

Aplicación de técnicas de inteligencia artificial para la detección de tuberculosis pulmonar en Colombia



Revista EIA
ISSN 1794-1237
e-ISSN 2463-0950
Año XIX/ Volumen 20/ Edición N.39
Enero-Junio de 2023
Reia3909 pp. 1-23

Publicación científica semestral
Universidad EIA, Envigado, Colombia

PARA CITAR ESTE ARTÍCULO / TO REFERENCE THIS ARTICLE /

Bedoya-Leiva, O.; Guarín-Aristizábal H.; Agudelo, J.

Aplicación de técnicas de inteligencia artificial para la detección de tuberculosis pulmonar en Colombia
Revista EIA, 20(39), Reia3909.
pp. 1-23.
<https://doi.org/10.24050/reia.v20i39.1617>

✉ Autor de correspondencia:

Bedoya-Leiva, O. (Oscar)
Ingeniería de sistemas. Maestría en Ingeniería. Doctorado en Ciencias de la computación.
Universidad del Valle
Correo electrónico:
oscar.bedoya@correounivalle.edu.co

Recibido: 19-07-2022

Aceptado: 12-09-2022

Disponibile online: 01-01-2023

✉ OSCAR FERNANDO BEDOYA-LEIVA¹

HARRY SANTIAGO GUARÍN-ARISTIZÁBAL¹

JARED AGUDELO²

1 Universidad del Valle
2 Universidad Libre Seccional Cali

Resumen

La tuberculosis es una enfermedad respiratoria que afecta los pulmones y que es causada por el bacilo *Mycobacterium tuberculosis* (MTB), el cual se expande cuando una persona enferma con tuberculosis expulsa la bacteria al aire al toser. Antes de la pandemia del COVID-19, la tuberculosis era considerada la principal causa de muerte por agentes infecciosos en el mundo incluso por encima del VIH/SIDA. El cultivo de MTB en medio sólido es el principal método de referencia diagnóstica. Además, es el método estándar para identificar el perfil de resistencia bacteriana. Sin embargo, se requiere de mucho tiempo para la obtención del resultado; tiempo durante el cual un paciente puede ser altamente contagioso. Por otro lado, la demora en instaurar un tratamiento para la tuberculosis puede conducir a mayor gravedad de la enfermedad, pero de tomarse una medida de sobretratamiento independiente de la confirmación diagnóstica, se generaría resistencia bacteriana, mayor tasa de eventos adversos y mayores costos. Por lo tanto, es necesaria una estrategia que permita hacer un diagnóstico rápido de la tuberculosis. Para abordar este reto, en este artículo se proponen modelos para la detección de tuberculosis pulmonar utilizando diferentes técnicas de inteligencia artificial. Estos modelos se pueden usar como apoyo a la toma de decisiones por parte de los médicos y tienen como objetivo identificar si un paciente padece de tuberculosis. En particular, se proponen modelos basados en cuatro técnicas de aprendizaje supervisado (redes neuronales, árboles de decisión, y dos métodos de ensamble) que permiten realizar un diagnóstico, positivo o negativo, de tuberculosis pulmonar a partir de unas variables de entrada y de diagnósticos anteriormente registrados de pacientes sanos y otros con tuberculosis pulmonar de la ciudad de Cali, Colombia. De acuerdo con los resultados obtenidos, el método de ensamble Extra Trees resulta ser el más exacto comparado con las otras técnicas utilizadas para la predicción de tuberculosis pulmonar alcanzando un área bajo la curva ROC de 95,63%.

Palabras clave: Aprendizaje supervisado; Árboles de decisión; Inteligencia artificial; Métodos de ensamble; Tuberculosis; Redes neuronales.

Application of artificial intelligence techniques for the detection of pulmonary tuberculosis in Colombia

Abstract

Tuberculosis is a respiratory disease that affects lungs and it is caused by the bacillus *Mycobacterium tuberculosis* (MTB), which is spread when people who are sick with tuberculosis expel bacteria into the air by coughing. Before the coronavirus (COVID-19) pandemic, tuberculosis was the leading cause of death from infectious agents even ranking above HIV/AIDS. Growing MTB on solid medium is the main diagnostic reference method. In addition, this is the standard method for identifying the bacterial resistance profile. However, the waiting period for the results is long, during which time a patient can be highly contagious. In turn, delaying treatment for tuberculosis can increase disease severity, but treating without diagnostic confirmation can lead to bacterial resistance, a higher rate of adverse events, and higher costs. Therefore, pulmonary tuberculosis disease must be rapidly diagnosed by a cost-effective method. This paper proposes models for the detection of pulmonary tuberculosis by using different artificial intelligence techniques. These models can be used to support decision making of doctors and are intended to identify whether a patient has tuberculosis or not. In particular, four supervised learning techniques are used (neural networks, decision trees, and two ensemble methods). Each model allows predicting a positive or negative diagnosis of pulmonary tuberculosis based on previously recorded diagnostic variables taken from patients in Cali, Colombia. According to the results, the Extra Trees method reaches the highest accuracy compared to the other techniques used for the prediction of pulmonary tuberculosis with an area under the ROC curve of 95,63%.

Keywords: Artificial intelligence; Ensemble methods; Decision trees, Neural networks; Supervised learning; Tuberculosis.

1. Introducción

La tuberculosis es una enfermedad infecciosa de cualquier órgano producido por el *Mycobacterium Tuberculosis* (MTB) que se transmite casi exclusivamente por vía aérea entre seres humanos. Esta enfermedad es una de las principales causas de muerte en el mundo y presenta gran dificultad para su detección temprana debido a que los métodos de identificación poseen algunas limitaciones que impiden su fácil implementación. Para el año 2021, la Organización Mundial de la Salud (World Health Organization, WHO) estimó una incidencia de infección por MTB de aproximadamente 10 millones. De estos, 1,2

millones fallecieron con estado VIH negativos y 208 mil VIH positivos (WHO, 2021). De esta forma, la tuberculosis se convierte en una de las 10 principales causas de muerte en el mundo, mayor incluso al VIH/SIDA. Además, debido a los mecanismos de patogenicidad del MTB, este puede permanecer en estado latente en los seres humanos. Se estima que alrededor de 1.700 millones de personas, el 23% de la población mundial, tienen una infección tuberculosa latente y, por lo tanto, presentan riesgo de desarrollar una enfermedad tuberculosa activa. La tuberculosis toma mayor impacto en países subdesarrollados o en vía de desarrollo, debido no sólo a los factores socioeconómicos que favorecen la transmisibilidad de la enfermedad, sino también a la dificultad metodológica para la detección del MTB.

La tuberculosis es prevenible y curable. Alrededor del 85% de las personas que tienen tuberculosis pueden ser tratados de forma exitosa con un régimen de medicamentos de 1 a 6 meses. Cada año se diagnostican y tratan eficazmente millones de personas con tuberculosis. Sin embargo, según Kitonsa et al. (2020) aún siguen existiendo aspectos por mejorar en la detección y el tratamiento temprano debido no sólo a factores socioeconómicos que afectan las zonas más vulnerables y que favorecen la transmisibilidad de la enfermedad, sino también a la dificultad metodológica para la detección del MTB. Lo anterior se debe principalmente a que los métodos de identificación comúnmente utilizados poseen algunas limitaciones que impiden su fácil implementación tales como problemas de rendimiento diagnóstico, tiempo en obtención de resultados, personal altamente calificado o tecnologías de alta complejidad que implican costos elevados (Joshi, 2018). Dicho lo anterior, se evidencia la importancia de obtener un diagnóstico rápido con una metodología de costo efectiva que permita determinar si una persona sufre, o no, de tuberculosis pulmonar. La detección de esta enfermedad se ha abordado usando diversas técnicas de inteligencia artificial sobre conjuntos de datos recolectados de pacientes con sospecha de tuberculosis para así obtener modelos que permitan hacer el diagnóstico. Entre los trabajos más destacados se encuentran Khan et al. (2019); Solh et al. (1999); Wu et al. (2017); Evangelista y Guedes (2019); Ramana y Basha (2004); Postnikov et

al. (2020); Li et al. (2018); Temurtas y Tanrikulu (2010); Yang et al. (2019); Rajaraman et al. (2018); Fojnica et al. (2016); Shamshirband et al. (2014); Rusdah et al. (2015); Xing et al. (2020); Vijayaraj et al. (2020); Hwang et al. (2019); Souza et al. (2018).

Khan et al. (2019) proponen una red neuronal para la predicción de tuberculosis pulmonar. Cada paciente se representa en el conjunto de datos por medio de siete valores que son el género, historia clínica, tipo de enfermedad (pulmonar o extrapulmonar), VIH, edad, tipo de muestra y grupo étnico. Para este estudio se recolectaron un total de 12636 registros entre 2016 y 2017 en Pakistán, de los cuales se usaron el 70% para entrenar la red neuronal y el 30% para las pruebas. La exactitud que alcanza la red neuronal es del 94%.

Evangelista y Guedes (2019) plantean un modelo para el diagnóstico de tuberculosis pulmonar usando un sistema inmune artificial que combina las técnicas de lógica difusa con redes neuronales artificiales. Se emplea un total de 20 características para representar a cada paciente entre las que se incluyen dolor en el pecho, recuento de leucocitos, pérdida de peso, sudoración nocturna, fiebre, falta de aliento, VIH, concentración de hemoglobina, plaquetas, entre otras. Se utilizó un conjunto de datos de 175 pacientes del laboratorio Pasteur ubicado en el norte de Irán de los cuales el 70% hizo parte del entrenamiento del sistema inmune artificial y el 30% restante se reservó para las pruebas. Con esta estrategia se alcanza una sensibilidad del 87% y una especificidad del 86,12%.

Temurtas y Tanrikulu (2010) presentan modelos basados en redes neuronales multicapa y redes neuronales de regresión general para el problema del diagnóstico de tuberculosis pulmonar. En el caso de las redes neuronales multicapa se hicieron pruebas usando una y dos capas ocultas. Cada paciente se representa por medio de 38 características que incluye datos tales como temperatura, dolor en el pecho, debilidad, disnea por esfuerzo, presión sobre el pecho, esputo, sonido tracto respiratorio, hábito del cigarrillo, leucocitos, eritrocitos, trombosis, entre otros. Un total de 150 pacientes del hospital Diyarbakir al sureste de Turquía

fueron utilizados como conjunto de datos. La configuración que obtuvo la mayor exactitud es una red neuronal multicapa con dos capas ocultas alcanzando un 95,08%.

Shamshirband et al. (2014) proponen una red neuronal artificial utilizando información clínica y radiográfica para predecir la tuberculosis pulmonar activa. Se empleó un grupo de 563 episodios de aislamiento y un conjunto de validación de 119 episodios para un total de 682 pacientes del centro médico Erie County en Estados Unidos. La técnica empleada para obtener el modelo predictivo es una red neuronal de regresión general (GRNN). La topología de la red tiene tres capas, una capa de entrada, una sola capa oculta y una capa de salida. Cada paciente se representa como un conjunto de 21 variables dentro de las que se encuentran variables demográficas como la edad y si el paciente padece de VIH, variables relacionadas a síntomas como dolor en el pecho, pérdida de peso, tos, fiebre, y finalmente hallazgos radiográficos tales como cavidad en los lóbulos inferior y superior, adenopatía, entre otros. Se probaron diferentes cantidades de neuronas en la capa oculta intentando obtener los mejores resultados en el diagnóstico de tuberculosis pulmonar. Para las pruebas se usó un nuevo conjunto de 119 pacientes aislados por sospecha de tuberculosis pulmonar activa. El área bajo la curva ROC obtenida fue del 92,3%.

En este artículo se exploran otras técnicas de inteligencia artificial además de las redes neuronales que ya han sido usadas en otros trabajos para la predicción de tuberculosis pulmonar. Además, se usa un conjunto de datos de pacientes de Cali-Colombia que recoge características propias de los casos encontrados en esta ciudad, lo que hace necesario crear modelos propios que representen su comportamiento. Finalmente, también se presenta una aplicación móvil que permite usar los modelos de inteligencia artificial propuestos y que intenta acercar a los médicos al uso de los modelos para la toma de decisiones.

2. Materiales y métodos

En esta sección se presenta la metodología usada para obtener los modelos propuestos para la predicción de tuberculosis pulmonar. Los modelos se generaron utilizando cuatro algoritmos de clasificación supervisada (redes neuronales, árboles de decisión, y dos métodos de ensamble) y tienen como objetivo predecir si un paciente padece tuberculosis, o no.

2.1. *Diseño de estudio*

Estudio de prueba diagnóstica que utilizó la base de datos de los pacientes con tuberculosis pulmonar obtenida por la secretaría de salud del departamento del Valle del Cauca y aprobado por el comité de ética de la Universidad Libre. Esta investigación incluyó únicamente variables clínicas de los pacientes, sin especificación de nombres, número de identificación u otra información confidencial. Los autores manifiestan que esta investigación se acoge a los estándares internacionales de investigación biomédica según el acuerdo de Helsinki y definen el nivel de riesgo como: investigación sin riesgo, basados en el artículo 11 numeral a) de la Resolución 8430 de 1993 del Ministerio de la protección social de Colombia.

2.2. *Selección de datos*

Durante los años 2017 a 2019 la Secretaría de salud de la ciudad de Cali reportó aproximadamente 3152 pacientes con sospecha epidemiológica, clínica, microbiológica o radiológica a quienes se les realizó cultivo para MTB en cualquier muestra. Se seleccionó únicamente a los pacientes con cultivos de muestras en esputo; excluyendo a los casos cuyo cultivo se realizó después de 24 horas de iniciado el tratamiento antituberculoso, a pacientes con diagnóstico previo de tuberculosis pulmonar, y a pacientes con resultados de cultivos contaminados.

En la Tabla 1 se presentan los atributos o características usadas en esta investigación. Se tiene un total de 14 variables independientes sumado a la variable Tuberculosis que corresponde a la etiqueta de clase que se quiere predecir. Se utilizaron 1455 registros, de los cuales 1013 (69,6%) son diagnósticos positivos y 442 (30,4%) corresponden a diagnósticos negativos. Todos los registros pertenecen a pacientes de la ciudad de Cali, Colombia. Las muestras seriadas de baciloscopia fueron tomadas de esputo y evaluadas con tinción Ziehl-Neelsen o Auramina. Para pacientes menores de 5 años las muestras son tomadas de jugo gástrico. Los cultivos de las muestras fueron procesados en medio de Löwenstein-Jensen.

Tabla 1. Variables usadas para la predicción de tuberculosis

Atributo	Valores	Tipo de Dato
Sexo	Femenino=0, Masculino=1	Nominal
Edad	Edad en años del paciente	Numérico
Etnia	0=Indígena, 1=Afro descendiente, 2=Otro	Nominal
Grupo poblacional	0= Trabajador de la salud, 1 = Habitante de calle, 2 = Fármaco dependiente, 3 = Población carcelaria, 4 = Ningún grupo poblacional	Nominal
Comorbilidad	0 = Enfermedad renal, 1 = Otras Inmunosupresiones, 2 = Hipertensión, 3 = Diabetes, 4 = Otra enfermedad pulmonar, 5 = Ninguna	Nominal
Malnutrición	0 = No, 1 = Si	Nominal
Cantidad contactos	Cantidad de personas con las que tuvo contacto el paciente	Numérico
Cantidad contactos sintomáticos respiratorios	Cantidad de personas con síntomas respiratorios con las que tuvo contacto el paciente	Numérico

Atributo	Valores	Tipo de Dato
Cantidad contactos sintomáticos respiratorios con baciloscopia positiva	Cantidad de personas con síntomas respiratorios con baciloscopia positiva con las que tuvo contacto el paciente	Numérico
Cantidad contactos con menores de 5 años	Cantidad de personas menores de 5 años con las que tuvo contacto el paciente	Numérico
VIH	0 = No realizado, 1 = Positivo, 2 = Negativo	Nominal
Recibe terapia antirretroviral	0 = No, 1 = Si, 2 = No aplica	Nominal
Recibe Trimetoprim	0 = No, 1 = Si, 2 = No aplica	Nominal
Resultado de Baciloscopia	0 = No realizado, 1 = Positivo 1 a 9 BAAR, 2 = +++, 3 = ++, 4 = +, 5 = -	Nominal
Tuberculosis	0=Negativo, 1=Positivo	Nominal

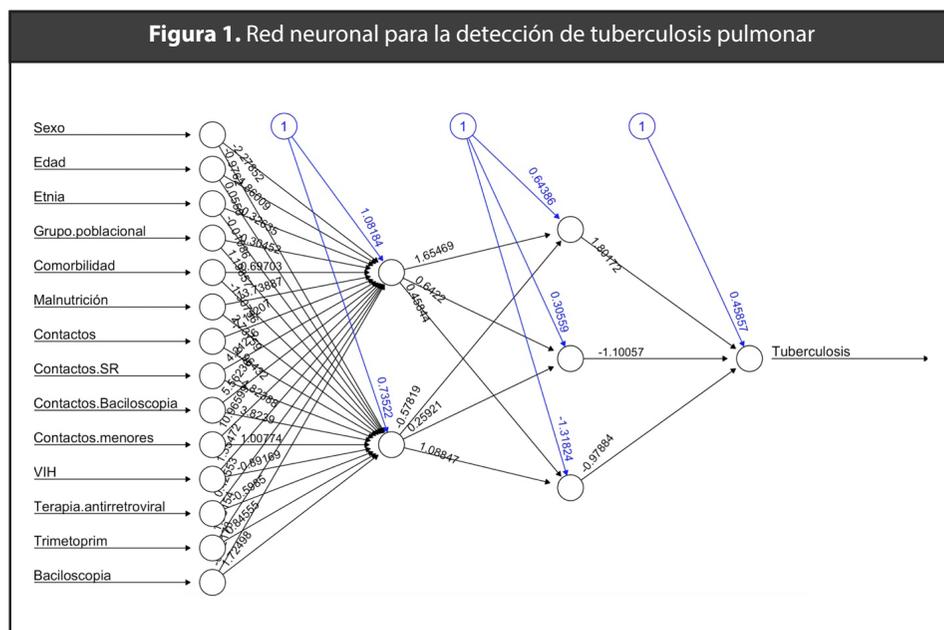
2.3. Construcción de los modelos de predicción

Para obtener los modelos se hizo uso de Python como lenguaje de programación y de scikit-learn como herramienta de machine learning, la cual permite aplicar diferentes algoritmos y utilizar diversas herramientas para el preprocesamiento de los datos.

2.3.1. Modelo propuesto usando redes neuronales

Para obtener el modelo predictivo usando redes neuronales se utilizó una red perceptrón multicapa (MLP) y se probaron diferentes topologías haciendo uso de reglas adhoc. En la capa de entrada se tienen 14 neuronas que corresponden a la cantidad de variables independientes que se presentaron en la Tabla 1. Además, se utilizaron

dos capas ocultas y se varió la cantidad de neuronas en dichas capas usando como heurística la regla de la potencia más cercana de 2, la cual indica que la cantidad de neuronas en la primera capa oculta debe ser la potencia más cercana de 2 al valor $2N$, donde N es la cantidad de neuronas en la capa de entrada. Por lo tanto, para la primera capa oculta se utilizaron hasta 32 neuronas y en la segunda capa oculta hasta 16. Además, se usaron 4 funciones de activación (identity, logistic, tan, relu) y 2 solvers (sgd y adam). En la capa de salida se tiene una sola neurona que indica si el paciente padece tuberculosis, o no. En total, se probaron $32 \times 16 \times 4 \times 2 = 4096$ configuraciones distintas. Este valor se obtiene al multiplicar la cantidad de neuronas en las capas ocultas (32 y 16) y la cantidad de funciones de activación y solvers que se probaron (4 y 2). En la Figura 1 se muestra la topología 14-3-2-1 que tiene 14 neuronas en la capa de entrada, 3 neuronas en la primera capa oculta, 2 neuronas en la segunda capa oculta, y 1 neurona de salida.



2.3.2. Modelo propuesto usando árboles de decisión

Para obtener el modelo predictivo con árboles de decisión se utilizó la implementación del algoritmo Decision Tree Classifier disponible

en scikit-learn. Este algoritmo permite obtener un árbol en cuyas hojas se encuentran las etiquetas de clase que en esta investigación corresponden al diagnóstico positivo o negativo de la tuberculosis pulmonar y en sus nodos internos se tienen condiciones sobre las 14 variables independientes. La clasificación de un nuevo paciente se realiza al someter sus 14 valores a los nodos internos del árbol hasta alcanzar una hoja con una etiqueta de clase que le será asignada. Durante las pruebas se modificaron tres parámetros, la profundidad máxima del árbol, la función usada para medir la calidad de un split, y la estrategia usada para elegir el split en cada nodo del árbol. La profundidad máxima se varió desde 50 a 2.000 con incrementos de 50. Se probaron las funciones Gini y Entropy para medir la calidad de un split. Finalmente, como estrategia para elegir el split en cada nodo se emplearon los métodos best y random disponibles en scikit-learn.

2.3.3. Modelo propuesto usando Random Forest

Adicional a las técnicas presentadas anteriormente, se emplearon dos métodos de ensamble para el problema del diagnóstico de tuberculosis pulmonar. En este tipo de técnicas se construyen estratégicamente varios modelos base y se utilizan sus salidas para obtener una clasificación más acertada. Los métodos de ensamble usan internamente algún tipo de clasificador. En el método Random Forest se utilizan árboles de decisión para construir los modelos base, es decir, se construyen N árboles que de forma independiente intentan hacer el diagnóstico de tuberculosis con base en las 14 variables independientes, o un subconjunto de ellas, tal como se explicó en la sección anterior. Al momento de construir cada árbol de decisión se seleccionan aleatoriamente cuáles atributos se utilizan en cada split de dicho árbol, por lo que es posible que no todos los atributos se usen en cada árbol que se emplea como modelo base. Para clasificar un nuevo paciente, se someten los 14 valores que lo representan a los N árboles para que cada uno asigne una etiqueta de clase indicando si la persona padece, o no, de tuberculosis. Finalmente, la decisión del método de ensamble basado en el algoritmo Random Forest será la mayoría entre las clasificaciones hechas por los N árboles. En este trabajo se varía la

cantidad N de árboles que se usan para hacer el ensamble desde 150 hasta 1.000 con incrementos de 50.

2.3.4. Modelo propuesto usando Extra Trees

Como segunda técnica de ensamble se utilizó el algoritmo Extra Trees. Al igual que el método Random Forest, el algoritmo Extra Trees crea un gran número de árboles de decisión a partir de un conjunto de entrenamiento. La predicción se realiza usando el criterio de mayoría para los problemas de clasificación y la media en el caso de regresión. Ambas técnicas coinciden en que seleccionan aleatoriamente cuáles atributos se utilizan en cada split de los árboles de decisión base. Esto permite que exista mayor variabilidad en los árboles resultantes y que la decisión de cada árbol pueda ser diferente, lo que favorece la clasificación final en el método de ensamble. Sin embargo, a diferencia de Random Forest donde cada árbol de decisión se construye usando un subconjunto del conjunto de entrenamiento, en el algoritmo Extra Trees se utiliza el conjunto de entrenamiento completo al momento de entrenar cada modelo base. Además, una vez seleccionados los atributos que se utilizan en cada split, el método Random Forest utiliza un algoritmo voraz para seleccionar el atributo óptimo que se debe usar en el split mientras que el algoritmo Extra Trees selecciona dicho atributo de forma aleatoria. Al igual que se hizo en el caso del algoritmo Random Forest, en este trabajo se varía la cantidad N de árboles que se usan para hacer el ensamble desde 150 hasta 1.000 con un paso de 50.

3. Resultados

En esta sección se muestran los valores obtenidos por los modelos propuestos para la detección de tuberculosis pulmonar. En las pruebas realizadas se utilizó validación cruzada con $k=5$ dividiendo el conjunto de datos en 20% para pruebas y 80% para entrenar los modelos. Esta división se repite cinco veces cambiando los datos que hacen parte del conjunto de pruebas de tal forma que cada instancia del conjunto de datos en algún momento haga parte de las pruebas. Además, en

esta sección se analizan y comparan los resultados obtenidos entre los modelos propuestos y algunos de los trabajos más relevantes enfocados al diagnóstico de tuberculosis pulmonar.

3.2. Métricas

Dado que en este artículo se aborda un problema de clasificación se hace uso de la matriz de confusión. Esta matriz permite conocer la capacidad de los modelos propuestos al momento de diagnosticar si un paciente tiene, o no, tuberculosis pulmonar. Además, se calculan diversas métricas que permiten comprender y comparar dicha capacidad predictiva.

Inicialmente, los parámetros calculados en cada prueba son los verdaderos positivos (VP) que corresponden a la cantidad de pacientes con tuberculosis que fueron clasificados por el modelo correctamente como positivos para dicha enfermedad, los verdaderos negativos (VN) que indica la cantidad de pacientes sanos que fueron clasificados por el modelo correctamente como negativos para tuberculosis pulmonar, los falsos positivos (FP) que en este trabajo indican la cantidad de pacientes sin tuberculosis que fueron clasificados por el modelo como positivos para dicha enfermedad, y finalmente los falsos negativos (FN) que corresponden a la cantidad de pacientes con tuberculosis que el modelo no logra detectar y los clasifica como negativos para tuberculosis.

Luego, a partir de los valores calculados en la matriz de confusión se obtiene la exactitud $((VP+VN)/(VP+VN+FN+FP))$ que corresponde a la cantidad de predicciones positivas que fueron correctas, la precisión $(VP/(VP+FP))$ que permite conocer la fracción de los verdaderos positivos entre los casos que se prevén positivos, el Recall o sensibilidad que se define con la fórmula $VP/(VP+FN)$ e indica la tasa de los verdaderos positivos, y el F1-score $(2*(precision*recall)/(precision + recall))$ que resume la precisión y sensibilidad de cada modelo. Además, se utiliza el área bajo la curva ROC (AUC) que es una medida de rendimiento para problemas de clasificación en varios

ajustes umbrales. El valor AUC indica qué tanto es capaz el modelo de distinguir entre las diferentes clases. Cuanto mayor sea el AUC, mejor será el modelo para predecir las clases correctas, bien sea el diagnóstico positivo o negativo de la enfermedad. Cuando el área bajo la curva ROC es 0.5, significa que el modelo no tiene capacidad de separación de clases. Por analogía, cuanto mayor es el AUC, mayor es la capacidad del modelo para distinguir entre pacientes con tuberculosis y aquellos sanos.

Tabla 2. Mejores resultados usando redes neuronales

Neuronas en la primera capa oculta	Neuronas en la segunda capa oculta	Sol-ver	Función de activación	Exactitud	Precisión	Recall	F1-score	AUC
8	9	adam	relu	0,7297	0,7416	0,9738	0,8413	0,630839
31	13	adam	relu	0,7150	0,7434	0,9363	0,8282	0,627402
19	5	adam	relu	0,7302	0,7419	0,9747	0,8417	0,626360
14	14	adam	relu	0,7302	0,7467	0,9587	0,8390	0,626170
28	14	adam	relu	0,7144	0,7524	0,9126	0,8244	0,624875
17	14	adam	relu	0,7270	0,7468	0,9564	0,8377	0,623549
25	15	adam	relu	0,7193	0,7500	0,9288	0,8293	0,622858
29	13	adam	relu	0,7193	0,7464	0,9408	0,8313	0,621704
8	11	adam	relu	0,7362	0,7437	0,9809	0,8455	0,621413
22	7	adam	relu	0,7308	0,7394	0,9809	0,8427	0,621380

3.3. Resultados usando redes neuronales

Para cada una de las 4096 configuraciones distintas de redes neuronales que se probaron usando esta técnica se calcularon las métricas explicadas en la sección anterior. La Tabla 2 muestra los 10 mejores resultados usando como criterio de selección el área bajo la curva ROC. El valor AUC más alto alcanzado con la técnica de redes neuronales para el problema del diagnóstico de pacientes con

tuberculosis pulmonar es de 63.08% y se logra con una topología 14-8-9-1, un solver adam y una función de activación relu. El recall del 97,38% indica que el modelo es capaz de predecir casi la totalidad de los casos positivos de tuberculosis. Sin embargo, la exactitud cae al 72,97%, lo cual refleja que el modelo no logra identificar correctamente los verdaderos negativos, es decir, los pacientes sanos, evidenciando así que el modelo tiende a clasificar como positivos a la mayoría de los casos de prueba. Esto se puede confirmar con el 74,16% de la precisión, la cual se ve afectada por la cantidad de falsos positivos, es decir, cuando el modelo clasifica a un paciente como positivo para tuberculosis a pesar de estar sano. Finalmente, el AUC tan solo alcanza un 63,08%, lo cual es consecuencia de la baja capacidad del modelo para clasificar los verdaderos negativos.

3.4. Resultados usando arboles de decisión

En la Tabla 3 se muestran los resultados obtenidos usando árboles de decisión y variando la profundidad máxima del árbol, la función empleada para medir la calidad de un split (Gini y Entropy), y la estrategia utilizada para elegir el split en cada nodo del árbol (best y random). En la tabla se presentan los 10 mejores resultados de acuerdo al área bajo la curva ROC. El valor AUC más alto usando la técnica de árboles de decisión para el diagnóstico de tuberculosis se alcanza con una profundidad de 800, la función gini y la estrategia best. Sin embargo, el área bajo la curva ROC más alto alcanzado por esta técnica (57,74%) es menor que el que se obtiene con el modelo de redes neuronales (63,08%). En la Figura 2 se muestra cómo varía el área bajo la curva ROC al utilizar las profundidades del árbol desde 50 hasta 2000 con incrementos de 50. Se puede observar que la profundidad del árbol no es proporcional con el AUC obtenido y que el pico más alto se alcanza en una profundidad de 800.

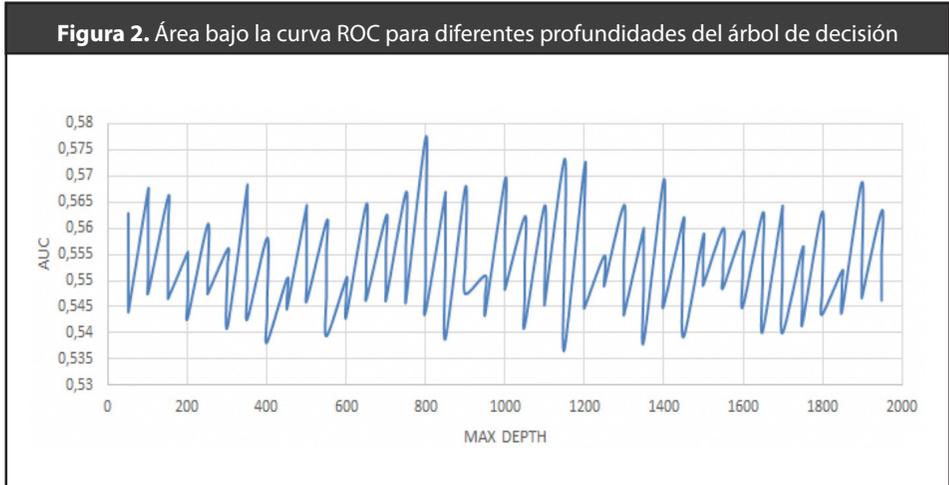


Tabla 3. Mejores resultados usando árboles de decisión

Función	Estrategia	Profundidad	Exactitud	Precisión	Recall	F1-score	AUC
gini	best	800	0,6396	0,7734	0,7137	0,7423	0,5774
gini	best	1.150	0,6380	0,7708	0,7161	0,7418	0,5732
gini	best	1.200	0,6422	0,7680	0,7289	0,7476	0,5716
gini	best	1.000	0,6343	0,7685	0,7150	0,7399	0,5693
gini	best	1.400	0,6396	0,7665	0,7270	0,7455	0,5692
gini	best	1.900	0,6375	0,7677	0,7201	0,7428	0,5688
gini	best	350	0,6364	0,7677	0,7214	0,7427	0,5682
gini	best	900	0,6412	0,7662	0,7282	0,7466	0,5679
gini	best	100	0,6401	0,7672	0,7270	0,7461	0,5675
gini	best	850	0,6375	0,7659	0,7229	0,7436	0,5669

3.5. Resultados usando Random Forest

En la Tabla 4 se muestran los resultados obtenidos usando la técnica Random Forest para diferentes cantidades de árboles al momento de hacer el ensamble. En la tabla se presentan los 10 mejores resultados de acuerdo al área bajo la curva ROC. El valor AUC más alto usando la técnica Random Forest es de 89,33% y se alcanza al emplear 450 árboles. Se puede observar que el recall sigue siendo alto (98,12%),

tal como ocurrió con la técnica de redes neuronales (97,38%), lo cual indica que el modelo es capaz de predecir casi la totalidad de los casos positivos de tuberculosis. Además, se observa una diferencia notable en la exactitud (92,78%) y la precisión (92,03%) en comparación con los resultados obtenidos con la técnica de redes neuronales. Con redes neuronales la exactitud y la precisión cayeron al 72,97% y 74,16%, respectivamente. Una exactitud del 92,78% como la obtenida por la técnica Random Forest indica que el modelo logra identificar correctamente gran parte de los verdaderos negativos, es decir, los pacientes que no tienen tuberculosis. Por esta misma razón, la precisión sube al 92,03%, ya que el modelo no presenta tantos falsos positivos como ocurrió con la técnica de redes neuronales. Sin embargo, para problemas de clasificación en el ámbito médico, un AUC del 89,33% puede ser considerado como bajo dadas las implicaciones que tiene hacer una clasificación incorrecta de los casos positivos y negativos.

Tabla 4. Mejores resultados usando Random Forest

Cantidad de árboles en el ensamble	Exactitud	Precisión	Recall	F1-score	AUC
450	0,927835052	0,920370370	0,981243830	0,949832776	0,893336847
300	0,927835052	0,915750916	0,987166831	0,950118765	0,889511017
900	0,927147766	0,915673694	0,986179664	0,949619772	0,889017434
150	0,927835052	0,914990859	0,988153998	0,950166113	0,888873379
800	0,927805012	0,914655218	0,988096443	0,950151373	0,888163960
1.000	0,927147766	0,913400182	0,989141165	0,949763033	0,887104519
350	0,923024055	0,912168344	0,984205331	0,946818613	0,883505380
700	0,923024055	0,912168344	0,984205331	0,946818613	0,883505380
600	0,924398625	0,910828025	0,988153998	0,947916667	0,883217271
750	0,922336770	0,911334552	0,984205331	0,946369245	0,882374159

3.6. Resultados usando Extra Trees

En la Tabla 5 se muestran los resultados obtenidos usando la técnica de ensamble Extra Trees para diferentes cantidades de árboles. En la tabla se presentan los 10 mejores resultados de acuerdo al área bajo la curva ROC. El valor AUC más alto usando la técnica Extra Trees es del 95,63% y se alcanza al utilizar 650 árboles en el ensamble. Este es el valor AUC más alto de las cuatro técnicas usadas en este trabajo. Además, la exactitud (96,76%) y la precisión (96,89%) también son las más altas alcanzadas durante la experimentación. Esto indica que el modelo logra una mejor capacidad para predecir los verdaderos negativos, es decir, las personas sanas, y que no tiende a clasificar como positivos para tuberculosis a estos pacientes sanos, disminuyendo en gran medida la tasa de falsos positivos. Además, el recall mantiene un valor alto (98,51%) reflejando así la capacidad del modelo para predecir los verdaderos positivos. En general, esta técnica de ensamble alcanza los mejores resultados y se perfila como la mejor alternativa para la detección de tuberculosis pulmonar entre las técnicas utilizadas en este trabajo.

Tabla 5. Mejores resultados usando Extra Trees

Cantidad de árboles en el ensamble	Exactitud	Precisión	Recall	F1-score	AUC
650	0,967697595	0,968932039	0,985192498	0,976994616	0,956397154
900	0,967697595	0,967117988	0,987166831	0,977039570	0,955121877
150	0,964261168	0,966958212	0,982230997	0,974534770	0,952653960
700	0,964948454	0,964285714	0,986179664	0,975109810	0,951234629
750	0,968456401	0,964572181	0,985041118	0,975063440	0,951204517
800	0,965744948	0,965717010	0,987179107	0,975039198	0,951185520
400	0,964948454	0,962500000	0,988153998	0,975158305	0,949959352
850	0,962199313	0,963249516	0,983218164	0,973131412	0,948622657
600	0,962886598	0,961501444	0,986179664	0,973684211	0,947840963
300	0,962886598	0,960614793	0,987166831	0,973709834	0,947203325

4. Discusión

El consolidado de las pruebas realizadas se muestra en la Tabla 6. El modelo que se obtiene con la técnica de ensamble Extra Trees resulta ser más acertado que los demás modelos propuestos para el diagnóstico de tuberculosis pulmonar. Los valores obtenidos en esta investigación con la técnica de ensamble Extra Trees son comparables y en algunos casos superan la exactitud reportada en trabajos similares del estado del arte. Por ejemplo, Khan et al. (2019) utilizan un total de siete atributos y la técnica de redes neuronales alcanzando una exactitud del 94%. Con el modelo de ensamble propuesto en este artículo se obtiene un 96,76% de exactitud. De igual forma, Evangelista y Guedes (2019) utilizan un sistema inmune artificial y se reporta un recall de 87%, muy por debajo de los valores obtenidos con las dos técnicas de ensamble e incluso con las redes neuronales. Por su parte, Temurtas y Tanrikulu (2010) emplean redes neuronales y un conjunto de datos con 150 pacientes logrando una exactitud del 95,08%. Es importante aclarar que los conjuntos de datos utilizados son diferentes en cada investigación, así como los atributos que representan a cada persona. De igual manera, el comportamiento de los pacientes puede variar de un país a otro así se esté analizando la misma enfermedad y eso también puede explicar el hecho de que las técnicas y los modelos obtenidos presenten exactitudes diferentes.

De nuestro conocimiento, este es el primer estudio en abordar el problema del diagnóstico de tuberculosis con técnicas de machine learning basado en árboles de decisión y estrategias de ensamble. La mayoría de estudios al respecto han utilizado redes neuronales empleando la exactitud como medida de generalización del modelo. En nuestro estudio, el desarrollo de modelos basados en redes neuronales artificiales no alcanzó una exactitud similar a los trabajos previos. Esta diferencia se debe en parte al objetivo que se consideró en la selección de los valores de entrada, pues fueron elegidos en función de que el modelo pudiese ser empleado de manera oportuna aún en los escenarios de práctica con menos recursos y con difícil acceso a tecnología. Estos escenarios suelen coincidir con lugares

donde la enfermedad es más prevalente. Por lo tanto, las variables debían ser fácilmente accesibles, reproducibles, precisas y de bajo costo. Los estudios hasta ahora desarrollados utilizan un amplio número de parámetros de entrada o incluyen parámetros que pueden ser de difícil acceso o con alto grado de intra-variabilidad. Por otro lado, las redes neuronales cuyas variables de entrada son pocas y fácilmente accesibles, reportan valores de rendimiento diagnóstico similares a los de nuestro estudio, principalmente en la sensibilidad. Sin embargo, debe tenerse en cuenta que la diferencia de resultados es debido principalmente a los criterios de selección de pacientes, al comparador de referencia, y al objetivo a evaluar, así como a las condiciones individuales e idiosincrásicas de la población de análisis, lo que hace que varíe el comportamiento de la misma enfermedad de un lugar a otro.

Tabla 6. Consolidado de los mejores resultados

Técnica	Configuración	Exactitud	Precisión	Recall	F1-score	AUC
Redes neuronales	Topología: 14-8-9-1 Solver: adam Función de activación: relu	0,7297	0,7416	0,9738	0,8413	0,6308
Árboles de decisión	Profundidad: 800 Función: gini Estrategia: best	0,6396	0,7734	0,7137	0,7423	0,5774
Método de ensamble Random Forest	Cantidad de árboles: 450	0,9278	0,9203	0,9812	0,9498	0,8933
Método de ensamble Extra Trees	Cantidad de árboles: 650	0,9676	0,9689	0,9851	0,9769	0,9563

Figura 3. Aplicación móvil para la predicción de tuberculosis pulmonar

Realizar Test

Fecha Nacimiento: 08/08/1994

Género: Masculino

Etnia: [dropdown]

Grupo poblacional: [dropdown]

Comorbilidad: [dropdown]

Malnutrición: No

VIH Confirmado: [dropdown]

Recibe Terapia Antiretroviral: [dropdown]

Recibe Trimetoprim: [dropdown]

Contactos: 0

Nro de Contactos sintomáticos respiratorios: 0

Nro Contactos sintomáticos respiratorios con baciloscopia positiva: 0

Nro de Contactos menores de 5 años: 0

Resultado de Baciloscopia: [dropdown]

Cerrar Obtener Predicción

Para permitir que el modelo basado en la técnica de ensamble Extra Trees esté disponible para el diagnóstico de tuberculosis pulmonar y puede servir de apoyo a la toma de decisión por parte del personal médico calificado, se desarrolló una aplicación móvil. La aplicación se muestra en la Figura 3 y permite, entre otras tareas, el registro de pacientes, el ingreso de las 14 variables de entrada al modelo, el diagnóstico anónimo para pacientes no registrados, la predicción usando la técnica de ensamble Extra Trees, el almacenamiento de los diagnósticos realizados y la aprobación del médico al diagnóstico realizado por la aplicación móvil. Esta última opción permite recolectar más datos que podrán ser usados para reentrenar el modelo.

5. Conclusiones

En este artículo se abordó el problema de la detección de tuberculosis pulmonar mediante el uso de varias técnicas de aprendizaje supervisado. Se propusieron cuatro modelos que permitan predecir si un paciente padece, o no, tuberculosis pulmonar usando 14 atributos

para representar a cada paciente. Las técnicas usadas fueron redes neuronales, árboles de decisión, y dos métodos de ensamble llamados Random Forest y Extra Trees.

Para evaluar el rendimiento de cada modelo se hizo uso de la matriz de confusión y se obtuvieron métricas tales como la exactitud, precisión, recall, f1-score y el área bajo la curva ROC. Los métodos de ensamble mostraron un área bajo de la curva mayor que las redes neuronales y los árboles de decisión. El método de ensamble Extra Trees obtuvo un área bajo la curva ROC de 95.63%, superior a los valores 63,08%, 57,64% y 89,33% de las redes neuronales, árboles de decisión y Random Forest, respectivamente. La exactitud y precisión de algunos modelos se vio afectada por su poca capacidad para predecir los verdaderos negativos. En general, los modelos propuestos con redes neuronales y árboles de decisión tienden a clasificar como positivos para tuberculosis a personas sanas, eso afectó significativamente su exactitud. Sin embargo, el modelo obtenido con la técnica Extra Trees mostró ser capaz de mantener un alto recall, clasificando correctamente los casos positivos, y al mismo tiempo identificar los verdades negativos, es decir, las personas sanas. Finalmente, se presentó una aplicación móvil que permite usar el modelo obtenido con la técnica de ensamble Extra Trees que intenta acercar a los médicos al uso de los modelos para la toma de decisiones, en este caso particular para la detección de tuberculosis pulmonar.

6. Referencias

- Evangelista, L.; Guedes, E. (2019). Ensembles of Convolutional Neural Networks on Computer-Aided Pulmonary Tuberculosis Detection. *IEEE Latin America Transactions*, 17(12), pp. 1954-1963. <https://doi.org/10.1109/TLA.2019.9011539>
- Fojnica, A.; Osmanović, A.; Badnjević, A. (2016). Dynamical model of Tuberculosis-Multiple Strain Prediction based on artificial neural network. *5th Mediterranean Conference on Embedded Computing (MECO)*, pp. 290-293. <https://doi.org/10.1109/MECO.2016.7525763>
- Hwang, E.J.; Park, S.; Jin, K-N.; Kim, J.I.; Choi, S.Y.; Lee, J.H.; Goo, J.M.; Aum, J.; Yim, J-J.; Park, C.M. (2019). Development and Validation of a Deep Learning-based Automatic Detection Algorithm for Active Pulmonary Tuberculosis on Chest Radiographs. *Clinical infectious diseases*, 69(5), pp. 739-747. <https://doi.org/10.1093/cid/ciy967>
- Joshi, B.; Lestari, T.; Graham, S.M.; Baral, S.C.; Verma, S.C.; Ghimire, G.; Bhatta, B.; Dumre, S.P.; Utarini, A. (2018). The implementation of Xpert MTB/RIF assay for diagnosis of tuberculosis in Nepal: A mixed-methods analysis. *PLoS One*, 13(8). <https://doi.org/10.1371/journal.pone.0201731>
- Khan, M.T.; Kaushik, A.C.; Ji, L.; Malik, S.I.; Ali, S.; Wei, D-Q. (2019). Artificial Neural Networks for Prediction of Tuberculosis Disease. *Frontiers in Microbiology*, 10(395). <https://doi.org/10.3389/fmicb.2019.00395>
- Kitonsa, P.J.; Nalutaaya, A.; Mukiibi, J.; Nakasolya, O.; Isooba, D.; Kamoga, C.; Baik, Y.; Robsky, K.; Dowdy, D.W.; Katamba, A.; Kendall, E.A. (2020). Evaluation of underweight status may improve identification of the highest-risk patients during outpatient evaluation for pulmonary tuberculosis. *PLoS One*, 15(12). <https://doi.org/10.1371/journal.pone.0243542>
- Li, L.; Huang, H.; Jin, X. (2018). AE-CNN Classification of Pulmonary Tuberculosis Based on CT Images. *9th International Conference on Information Technology in Medicine and Education (ITME)*, pp. 39-42. <https://doi.org/10.1109/ITME.2018.00020>
- Postnikov, E.; Esmedljaeva, D.; Lavrova, A. (2020). A CatBoost machine learning for prognosis of pathogen's drug resistance in pulmonary tuberculosis. *IEEE 2nd Global Conference on Life Sciences and Technologies (LifeTech)*, pp. 86-87. <https://doi.org/10.1109/LifeTech48969.2020.1570619054>
- Rajaraman, S.; Candemir, S.; Xue, Z.; Alderson, O.; Kohli, M.; Abuya, J.; Antani, S. (2018). A novel stacked generalization of models for improved TB detection in chest radiographs. *40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 718-721. <https://doi.org/10.1109/EMBC.2018.8512337>
- Ramana, K.; Basha, S. (2004). Neural image recognition system with application to tuberculosis detection, *International Conference on Information Technology: Coding and Computing*, 2, pp. 694-698. <https://doi.org/10.1109/ITCC.2004.1286735>
- Rusdah; Winarko, E.; Wardoyo, R. (2015). Preliminary diagnosis of pulmonary tuberculosis using ensemble method. *International Conference on Data and Software Engineering (ICoDSE)*, pp. 175-180. <https://doi.org/10.1109/ICoDSE.2015.7436993>
- Shamshirband, S.; Hessam, S.; Javidnia, H.; Amiribesheli, M.; Vahdat, S.; Petković, D.; Gani, A.; Kiah, MLM. (2014). Tuberculosis Disease Diagnosis Using Artificial Immune Recognition System. *International Journal of Medical Sciences*, 11(5), pp. 508-514. <https://doi.org/10.7150/ijms.8249>
- Solh, El.; Hsiao, A.; Goodnough, C.; Serghani, S.; Grant, B. (1999). Predicting Active Pulmonary Tuberculosis Using an Artificial Neural Network. *Chest Journal*, 10(4), pp. 968-973. <https://doi.org/10.1378/chest.116.4.968>
- Souza, F.; Sanchez, M.; Seixas, J.M.; Maidantchik, C.; Galliez, R.; Moreira, ADSR.; da Costa, P.A.; Oliveira, M.M.; Harries, A.D.; Kritski, A.L. (2018). Screening for active pulmonary tuberculosis: Development and applicability of artificial neural network models. *Tuberculosis*, 111, pp. 94-101. <https://doi.org/10.1016/j.tube.2018.05.012>
- Temurtas, F.; Tanrikulu, A. (2010). Tuberculosis disease diagnosis using artificial neural

- networks. *Journal of Medical Systems*, 34(3), pp. 299-302. <https://doi.org/10.1007/s10916-008-9241-x>
- Vijayaraj, M.; Abhinand, P.A.; Venkatesan, P.; Ragunath, P.K. (2020). An ANN model for the differential diagnosis of tuberculosis and sarcoidosis. *Bioinformation*, 16(7), pp. 539-546. <https://doi.org/10.6026/97320630016539>
- World Health Organization. (2021). Global tuberculosis report 2021, Global Tuberculosis Programme, 25 p. <https://www.who.int/publications/i/item/9789240037021>
- Wu, Y.; Wang, H.; Wu, F. (2017). Automatic classification of pulmonary tuberculosis and sarcoidosis based on random forest. *10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, pp. 1-5. <https://doi.org/10.1109/CISP-BMEI.2017.8302280>
- Xing, Z.; Ding, W.; Zhang, S.; Zhong, L.; Wang, L.; Wang, J.; Wang, K.; Xie, Y.; Zhao, X.; Li, N.; Ye, Z. (2020). Machine Learning-Based Differentiation of Nontuberculous Mycobacteria Lung Disease and Pulmonary Tuberculosis Using CT Images. *Biomed Research International*. <https://doi.org/10.1155/2020/6287545>
- Yang, A.; Jin, X.; Li, L. (2019). CT Images Recognition of Pulmonary Tuberculosis Based on Improved Faster RCNN and U-Net. *10th International Conference on Information Technology in Medicine and Education (ITME)*, pp. 93-97. <https://doi.org/10.1109/ITME.2019.00032>