueden cubrir más partes del problema, lo que resulta en una mejora de la capacidad de clasificación.

Los mejores resultados corresponden a utilizar 1.000 y 10.000 árboles de decisión cuya diferencia en los errores promedios es mínima. Sin embargo, existe una gran diferencia en el tiempo de entrenamiento entre ambas configuraciones. Para la configuración con 1.000 árboles de decisión el tiempo de entrenamiento fue de aproximadamente una hora. Mientras que para la configuración con 10.000 árboles de decisión el tiempo de entrenamiento fue mucho mayor, aproximadamente 12 horas. Por lo que se decidió utilizar 1.000 árboles para RF en este problema en particular.

5.4 Ajustes de parámetros para CNN

Debido a la complejidad de las CNN, existen una extensa cantidad de parámetros que pueden configurarse para influir en el rendimiento del entrenamiento. Para este modelo en particular los parámetros que se ajustaron son, el número de neuronas por capa, el número de capas, el *learningrate* y las iteraciones o *epochs*.

Si bien existe la idea de que con más capas y neuronas se puede mejorar las predicciones en problemas complejos, no existe una regla general para determinar la arquitectura de una CNN. El mejor ajuste se debe encontrar mediante la experimentación y no siempre es aquel con mayor número de capas y neuronas. Otro parámetro importante es el *learningrate*, mientras mayor sea el valor del parámetro menos estable se vuelve el modelo por lo tanto el resultado se ve afectado. Si se utiliza valores pequeños de *learningrate* el resultado mejorará, pero tardará en converger por lo que se debe encontrar un equilibrio entre estos dos aspectos. Por último, se ajustó el parámetro *epochs* que especifica el número de iteraciones a realizar sobre todo el conjunto de entrenamiento. En todas las experimentaciones de esta variante se calculó el error de clasificación promedio de 5 corridas.

5.4.1 Evaluación de las arquitecturas

En primer lugar, se evaluaron nueve arquitecturas diferentes para determinar cuál de ellas obtiene el mejor desempeño en este problema. Para ello se evaluaron diferentes cantidades de capas y neuronas manteniendo constante los demás parámetros.

En la Tabla 3 se exponen los resultados de las experiencias realizadas sobre CNN para las distintas arquitecturas. Los números comprendidos en los paréntesis corresponden a la cantidad de neuronas de cada capa. Por ejemplo, la primera arquitectura contiene cuatro capas de las cuales la primera contiene 190 neuronas, la segunda 63 neuronas, la tercera 21 neuronas y última capa 7 neuronas. Además, en esta tabla se muestran los errores promedio de clasificación de cinco corridas por cada configuración sobre el conjunto de datos de prueba.

Tabla 3. Resultados de CNN con distintas arquitecturas.

| Arquitectura | Error promedio de clasificación |
|------------------------|---------------------------------|
| $A_1(190,63,21,7)$ | 0,2905457 |
| $A_2(380,126,42,14)$ | 0,2687648 |
| $A_3(570,189,63,21)$ | 0,2618187 |
| $A_4(190,63,21,7,3)$ | 0,264738 |
| $A_5(380,126,42,14,6)$ | 0,2960904 |
| $A_6(570,189,63,21,9)$ | 0,2783988 |
| $A_7(190, 63, 21)$ | 0,2602803 |
| $A_8(380,126,42)$ | 0,2686285 |
| $A_9(570,189,63)$ | 0,2654643 |

Se puede observar en la Tabla 3, que el error promedio de clasificación más bajo corresponde a la arquitectura A_7 que es la más simple, la cual está formada por tres capas, donde la primera capa tiene 190 neuronas, la segunda 63 y la última 21 neuronas. Por lo que esta arquitectura será la elegida para las siguientes experimentaciones de este del modelo.

5.4.2 Ajuste de los parámetros *epochs* y *learningrate*

En estas experimentaciones se evaluaron distintos valores para los parámetros *epochs* y *learningrate* sobre la arquitectura A_7 . Los resultados de las experiencias realizadas se ilustran en la Tabla 4. Se muestra el error promedio de clasificación al realizar cinco corridas por cada configuración sobre el conjunto de datos de prueba.

Tabla 4. Resultados de CNN con arquitectura A_7 y distintas configuraciones para los parámetros *epochs* y *learningrate*.

| <i>Epochs</i> | <i>LearningRate</i> | Error promedio de clasificación |
|---------------|---------------------|---------------------------------|
| 50 | 0,05 | 0,2606039 |
| | 0,005 | 0,2506185 |
| | 0,0005 | 0,2478248 |
| 100 | 0,05 | 0,2586947 |
| | 0,005 | 0,2637409 |
| | 0,0005 | 0,2744089 |

De los errores promedio de clasificación ilustrados en la Tabla 4, el menor corresponde a un valor de 50 para *epochs* y un *learningrate* de 0,0005. Por lo que podemos concluir que, a partir de las configuraciones experimentadas anteriormente, la que mejor configuración es aquella con la arquitectura A_7 , *learningrate* 0,0005 y 50 *epochs*.

5.5 Comparación de las variantes del modelo propuesto

En este apartado se realiza una comparación de los mejores resultados en cuanto a desempeño/rendimiento de todas las variantes del modelo con el fin de determinar aquella que obtenga el mejor resultado de todas. En la Tabla 5 se presenta un resumen de los errores promedios correspondientes a dichas variantes.

Tabla 5. Resultados de las mejores configuraciones de las variantes del modelo.

| Variante | Error promedio de clasificación |
|--|---------------------------------|
| SVM con <i>kernel</i> radial y $C = 10$ | 0,2472727 |
| + | 0,2197497 |
| CNN con arquitectura A_7 , <i>epochs</i> = 50 y <i>learningrate</i> = 0,0005 | 0,2478248 |

Se puede observar en la Tabla 5, que la variante con RF es la que obtuvo el menor error promedio de clasificación. Por lo que se puede concluir que para este problema en particular y en comparación con las distintas configuraciones de las demás variantes mencionadas previamente, la variante con RF obtuvo mejores resultados. Esto podría deberse a que SVM y CNN generalmente necesitan un ajuste de parámetros más

exhaustivo. Además, las CNN necesitan una gran cantidad de imágenes para obtener un mejor rendimiento. Por lo que la variante con RF formará parte de la versión final del modelo propuesto en el presente trabajo.

5.6 Sensibilidad y especificidad

Existen dos medidas muy utilizadas para evaluar sistemas de diagnósticos asistidos por computadora, la sensibilidad y la especificidad. En este caso de estudio la sensibilidad es el porcentaje de personas que tienen tuberculosis y dan positivas con el modelo. Mientras que la especificidad es el porcentaje de personas que no tienen la enfermedad y que el modelo clasifica como negativos.

El modelo con RF obtuvo como porcentaje de sensibilidad el valor de 77%. Esto significa en términos numéricos, que de 100 personas que presenten signos de la enfermedad tuberculosis, el sistema detectaría como casos positivos verdaderos a 77 personas y clasificaría como falsos positivos a 23 personas. Mientras que el porcentaje de especificidad obtenido es del 80%. Lo que resulta en que de 100 personas que no presentan la enfermedad, el sistema clasificaría 80 personas como negativas para tuberculosis y las restantes 20 personas las clasificaría como falsos negativos para tuberculosis.

Estos resultados son bastantes prometedores considerando lo problemático que es diagnosticar la enfermedad mediante imágenes radiográficas por lo que se considera que el modelo cumplió con las expectativas ya que realiza la clasificación con una precisión aproximada a la de los trabajos consultados.

5.7 Comparación del modelo y el algoritmo de Stirenko y colaboradores

En el trabajo de Stirenko *et al.* (2018) se propone un algoritmo basado en redes convolucionales junto con métodos como la segmentación y la aumentación con y sin pérdida de datos. Esta propuesta tiene la desventaja de demandar un gran poder de cómputo ya que utiliza una red convolucional con una arquitectura compleja. Esto implica contar con un equipamiento costoso para poder implementarla. Con la intención de lograr resultados comparables e incluso mejorarlos, en el presente trabajo se utilizó el

mismo conjunto de imágenes y máscaras propuestas por Stirenko *et al.* (2018), pero sin aumentación de datos y utilizando métodos más simples de *machinelearning* que puedan ser implementados en computadoras convencionales. Las experimentaciones realizadas permitieron demostrar que el modelo con RF propuesto obtuvo una exactitud (*accuracy*) 0,8. Mientras que el modelo de Stirenko *et al.* (2018) obtuvo una exactitud de 0,7. Por lo que se puede concluir que se logró una mejora significativa en los resultados obtenidos con respecto a las CNN evaluadas en el trabajo mencionado. Esto podría deberse a que RF al no necesitar una gran cantidad de imágenes en su entrenamiento logra acercarse a una mejor solución.

6 CONCLUSIONES

En este trabajo se presentó la construcción de un modelo capaz de diagnosticar Tuberculosis de forma automática a través de imágenes radiográficas de tórax. El modelo desarrollado posee tres variantes que corresponden a la aplicación de distintas técnicas de *machine learning* tales como: SVM, RFy CNN. Las experimentaciones realizadas permitieron concluir lo siguiente:

- La utilización de un *kernel* radial mejoró el desempeño del modelo con SVM. Debido a que este tipo de *kernel* es más adecuado en problemas no lineales.
- El tiempo de entrenamiento fue decisivo al momento de definir la cantidad de árboles en RF debido a que el mismo fue mucho menor y la diferencia de la tasa de error no fue significativa al usar 1.000 o 10.000 árboles.
- Para este caso en particular la arquitectura más simple de CNN presentó la menor tasa de error. Consiste en tres capas donde la primera consta de 190 neuronas, la segunda de 63 neuronas y la última capa de 21 neuronas.
- El estudio comparativo entre las tres variantes del modelo propuesto en el presente trabajo, permitió concluir que con RF se obtiene el mejor desempeño para este problema en particular. Por lo que este método de clasificación fue el elegido para versión final del modelo.
- Se obtuvo un valor de sensibilidad del 77% y un valor de especificidad del 80% con la versión

final del modelo. Estos valores reflejan el buen desempeño del clasificador RF para este problema, ya que corresponden a valores aproximados a los obtenidos en trabajos similares.

- En base a las experimentaciones realizadas, se puede concluir que el modelo propuesto en este trabajo mejora significativamente los resultados de la propuesta de Stirenko *et al.* (2018).
- Se demostró que utilizando un método simple de *machinelearning*, capaz de correr en un sistema convencional, se obtienen resultados comparables con los existentes en el estado del arte sin la necesidad de la aplicación de técnicas sofisticadas de aumentación de datos y extracción de características constituyendo esto el principal aporte del trabajo.

7 TRABAJOS FUTUROS

Existen muchas vías de investigación para continuar este trabajo, por un lado, están las que contemplan aspectos a mejorar:

- Realizar un ajuste de parámetros más profundo de los métodos de clasificación, en especial de la variante con CNN en la cual se podría experimentar con numerosas arquitecturas en busca de reducir la tasa de error.
- Evaluar el desempeño de todas las variantes del modelo al aplicar técnicas de preprocesamiento a las imágenes como, aumentación de datos, filtros, entre otros.
- Aplicar técnicas de extracción de características como el análisis de textura, el análisis de la forma del pulmón y el análisis de la simetría para las variantes con SVM y RF.

Por otro lado, algunas direcciones en las que se puede extender el presente trabajo a manera de continuar contribuyendo con el desarrollo de la investigación son:

- Utilizar otro tipo de imágenes para el diagnóstico de Tuberculosis como las resonancias magnéticas las cuales presentan más información.
- Extender el modelo a otros problemas de enfoque similar como por ejemplo el diagnóstico de cáncer de pulmón mediante radiografías o incluso tomografías.

8 REFERENCIAS

- Acea Mena, L. (2012). *Efectos de la interpolación sobre la calidad de la segmentación empleando la transformada watershed* (Doctoral dissertation, Universidad Central "Marta Abreu" de Las Villas).
- Bosch, A., Zisserman, A., & Munoz, X. (2007, October). Image classification using random forests and ferns. In *2007 IEEE 11th international conference on computer vision* (pp. 1-8). Ieee.
- Bradski, G. R. (1998). Computer vision face tracking for use in a perceptual user interface.
- Cáceres-Manrique, F. D. M., & Orozco-Vargas, L. C. (2008). Demora en el diagnóstico de tuberculosis pulmonar en una región de Colombia. *Revista de Salud Pública*, *10*, 94-104.
- Calandria, D. J., Cañas, A., & Díaz, A. F. (2007). Detección de rostros humanos mediante SVM y generación de fotografías de tipo carné. In *Actas del Simposio de Inteligencia Computacional SICO'2007*.
- Cao, Y., Liu, C., Liu, B., Brunette, M. J., Zhang, N., Sun, T., ... & Curioso, W. H. (2016, June). Improving tuberculosis diagnostics using deep learning and mobile health technologies among resource-poor and marginalized communities. In *2016 IEEE first international conference on connected health: applications, systems and engineering technologies (CHASE)* (pp. 274-281). IEEE.
- Dasanayaka, C., & Dissanayake, M. B. (2021). *Deep learning methods for screening pulmonary tuberculosis using chest X-rays*. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, *9*(1), 39-49.
- Evalgelista, L. G. C., & Guedes, E. B. (2018, October). Computer-Aided Tuberculosis Detection from Chest X-Ray Images with Convolutional Neural Networks. In *Anais do XV Encontro Nacional de Inteligência Artificial e Computacional* (pp. 518-527). SBC.
- Fati, S. M., Senan, E. M., & ElHakim, N. (2022). *Deep and hybrid learning technique for early detection of tuberculosis based on X-ray images using feature fusion*. *Applied Sciences*, *12*(14), 7092.
- Freeman, W. T., Tanaka, K. I., Ohta, J., & Kyuma, K. (1996). Computer vision for computer games. *fg*, *96*, 100-105.

- Heo, S. J., Kim, Y., Yun, S., Lim, S. S., Kim, J., Nam, C. M., ... & Yoon, J. H. (2019). Deep Learning Algorithms with Demographic Information Help to Detect Tuberculosis in Chest Radiographs in Annual Workers' Health Examination Data. *International journal of environmental research and public health*, 16(2), 250.
- Hou, R., Chen, C., & Shah, M. (2017). Tube convolutional neural network (T-CNN) for action detection in videos. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 5822-5831).
- Jiang, N., & Wang, L. (2015). Quantum image scaling using nearest neighbor interpolation. *Quantum Information Processing*, 14(5), 1559-1571.
- Kwan, B. Y., & Kwan, H. K. (2011). Improved lung nodule visualization on chest radiographs using digital filtering and contrast enhancement. *World AcadSciEngTechnol*, 110, 590-3.
- Lakhani, P., & Sundaram, B. (2017). Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology*, 284(2), 574-582.
- Liu, C., Cao, Y., Alcantara, M., Liu, B., Brunette, M., Peinado, J., & Curioso, W. (2017, September). TX-CNN: Detecting tuberculosis in chest X-ray images using convolutional neural network. In *2017 IEEE International Conference on Image Processing (ICIP)* (pp. 2314-2318). IEEE.
- Lopes, U. K., & Valiati, J. F. (2017). Pre-trained convolutional neural networks as feature extractors for tuberculosis detection. *Computers in biology and medicine*, 89, 135-143.
- Mader, K. M. (2017). Pulmonary Chest X-Ray Abnormalities. Recuperado 2 noviembre, 2019, de <https://www.kaggle.com/kmader/pulmonary-chest-xray-abnormalities>
- Melo, S. B. (2009). Transformaciones geométricas sobre imágenes digitales. *Facultad de Ciencias-Carrera de Matemáticas. Universidad Distrital Francisco José de Caldas*.
- Miranda, G., Díaz, J. C., Arancibia, P., Antolini, M., Díaz, C., & Vidal, A. (2004). Manifestaciones radiográficas de la tuberculosis pulmonar. *Revista chilena de radiología*, 10(4), 178-182.
- Quezada Moncayo, C., & Mera Collantes, M. (2009). *Programa demostrativo del escalamiento de una imagen monocromática usando los métodos de replicación de píxeles (filtro de cajas) e interpolación bilineal (filtro bartlett)* (Bachelor's thesis).
- Rohilla, A., Hooda, R., & Mittal, A. (2017). TB detection in chest radiograph using deep learning architecture. In *Proceeding of 5th International conference on Emerging Trends in Engineering, Technology, Science and Management (ICETETSM-17)* (Vol. 136147).
- Savitha, S. K., & Naveen, N. C. (2016, March). Algorithm for pre-processing chest-x-ray using multi-level enhancement operation. In *2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)* (pp. 2182-2186). IEEE.
- Sherrier, R. H., & Johnson, G. A. (1987). Regionally adaptive histogram equalization of the chest. *IEEE transactions on medical imaging*, 6(1), 1-7.
- Stirenko, S., Kochura, Y., Alienin, O., Rokovyi, O., Gordienko, Y., Gang, P., & Zeng, W. (2018, April). Chest X-ray analysis of tuberculosis by deep learning with segmentation and augmentation. In *2018 IEEE 38th International Conference on Electronics and Nanotechnology (ELNANO)* (pp. 422-428). IEEE.
- Vallejo, P., Searle, A., & Farga, V. (2015). Ensayo Xpert MTB/RIF en el diagnóstico de tuberculosis. *Revista chilena de enfermedades respiratorias*, 31(2), 127-131.
- van Ginneken, B., Philipsen, R. H. H. M., Hogeweg, L., Maduskar, P., Melendez, J. C., Sánchez, C. I., ... & Adetifa, I. M. (2013). Automated scoring of chest radiographs for tuberculosis prevalence surveys: a combined approach. In *Fifth International Workshop on Pulmonary Image Analysis* (pp. 9-19).
- Weiss, S., & Kulikowski, C. (1991). Computer systems that learn.