UNIVERSIDAD POLITÉCNICA SALESIANA ECUADOR

POSGRADOS

Maestría en

TELEMÁTICA

RPC-SO-01-NO.025-2021

Opción de Titulación:

Proyecto de titulación con componentes de investigación aplicada y/o de desarrollo

Tema:

Evaluación del rendimiento de sistemas inteligentes basados en aprendizaje automático y aprendizaje profundo para la detección del asma en una señal audible de tos

Autor(es)

Silvia Lorena Chuquín Balseca

Director:

Christian Raúl Salamea Palacios

QUITO – Ecuador

2023

Autor(es):

Silvia Lorena Chuquín Balseca
Ingeniero Electrónico Mención Telecomunicaciones
Candidata a Magíster en Telemática por la Universidad Politécnica Salesiana – Sede Quito.
Silvia_chuquin@hotmail.com

Dirigido por:

Christian Raúl Salamea Palacios
Profesor Titular de la Universidad Politécnica Salesiana
Doctor en Ingeniería de Sistemas Electrónicos
Master en Diseño, Gestión y Dirección de Proyectos
Ingeniero Electrónico
csalamea@ups.edu.ec

Silvia Lorena Chuquín Balseca

EVALUACIÓN DEL RENDIMIENTO DE SISTEMAS INTELIGENTES BASADOS EN APRENDIZAJE AUTOMÁTICO Y APRENDIZAJE PROFUNDO PARA LA DETECCIÓN DEL ASMA EN UNA SEÑAL AUDIBLE DE TOS

DEDICATORIA

Dedico el actual trabajo a todas las personas que estuvieron a mi lado en todo momento y a mi familia. Gracias a todos.

AGRADECIMIENTO

Agradezco principalmente a Dios, a mi familia y a las personas que estuvieron apoyándome en esta parte de mi vida.

Tabla de Contenido

Índice de figuras

Índice de tablas

Evaluación del rendimiento de sistemas inteligentes basados en aprendizaje automático y aprendizaje profundo para la detección del asma en una señal audible de tos

Autor(es):

SILVIA LORENA CHUQUÍN BALSECA

Resumen

La inteligencia artificial en las últimas décadas ha tenido un gran avance, se utiliza en varios campos, uno de ellos es la medicina, la cual permite la detección temprana de enfermedades. La tos es un sonido sonoro que se origina por varios factores en los cuales se encuentra incluido el asma, hasta el momento el análisis de la tos como inicios de alguna enfermedad grave tiene una limitante, ya que se detecta con herramientas de medición antiguas e incómodas para los usuarios.

Con respecto a las neuronas artificiales son semejantes a las neuronas cerebrales de las personas, se interconectan entre ellas para realizar diferentes procesos, reciben desde el exterior información o a su vez de otras neuronas, tienen la capacidad de aprender y en base a este realizan variedades de aplicaciones de aprendizaje automático, como la clasificación de imágenes, el procesamiento del lenguaje natural y la predicción de series de tiempo.

En el siguiente trabajo de titulación se busca analizar las diferentes neuronas tales como la neurona base que es la que pertenece al aprendizaje automático y la neurona convolucional del aprendizaje profundo, se determinara cuál de las dos neuronas tiene mejor rendimiento para la detección temprana de tos con asma.

Para realizar este procedimiento se tiene los audios audibles de tos de la base de datos concedida por la Universidad de Cambridge, se realiza un acondicionamiento de la señal de tos y se aplica a las diferentes neuronas.

Como resultados, se evaluó el accuracy de las redes neuronales artificiales dónde, la MPL tiene un valor de 0,9132, la CNN de una capa es igual a 0,8264 y la CNN de dos capas posee un valor de 0,9421, dando como conclusión que la red neuronal con mejor resultado es la red neuronal convolucional de 2 capas.

Palabras clave:

Asma, Aprendizaje, Neurona, Convolución, Tos.

Abstract

Artificial intelligence in recent decades has made a breakthrough, it is used in various fields, one of them is medicine, which allows early detection of diseases. Cough is a sonorous sound that originates from several factors in which asthma is included. Until now, the analysis of cough as the beginning of a serious disease has a limitation, since it is detected with old and uncomfortable measurement tools. For the users.

With respect to artificial neurons, they are similar to the brain neurons of people, they interconnect with each other to carry out different processes, they receive information from the outside or in turn from other neurons, they have the ability to learn and, based on this, they carry out varieties of machine learning applications such as image classification, natural language processing, and time series prediction.

In the following degree work, we seek to analyze the different neurons such as the base neuron that belongs to automatic learning and the convolutional neuron of deep learning, it will be determined which of the two neurons has the best performance for the early detection of cough with asthma.

To carry out this procedure, we have the audible cough audios from the database provided by the University of Cambridge, a conditioning of the cough signal is carried out and it is applied to the different neurons.

As results, the accuracy of the artificial neural networks was evaluated where the MPL has a value of 0.9132, the CNN of one layer is equal to 0.8264 and the CNN of two layers has a value of 0.9421, giving As a conclusion, the neural network with the best result is the 2-layer convolutional neural network.

Keywords:

Asthma, Learning, Neuron, Convolution, Cough.

1. Introducción

Desde ya hace un tiempo considerable, la ciencia ha ido creando diferentes sistemas de reconocimiento de voz, ya que se busca analizar la voz humana y así poder tener patrones, los cuales serán utilizados posteriormente para su análisis, la iniciativa de realizar este procedimiento es el poder entrelazar tanto a las personas y las máquinas.

En un inicio se buscaba ayudar a las personas que sufrían de problemas auditivos. Con los años se obtuvieron sistemas automáticos de reconocimiento del habla, los que, hoy en día se utilizan en diferentes ramas , tanto para inteligencia artificial como en la salud (Rouhiainen, 2018).

El sistema de reconocimiento de voz es la capacidad que conserva un computador, al cambiar las palabras del ser humano en un código binario, lo cual se vuelve compresible para el mismo.

Concerniente al tema de salud y más específicamente en las enfermedades respiratorias, se busca alertar de alguna anomalía en la tos. El sistema nervioso es tan complejo y diestro, que advierte por medio de una alerta al cerebro de que algo anda mal con la persona (Asensi, 2015).

El cerebro indica a los músculos del pecho y abdomen que se contraigan y expulsen ráfagas de aire, produciendo secuencias de tos que pueden ser procesadas por sistemas de reconocimiento de voz y pudiendo detectar pequeñas anomalías en las diferentes fases del procesamiento de la señal auditiva, en este caso de la tos. Así, se obtienen los diferentes datos que posteriormente son analizados y procesados (Asensi, 2015).

En el presente trabajo se evalúan diferentes sistemas de reconocimiento de voz, asumiendo que los mismos se pueden aplicar a los eventos de tos (Romero et al., 2021). Estructurado en 4 capítulos que justifican los objetivos planteados.

El siguiente trabajo contiene el acondicionamiento de la base de datos de las señales de tos, siendo estos audibles, teniendo así los requerimientos necesarios para el entrenamiento de los sistemas automáticos propuestos.

Se busca encontrar los coeficientes cepstrales de Mel(MFCCs). Estos parámetros poseen características importantes de cada una de las señales audibles de tos, contenidas en la base de datos acondicionada, que servirá de ingreso para el entrenamiento de los sistemas automáticos.

Por consiguiente, se realiza el sistema de reconocimiento de tos de aprendizaje automático con redes neuronales utilizado la red Perceptrón Multicapa (MLP) el cual será considerado como línea base.

Se compara el rendimiento del sistema de línea base con el conseguido por el sistema de reconocimiento de tos, el cual es el de un aprendizaje profundo llamado Red Neuronal Convolucional (CNN).

Para las diferentes redes neuronales se consideran como entradas los espectrogramas que se obtienen a partir de los parámetros cepstrales de Mel.

2. Determinación del Problema

Hasta la actualidad, el análisis del asma como progreso de una padecimiento se restringe a materiales de medición subjetivas, donde el profesional de la salud remite un diagnóstico por el sonido de la tos, o si talvez presenta silbidos , también considera el hecho que le falta la respiración al paciente, pero la misma no siempre se da por el asma y puede también ser por infecciones crónicas del pulmón, enfermedades pulmonares restrictivas y el tromboembolismo pulmonar, siendo estos monitoreos incómodos (Asensi, 2015).

En octubre del 2019 apareció un virus llamado SARS-CoV-2 que provocaba el padecimiento denominada Covid-19, que derivó en una pandemia a nivel mundial, donde la crisis aumentó y las personas fueron obligadas al confinamiento para evitar más contagios y muerte en la población.

Como por ejemplo, debido a este hecho, el Gobierno de Buenos Aires incorporó una tecnología llamada IATos, esta aplicación conlleva inteligencia artificial permitiendo el testeo de coronavirus a los pacientes. Se obtuvieron muestras de 554 personas voluntarias con 2687 audios. Este proyecto ayudó a bastantes personas que tenían dudas si talvez se contagiaron, ellas podía realizar este test para después realizarle la prueba del Covid-19 con el hisopado (Coronavirus, 2022)

La población más afectada fue aquella que sufría de enfermedades respiratorias como el asma, muchas veces tales pacientes no sabían que tenían asma y comenzaban a presentar síntomas de tos.

Por esta razón y para evitar que estos pacientes deban asistir a los centros de salud por el miedo del contagio y las aglomeraciones, nace la idea de realizar una evaluación temprana y rápida, parecido a lo que realizó el Régimen de la ciudad de Buenos Aires pero con el asma, cabe recalcar que no es un diagnóstico médico, es importante siempre acudir al profesional de la salud, esta evaluación es un dato previo (Barrios et al., 2018).

3. Marco teórico referencial

El estudio acústico de la voz es un instrumento no invasivo de investigación vocal, que posibilita la obtención de registros de la señal audible de tos. Los diferentes datos que se alcanza de dicho análisis sirven para fines investigativos. Existen varios programas que utilizan diferentes algoritmos para conseguir mediciones de la voz que al no tener una estandarización obstaculiza este proceso ya que se tienen diferentes datos y va a depender del programa que se utilice.

En este trabajo se va a analizar la señal acústica de la tos dirigido a la identificación de la tos normal o con asma.

3.1 Estudio del asma

El asma es una de los padecimientos crónicos que existe en la actualidad, afectando a la sociedad y a los familiares de la persona que la padece (Ibrahim, Razak, Tamil, Idna, & Yusoff, 2008). Es causada por la disminución de tamaño de las vías por donde pasa el aire. Del 30-50% de las personas adultas presentan casos de asma que no es por alergia (Rouhiainen, 2018).

3.1.1 Naturaleza del asma

El asma es lo que se llama una enfermedad multifactorial, esto quiere decir que se desarrolla en una persona susceptible a enfermedades respiratorias. Para que se desarrolle esta enfermedad interactúan entre sí, una serie de factores que incluyen dos tipos, el primero es aquel en el que el huésped o paciente lo desarrolla propiamente por su organismo, y el segundo es el que se desencadena por factores externos, tales como alergias, tabaquismo, virus respiratorio, frio ambiental, o incluso los antinflamatorios no esteroides que también constituyen un factor que puede influir para desarrollar asma (M, D, J, & M, 2017).

El asma causa tos y silbidos recurrentes característicos de esta enfermedad, uno de los patógenos más frecuentes es la inflamación crónica de la vía aérea dificultando que el paciente pueda respirar.

3.1.2 Producción de la tos en el organismo

La tos es una forma en la cual el cuerpo humano limpia la garganta y las vías respiratorias, es una forma de curarse y estar protegido a pesar de que puede ser molesta y ruidosa.

El cerebro indica a los diferentes músculos que se tiene en el pecho y en el abdomen que debe librar aire de los pulmones y procesa a expulsar al agente irritante, así es como se origina la tos.

La tos preocupante es cuando persiste por mucho tiempo, esto quiere decir por varias semanas y es conducida de una mucosidad descolorida o en algunos casos con sangre, significa que la persona tiene una enfermedad que necesita atención médica (MedlinePlus, 2020).

La tos se puede presentar como aguda o crónica, se vuelve crónica cuando dura más de 3 semanas por lo cual se debe asistir al médico inmediatamente, en el caso que no pasa de las 2 o 3 semanas, se considera este tipo de tos como un resfriado, una gripe común o la bronquitis aguda (Medlineplus, 2021).

La tos crónica puede ser causada por las siguientes razones:

- Bronquitis Crónica
- Asma
- Alergias
- EPOC
- Fumar
- Reflujos gastroesofágico

- Enfermedades de la garganta

3.1.3 Características acústicas de la tos asmática

Existen diversos tipos de sonidos en los pulmones normales en función de la zona auscultada o del origen del sonido formado durante la respiración. Como efecto de ello, los parámetros como la frecuencia y la amplitud de la señal acústica difiere en función del tipo de sonido que se examine.

Entre todos ellos se pueden destacar los sonidos vesiculares, que se emiten en las frecuencias más bajas, los bronco-vesiculares de ellos se obtiene una frecuencia media intrínsecamente del rango establecido y por último los bronquiales, que presentan frecuencias más altas y suelen ser más disonantes (Medlineplus, 2021).

Preexiste un subgrupo de sonidos pulmonares nombrados ruidos adventicios, que acumulan las definiciones de hasta 162 términos relacionados con los sonidos respiratorios y su estudio se lleva a cabo mediante diferentes aplicaciones informáticas.

Estos diferentes sonidos están incorporados a los sonidos respiratorios normales mencionados anteriormente y forman un vislumbre de patología pulmonar, así se llega a detectar el asma, en función de la frecuencia fundamental, duración, intensidad sonora, etc.

Se puede clasificar en los siguientes grupos: estertores, son sonidos continuos que parecen silbidos, éstos muestran frecuencias bajas; crepitantes, estos sonidos son discontinuos de corta duración y tienen frecuencias altas; por último, el estridor que presenta una alta frecuencia conocido como un sonido chillón (MedlinePlus, 2020).

El asma se encuentra en la característica de alta frecuencia y sonido chillón.

3.2 Reconocimiento Automático del Habla

El sistema automático que se va a utilizar para reconocer los patrones característicos del asma forma parte de sistemas de reconocimiento de patrones que sirven para descubrir la naturaleza profunda de un fenómeno, lo cual permite describir y seleccionar características esenciales, las cuales son utilizadas para clasificar en categorías determinadas. Los procedimientos automáticos de reconocimiento de patrones abordan diferentes inconvenientes en informática, en la rama de la ingeniería y en otras disciplinas que tengan que ver con lo científico, por lo cual el diseño está sujeto a diferentes análisis que validarán los resultados. Una de las aplicaciones más utilizadas del reconocimiento de patrones es la que identifica patrones en imágenes médicas, en la bioinformática y biométrica.

Un sistema automático tiene tres etapas las cuales son: sensor, selector de las características y clasificador. En la fase del sensor se obtiene la personificación más devota del fenómeno aprendido, y un módulo que admite crear características de este.

3.2.1 Grabación de los Audios

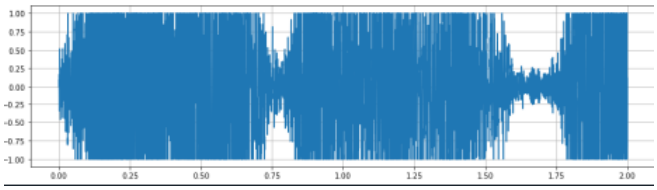
Con relación al proceso de la grabación, es transcendental el ambiente en que se efectúa este proceso. De preferencia que no tenga ruido ambiental, ni tampoco ruido electromagnético como el que se origina en los teléfonos celulares, también se debe considerar que los aparatos electrónicos como cafeteras, refrigeradoras, microondas, etc. Tienen ruidos propios de ellos, el cual es un ruido eléctrico, estos añaden o suman a la señal principal, de manera que se produce alteraciones en las señales originales, es importante evitar esos lugares para la realización de las grabaciones.

Como siguiente paso se debe considerar el hecho de que la máquina no va igualar al ser humano, por esa razón el uso del análisis acústico de una señal audible de tos va tener sus limitaciones, los diferentes datos conseguidos por este medio

deben ser interpretados y equiparados correctamente (Droguett & Droguett, 2017).

Una de las maneras más utilizadas para graficar un estudio acústico de la voz humana es el oscilograma, también llamado forma de onda (Figura 1), el cual representa la amplitud de la señal. Es la visualización directa que se tiene del sonido grabado, en este gráfico se puede observar y determinar las diferentes variaciones fonologías y los silencios que se tiene en el audio de tos (Droguett & Droguett, 2017).

Figura 1. Oscilograma de un audio de tos sin asma.



3.2.2 La representación de señales audibles por medio de MFCCs

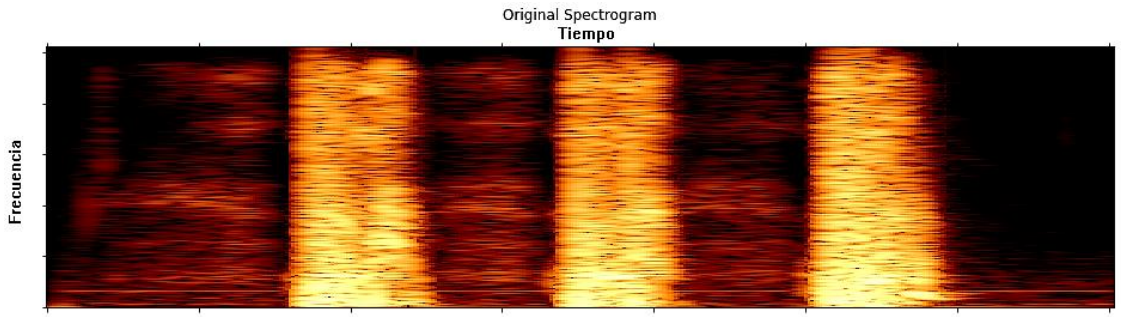
Los coeficientes cepstrales de frecuencia de Mel (MFCCs) se obtienen por medio de la aplicación de una técnica que extrae características spectrales de los audios dejando la señal del habla en la escala de frecuencia de Mel.

Esta escala dispone de un espaciado de frecuencia lineal con un valor por debajo de los 1000 Hz y con un espaciado logarítmico con un valor por encima de los 1000 Hz, la forma de onda del habla puede variar dependiendo de la condición física del ser humano. Los MFCCs son menos susceptibles a estas diferenciaciones por eso esta técnica se utiliza ampliamente en el reconocimiento de voz (Ibrahim, Razak, Tamil, Idna, & Yusoff, 2008).

El MFCC utiliza la escala de MEL para fraccionar la banda de frecuencia en sub bandas, para luego proceder a extraer los coeficientes cepstrales utilizando la transformada discreta del coseno (DCT).

Es importante saber cómo se comporta la voz humana, la frecuencia de la voz humana en los adultos es aproximadamente de 85 Hz a 255 Hz, para los hombres se tiene los rangos de 85 Hz a 180 Hz y para las mujeres se posee los rangos de 165 Hz a 255 Hz. Además de tener la frecuencia fundamental, hay armónicos de frecuencia fundamentales, estos armónicos son multiplicaciones enteras de la frecuencia fundamental. Para entender mejor lo de los armónicos, se tiene una frecuencia fundamental de 200 Hz, el segundo armónico tiene 300 Hz, el tercer armónico seria 400 Hz y así sucesivamente (kagglle, 2018).

Figura 2. Espectrograma de un audio de tos



Como se observa en figura 2 se tiene un espectrograma de un audio de tos, se muestra la frecuencia frente al tiempo, el color amarillo representa el nivel de energía de la señal de voz.

Las personas pueden escuchar sonidos aproximadamente entre 20 Hz Y 20 KHz, la percepción del sonido no es lineal, por esa razón los seres humanos pueden distinguir mejor entre sonidos de baja frecuencia que los sonidos de alta frecuencia, un ejemplo es que los seres humanos pueden escuchar claramente la diferencia entre 200 Hz Y 400 Hz, pero entre 14 y 15 KHz el oído humano no percibe ninguna diferencia al ser frecuencias altas.

3.3 Inteligencia artificial

La inteligencia artificial (IA) es la destreza que tienen los ordenadores para realizar acciones que regularmente demanda de la inteligencia humana, se puede decir que es la cabida de los aparatos para usar varios algoritmos, educarse de los datos

que se les otorga y manejar lo asimilado en la adquisición de decisiones a modo que lo forjaría un ser humano. Una de las ventajas de utilizar la inteligencia artificial (IA) es que los dispositivos no requieren descansar y en la actualidad ya pueden examinar volúmenes grandes de información a la vez.

Debido al avance de la tecnología actual las técnicas de inteligencia artificial logran muchas de las tareas que antiguamente solo conseguía efectuar los seres humanos. Las diferentes tecnologías de la IA son manejadas para ayudar a las personas en la vida diaria, en las tareas de oficina, en la salud, en los diagnósticos, etc.

La IA se puede utilizar en casi todas las aplicaciones técnicas también como, por ejemplo: reconocimiento de imágenes, mejorar la estrategia comercial, detección y clasificación de cosas, colocación de contenido en las redes sociales, amparo frente a amenazas de seguridad cibernética (Rouhiainen, 2018).

En la actualidad se instaura otras categorías de IA según Stuart J.Russell y Peter Norving entre ellos: Sistemas que llegan a tener razonamiento como los seres humanos, sistemas que operan como los seres humanos, sistemas que piensan lógicamente y sistemas que operan razonablemente.

En la inteligencia artificial abarca 3 etapas:

- **Etapa inicial (1956-1970).** - En este período se realiza las técnicas básicas.
- **Etapa de prototipo (1970-1981).** – En este período se estudia la posibilidad de las aplicaciones en la industria de la IA.
- **Etapa de difusión (1981 - ??).** – En esta etapa la inteligencia artificial ya no es solo para desarrollo, sino que se utiliza para la manufactura y el mundo corporativo (Ana, 2018).

3.3.1 Aprendizaje automático o machine learning

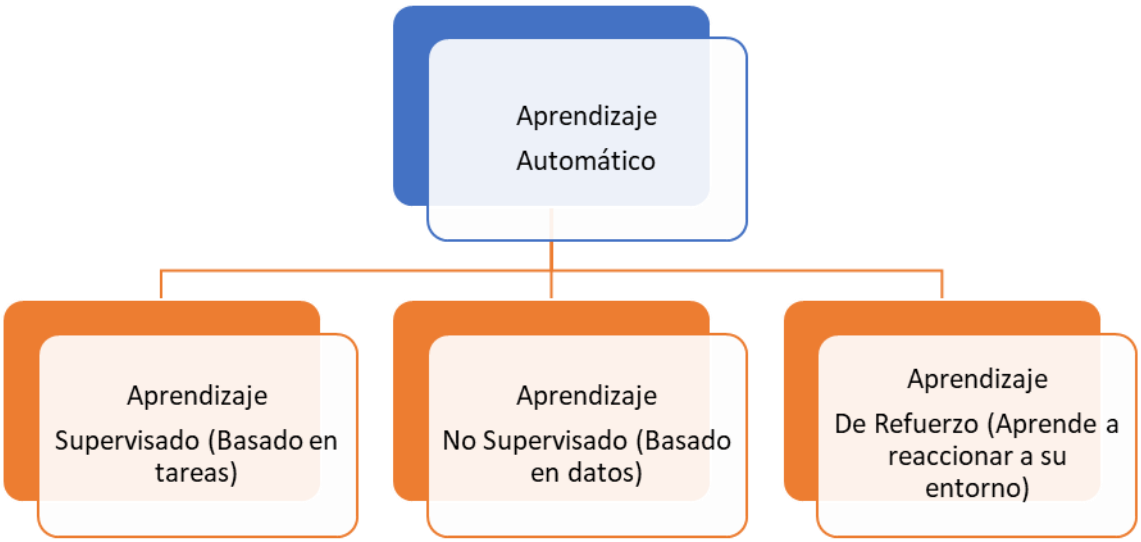
El aprendizaje automático consiste en que la maquina obtendrá experiencia y conocimiento a partir de las diferentes funciones que se le otorgue, ahí viene la diferencia comparada a las décadas anteriores, donde se programaba explícitamente para que realice esa función (Rouhiainen, 2018).

Uno de los objetivos del aprendizaje automático es que pueda predecir ejemplos futuros, sin que el ser humano intervenga en este proceso. El motor de Google que hoy en día es muy utilizado tiene aprendizaje profundo ya que usa algoritmos para poder aprender patrones de los diferentes datos que se le proporciona, este conocimiento adquirido por la maquina sirve para que pueda tomar decisiones.

Además de la capacidad de aprendizaje también anticipa comportamientos, esto ayuda en lo que es el reconocimiento facial, muy utilizado para la seguridad en las diferentes empresas y así no haya usurpadores, el aprendizaje automático en la actualidad está en todos lados. Muchas veces la población no sabe que atrás de alguna actividad cotidiana como la búsqueda de información en internet también tiene en el trasfondo la inteligencia artificial, ya que Google al saber los gustos de las personas le ofrece preferencias, también se utiliza en el reconocimiento de voz para las personas que no les gusta escribir lo pueden hacer hablando, lo cual también tiene inteligencia artificial (Ana, 2018).

En la figura 3 se puede observar los tres subconjuntos del aprendizaje profundo que se pueden utilizar.

Figura 3. Características de aprendizaje automatico



Con respecto al aprendizaje supervisado, los diferentes algoritmos utilizan los datos que han sido establecidos anticipadamente, así se muestra cómo debería ser categorizada la información reciente que proporciona al sistema, en este método si interviene el ser humano para que suministre la respectiva retroalimentación.

En el aprendizaje no controlado el algoritmo ya no usa datos constituidos previamente, al contrario, el algoritmo busca la manera en clasificar estos datos. Por lo tanto, se puede decir que este procedimiento no utiliza la mediación humana ni tampoco su retroalimentación.

Como último se tiene el aprendizaje por refuerzo, los algoritmos van a aprender de la práctica, se puede decir que se tiene un refuerzo positivo cada vez que este acierte. Uno de los ejemplos más claros para entender este tipo de aprendizaje, es cuando le damos una recompensa a nuestra mascota, si hace algo bien se le da el premio (Rouhiainen, 2018).

3.3.2 Aprendizaje profundo

Este tipo de aprendizaje propone modelar contemplaciones de alto nivel de los diferentes datos empleando arquitecturas formadas por un elevado número de las capas de transformaciones que consiguen ser tanto lineales o no lineales, al basarse en la arquitectura de las neuronas del cerebro humano estas técnicas se denominan redes neuronales artificiales (RNAs) (González Muñiz, 2018).

Se utiliza para solucionar problemas que tengan un grado de complejidad alto y que normalmente conllevan cantidades de datos grandes, su implementación requiere de enormes conjuntos de información y los equipos que tengan estos sistemas deben tener un procesamiento y capacidad amplia. En la actualidad el Deep learning es utilizado en el reconocimiento de voz, visión artificial y en la asistencia al conductor, esto quiere decir en los vehículos inteligentes que se tiene en la actualidad y ayuda a evitar accidentes de tránsito.

Otro ejemplo son las traducciones idioma como las generadas por Facebook quien reveló que se efectuaron hasta 4 500 millones de traducciones a diario. Si no existiera el aprendizaje profundo el costo de realizar esta acción de traducción sería bastante grande, además que se necesitaría más personal (Rouhiainen, 2018).

Estas técnicas de aprendizaje profundo son adelantadas en su escalabilidad, esto quiere decir que la cantidad de datos de entrada cuantos más sean es mejor el comportamiento, lo cual permite en la actualidad más beneficios y que siga avanzando la tecnología del Big Data o la Industria 4.0.

Con respecto de la industria el Deep learning aún no es muy avanzado debido a que es un campo por el cual se está explorado, pero se han tenido avances como por ejemplo el análisis de vibraciones en los motores que se utiliza en la industria (González Muñiz, 2018).

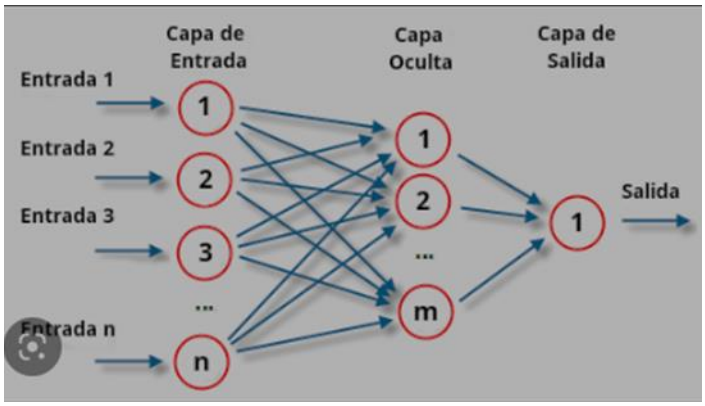
3.3.3 Neuronas Artificiales

En la actualidad existen varios estudios referentes a la detección de enfermedades respiratorias, para ellos se utilizan diferentes softwares , uno de ellos es MATLAB y se está efectuado nuevas tecnologías como las redes neuronales artificiales que se denominan ANN (Atificial Neuronal Networks). Estas neuronas se crearon por el hecho de simular el sistema nervioso biológico de los seres humanos y están constituidas por un contiguo de unidades llamadas neuronas que se localizan acopladas entre sí (Diego, 2019).

Al igual que las neuronas naturales de los seres humanos su primordial objetivo es desplegar diferentes operaciones de síntesis y procesamiento de la información, en la mayoría de las neuronas poseen estructuras de árbol, esto permite recoger caracteres de ingreso procedente de demás neuronas a través de las conexiones. El cerebro humano sujeta más de cien mil millones de neuronas y uniones en el sistema nervioso. A pesar del avance de la tecnología y de los buenos resultados con la tecnología computacional, no se logran tiempos de conmutación iguales a las neuronas del ser humano, que aún sigue siendo superior (Diego, 2019).

Una sola neurona en si no es de mucha importancia, pero cuando conmuta con las demás neuronas se tiene una red y puede resolver problemas muy complejos. Se puede indicar que una red neuronal está organizada e interconectada y así se organiza en tres capas, las cuales son la capa de ingreso (input), esta pasa por la capa oculta (layer1, layer2) y por último sale por la capa de salida (output) (Figura 4).

Figura 4. Red Neuronal Artificial



La neuronas, como se describió anteriormente, logran solucionar dificultades presentadas posteriormente de ser entrenadas. La resolución del problema propuesto se fundamentó en 5 principios los cuales son: Aprendizaje Adaptivo, Autoorganización, Tolerancia a fallos, Manipulación en tiempo real y Cómoda introducción en la tecnología existente hasta la fecha (Droguett, 2017).

La historia de las redes neuronales se suelen dividir en tres etapas, la primera etapa está comprendida en las décadas de los 40 hasta los 70, en donde saltó de la sorpresa de estos nuevos modelos hasta el escepticismo, para la segunda etapa se da 10 años después cuando ya es los 80, se torna los temas referentes a las redes neuronales y surgen mejores mecanismos, maneras, etc. Además, se consigue una planicie amplia en la que no se puede revelar la profundidad de aprendizaje que conservan las redes neuronales artificiales, esto se debe también porque en esa época no se tenía el procesamiento adecuado en las computadoras (Carlos & Marta, 2001).

En la tercera etapa es a partir del 2006 en la que se consigue superar la barrera y se aprovecha el poder de los equipos computacionales, se tienen mejores procesamientos, se consigue preparar cientos de capas jerárquicas que acceden y fomentan el aprendizaje profundo, dando una excelente capacidad de procesamiento de las redes neuronales y resolviendo problemas más complicados. En el trabajo de titulación se utiliza la red neuronal Perceptrón Multicapa (MLP) y la red neuronal Convolucional (CNN) (Carlos & Marta, 2001).

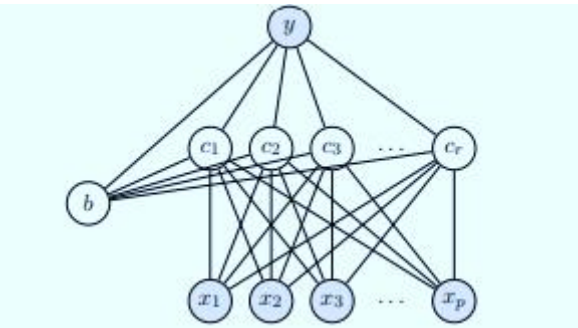
3.3.4 Perceptrón Multicapa (mlp)

El perceptrón simple fue creado en 1957 por Frank Rosenblatt, este tipo de red neuronal tiene limitaciones las cuales fueron publicadas en el libro de Marvin Minsky y Semour Papert (Minsky, 2017). Debido a la publicación de este libro en ese tiempo se pensó que era el fin de las redes neuronales. Uno de los inconvenientes que se refería en este libro es la abolida capacidad para corregir dificultades que no sean linealmente divisibles, se llegó incluso a demostrar en el libro que la red neuronal no era idónea de formar la función XOR (Julio, R, Diez, & Carlos, 2020).

La idea del perceptrón multicapa nace de la idea de los autores del libro que se mencionó anteriormente de Minsky y Papert, ellos indicaron que la mezcla de diferentes perceptrones simple podría ser un medio para indiscutibles problemas no lineales. Cabe recalcar que los autores no arreglaron el problema, ni tampoco dieron una solución de como adecuar los pesos de la capa de ingreso a la capa oculta.

El perceptrón multicapa fue afiliado como un piloto matemático ventajoso aproximado o interpolando relaciones no lineales entres los datos de entradas y salidas (Julio, R, Diez, & Carlos, 2020). Debido a su cómodo uso y a que las aplicaciones del perceptrón multicapa forman la organización característica de los modelos de aprendizaje, se ha constituido como base para el progreso de las demás redes neuronales. Se determina porque agrupa las neuronas en capas de desiguales paralelismos, cada capa tiene un conjunto de neuronas y estas se clasifican en tres tipos: la capa de entrada, las capas ocultas y la capa de salida, como se muestra en la figura 5 (Carlos & Marta, 2001).

Figura 5. Bosquejo de una red neuronal MLP con una sola capa de neuronas ocultas



En la capa de entrada, las neuronas son comisionadas uno por uno de recoger las señales o los diferentes esquemas del exterior y así poder trascender a todas las neuronas que siguen de la siguiente capa. Con respecto a las capas ocultas, estas tienen las funciones de intermedias, en esta capa se ejecuta un proceso no lineal de los diferentes esquemas recogidos.

Por último, la capa de salida es la que suministra la respuesta de la red al exterior para cada uno de los esquemas de la entrada (Julio, R, Diez, & Carlos, 2020).

Las conexiones del perceptrón multicapa se puede decir que perpetuamente están dirigidas hacia adelante, esto quiere indicar que las neuronas de una capa se conectan con las neuronas de la subsiguiente capa, por lo que se les llama redes neuronales prealimentadas que en ingles seria Feedforward Neural Networks (FNNs). Los enlaces entre las neuronas de la red transportan también asociado un umbral, en el caso de la red neuronal MLP es como un enlace más a la neurona, cuyo ingreso está dada o es igual al valor de 1.

Se dice que 28ienee conectividad total o que la red está completamente acoplada, cuando todas las neuronas de una capa están juntas a todas las neuronas de la siguiente capa (Carlos & Marta, 2001) (Julio, R, Diez, & Carlos, 2020).

3.3.5 Red neuronal convolucionales (CNN)

Una red neuronal convolucional conserva el conocimiento de capas, es un espécimen de red multicapa, esto quiere decir que tiene otras capas convolucionales y de submuestreo (pooling) variadas, en la parte final posee una serie de capas full-conected, esto quiere decir que es como una red perceptrón multicapa. Cada neurona de una capa no recoge enlaces entrantes de todas las neuronas de la capa anterior, sino de ciertas neuronas. Una de las ventajas es que una neurona se va a especializar en una sola zona de la enumeración que recibió de la capa anterior y así se consigue reducir drásticamente la cifra de los pesos, evitando tener duplicaciones innecesarias.

Estas redes neuronales se manipulan para el procesamiento de retratos, así se logra aparentar el comportamiento de las neuronas del cerebro humano que captan imágenes, poseen la cabida de instruirse con relación entrada-salida, en el cual se tiene como entrada una imagen y la misma está basada en operaciones de convolución de ahí su nombre de red neuronal convolucional (Jaime, 2017).

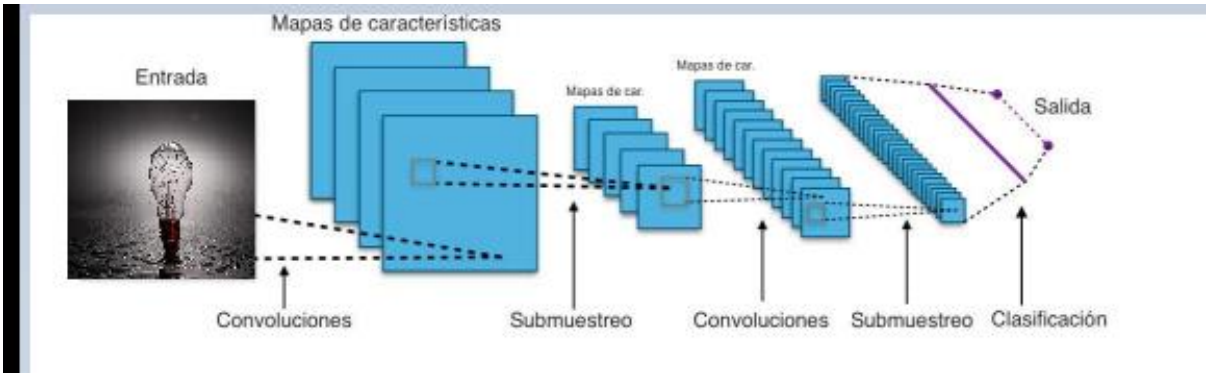
La entrada de una red convolucional ordinariamente es una imagen mxmxr, donde m es la altura, m el vasto de la imagen y r es la cifra de canales. Las capas convolucionales poseen k filtros (o kernels) cuyas extensiones son n x n x q, donde n y q son designadas por el diseñador de la red neuronal convolucional. Cada filtro forma mediante la convolución un mapa de rasgos con el tamaño (m – n + 1) x (m – n +1) x p, p es la cifra de filtros que el diseñador desea utilizar.

Cada mapa es sub-muestrado en la capa de pooling con la acción “mean pooling” o “max pooling” sobre zonas adyacentes de tamaño p x p donde p toma valores desde 2 hasta 5, 2 es para imágenes pequeñas y el 5 para las imágenes grandes, no se recomienda poner un valor más de 5.

Antes o después del submuestreo, se destina una función de activación sigmoidal más una ponderación para cada mapa de rasgos (Jaime, 2017) (Edgar, 2017).

En la figura 6 se observa un bosquejo elemental de una red neuronal convolucional.

Figura 6. Bosquejo elemental de una red neuronal convolucional



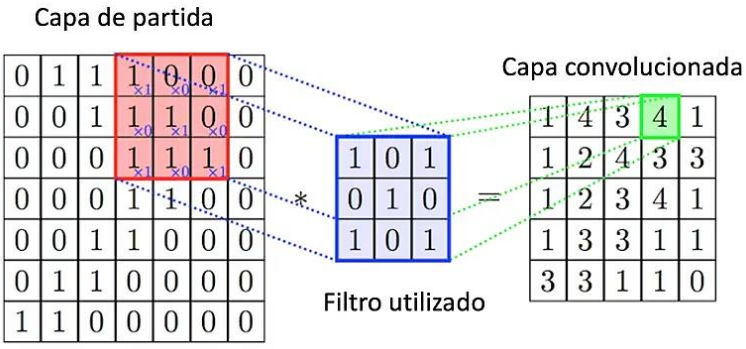
Capa convolucional: La capa convolucional busca sujetar la carga computacional del sistema, en esta etapa de la red, se limita el número de enlaces posibles entre las neuronas de la capa oculta y los compendios de la imagen de entrada.

Las imágenes conservan la propiedad de ser “estacionarias”, esto quiere decir que los rasgos de estas en algunas partes determinadas de la imagen pueden ser los mismos, a pesar de encontrase en otra zona distinta.

La convolución es una acción de efectos y sumas entre la imagen de entrada y el filtro para poder crear un mapa de características. Así se obtienen rasgos útiles que después serán utilizados para su posterior análisis.

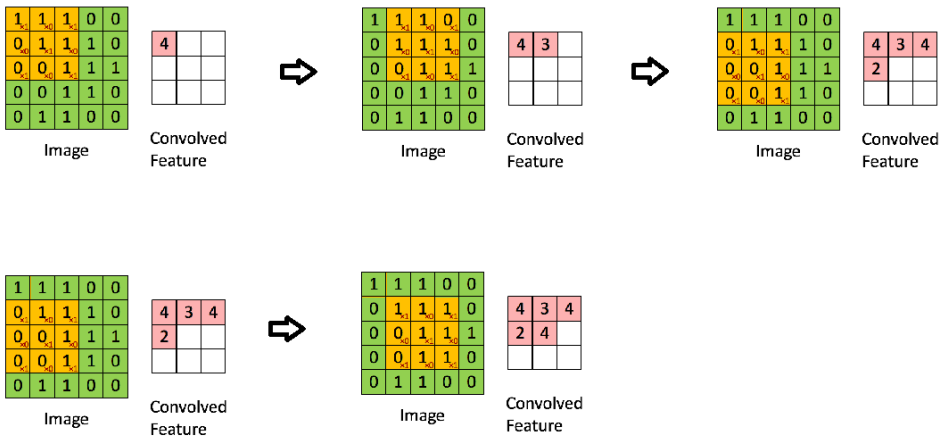
A la imagen de entrada a la red, se le sobrepone el filtro y se calcula la convolución de 2 dimensiones entre los pertinentes compendios de la imagen y el kernel, la consecuencia de esta maniobra se acumula en una posición en una matriz, esta se llama matriz de activación, como se observa en la figura 7.

Figura 7. Acción básica de convolución



El siguiente paso, se traslada el filtro a un enfoque a la derecha sobre la imagen de entrada y se retorna a computar la convolución, este efecto es almacenado en la subsiguiente posición de la matriz de activación, este paso se vuele a repetir en toda la imagen, trasladándose de izquierda a derecha, y descendiendo hasta conseguir a una unidad de borde. La matriz de activación se obtiene una vez se completa el camino en toda la imagen, y se obtiene las particularidades que se escudriñan en la imagen para cada filtro, este proceso se obtiene visualizar en la figura 8 (Jaime, 2017).

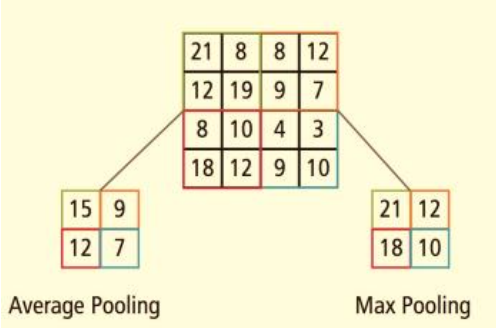
Figura 8. Matriz de activación



Capa de pooling o reducción. – Posteriormente de la capa de convolución se halla la capa de pooling, la salida es el máximo de la entrada en una ventana en el caso de Max Pooling y la media se aplica en el caso de Average Pooling, se traslada la ventana por la matriz de datos según una disposición que se le llama paso o stride. Esto sirve para sujetar el tamaño espacial de los datos obteniendo las características más comunes.

Al realizar este procedimiento se tiene un sistema menos preciso, pero eso conlleva a tener muchos beneficios como el sobreajuste y mejorar la compatibilidad, debido a que se logra reducir las características, y se obtiene un carácter invariable a pequeñas trasladaciones en la entrada, en la figura 9 se observa em Max pooling y Average Pooling con paso dos y la ventana de dos x dos (Jaime, 2017).

Figura 9. Max pooling y Average Pooling



Capa fully- connected. – La presente capa es la última del bosquejo de la red neuronal convolucional, es un clasificador que determinará la clase concierne a la imagen de entrada, su encargo es el de mostrar que numero 'cree' la red que se le entrego a la entrada.

Esta capa también tiene el cargo de activación, regularmente la función de activación denominado ReLu, se obtiene las probabilidades de cada clase posible.

Cada una de las componentes de salida simboliza la contingencia que tiene la imagen de entrada de corresponder a una explícita clase, en otras palabras, la probabilidad que posee de ser un número u otro (Jaime, 2017).

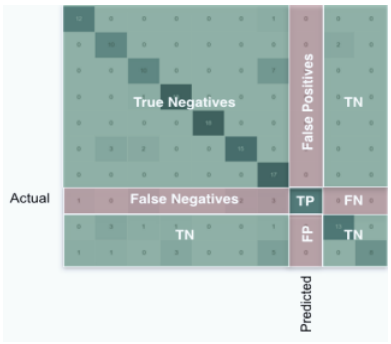
3.4 MATRIZ DE CONFUSIÓN

Llamada además como matriz de error o la tabla de contingencia, esta herramienta admite concebir el desempeño de un algoritmo de la Inteligencia artificial, en ese trabajo se verifica la matriz de confusión tanto de la red neuronal MLP y CNN, posterior a esto se realiza el análisis pertinente.

La cifra de pronósticos de cada clase es personificada en las columnas y las reclamaciones de la clase real está en las filas, el bien de las matrices de confusión es el poder ve si el sistema está enredando dos clases.

Para cuantificar el comportamiento de la matriz de confusión, se divide la matriz en regiones, para entender dichas regiones en la figura 10 se observa cómo está dividida la matriz de confusión en cuatro regiones (Ana, 2018).

Figura 10. Regiones de la matriz de confusión (Rouhiainen, 2018)



- Verdaderos positivos (TP), en esta parte de la región se encuentra en cuanto a los ciclos un clasificador ha anunciado que era de una concluyente clase y posterior a esto ha poseído triunfo en su predicción.
- Falso negativos (FN), en esta parte de la zona de la matriz de confusión se encuentra cuantas veces un clasificador ha adivinado que era una clase concluyente, pero en este caso que no habido acierto en el pronóstico.
- Falsos positivos (FP), para esta parte de la zona de la matriz de confusión se halla en cuanto a las veces un clasificador ha anunciado que no era de una concluyente clase, pero no tuvo acierto en la misma y por consiguiente si era de esa clase. Se dice que los falsos positivos son inexistentes alarmas de un clasificador.
- Verdaderos negativos (TN), para esta zona de la matriz de confusión se tiene las veces que un clasificador ha señalado que no era de una concluyente clase y si obtuvo acierto en esta predicción, son repercusiones correctas de un clasificador.

4. Materiales y metodología

La base de datos inicial de tos normal y asma utilizada en este trabajo contiene 262 grabaciones de tos con asma y 131 grabaciones de tos normal. Las grabaciones tienen una duración desde 0,3 segundos hasta 9 segundos, los colaboradores realizaron estas grabaciones de forma telemática y anónima, cada paciente era consciente que el objetivo de la grabaciones sería utilizarlas para realizar el procesamiento digital de las señales audibles.

En la ejecución de este proyecto, con respecto al hardware, se utiliza un ordenador laptop con las siguientes características (Tabla 1):

Tabla 1. Características del computador

Procesador:	Intel Core (TM9 i7-3610QM CPU @ 2.30GHz
Memoria instalada (RAM):	8,00 GB
Tipo de Sistema:	Sistema operative de 64 bits, procesador x64

Para la programación de las neuronas y el preprocesamiento de la base de datos, se maneja el software de Anaconda Navigator, dentro del cual se utiliza la aplicación de Spyder. Esta plataforma permite trabajar con el lenguaje de Python con una representación más amigable.

4.1 Anaconda Navigator

Anaconda Navigator es un ambiente de trabajo que se utiliza para la ciencia de datos, permite realizar varias aplicaciones, así como administrar de una manera fácil distintos paquetes que otorga esta plataforma. Así, Anaconda Navigator tiene la capacidad de indagar paquetes en Anaconda Cloud o en otros repositorios, la plataforma se encuentra disponible para sistemas de Window, macOS y Linux (Juan, 2020).

4.2 Data Augmentation

El data augmentation es una habilidad que se utiliza en las redes neuronales para agrandar la cifra de ejemplos con los que se va a entrenar la red neuronal.

Para entrenar un piloto de aprendizaje automático, realmente lo que se hace es concertar sus parámetros para fijar una entrada específica, en este caso, una imagen del espectrograma de la señal a alguna salida. Si se ajusta de manera correcta, se tiene como resultado un buen modelo.

Para ello como primer paso se realiza una etapa de preprocesamiento, donde las señales de audio se ajustan mediante el aumento de datos, la normalización, la eliminación de fragmentos silenciosos y la transformación a espectrogramas de Mel (Droguett, 2017).

En primer lugar, para el aumento de datos, se lleva a cabo un cambio de frecuencia, esto quiere decir, que se van escoger al azar los audios y se les va cambiar de frecuencia. Para ello, se utiliza el programa de Python, donde a los audios de tos normal y de asma se le varía el tono ajustado a un máximo de 20 %. Los nuevos audios se guardan en la equivalente carpeta de la base de datos, hasta tener un número similar de muestras de tos tanto con la tos normal y la de asma.

A continuación, se normaliza la amplitud y eliminación de silencio en cada grabación, que se explica más adelante, se definen los criterios para llevar a cabo esta tarea de forma heurística y por validación visual. Después de realizar este primer paso de preprocesamiento, las grabaciones iniciales de la base de datos quedan establecidas con una duración entre 0,3 segundos hasta 9 segundos. Lo que se busca es que todas las grabaciones tengan un valor modal de 2 segundos, para conseguir aquello igualmente se utiliza el programa de Python.

Por el contrario, las muestras de tos que tienen valores menores a 2 segundos se replican y se unen hasta alcanzar los 2 segundos.

4.3 Escala de MEL

Una escala de MEL es un mecanismo del tono propuesto por Stevens, Volkmann y Newmann en el años de 1937, ellos indicaron que es una escala de tonos juzgado por los diferentes oyentes como semejantes en distancia entre sí.

Debido a que los seres humanos perciben sonidos de una manera diferente cada uno, la escala de MEL es una escala no lineal y las distancias que se describió en el anterior párrafo van a aumentar con la frecuencia.

Para conseguir los MFCCs se siguen los siguientes pasos.

4.3.1 Preprocesado de la señal

Se normaliza la señal de audio ya que es la aplicación de una cantidad constante de ganancia, esto conlleva a que la amplitud de pico promedio sea llevada a un nivel imparcial normal, de ahí su nombre.

La normalización de un audio de tos es utilizada para poder conseguir el máximo volumen de una señal de audio seleccionada.

Una vez normalizado el audio, se realiza un encuadre de audio en el cual, se obtiene primero la transformada rápida de Fourier (FFT), que es un algoritmo que consiente obtener la transformada discreta de Fourier (DFT) y la inversa de la

misma, la FFT es utilizada en el tratamiento de las señales digitales (García & Mora, 2013).

Se debe tener en cuenta que el audio es un proceso no estacionario, por lo cual la FFT producirá distorsiones, por ello, se procede a dividir la señal en cuadros cortos, cada cuadro tendrá el mismo tamaño, también se desea que los marcos se superpongan, realizando esto los marcos van a tener correlación entre ellos, así se evita perder información en los bordes.

La frecuencia de muestreo es 44,1 KHz, esta es la frecuencia modelo para archivos de tos, las señales originales de tos de 44,1KHz pasan por un procesamiento de diezmado, donde el audio recibe un submuestreo para tener la nueva tasa de muestreo de 16 KHz y también se fraccionan en tramas de 20 ms con un desplazamiento de 10 ms (kagglle, 2018).

Después de realizar el encuadre se aplica una función de ventana en cada marco, si no se realiza este procedimiento se van a producir distorsiones de alta frecuencia. Es importante recalcar que primero se debe realizar el encuadre para después proseguir a la aplicación del FFT.

La ventana asevera que ambos extremos de la señal terminarán cerca de cero.

4.3.2 VENTANA DE HAMMING

La ventana de Hamming es una orientación que sirve para mejorar el filtrado de las señales, realiza el procedimiento de cortar los lugares de señal a cada lado, para que se pueda observar una imagen más despejada del espectro de frecuencia de la señal (Lima, s.f.).

Figura 11. VENTANA HAMMING

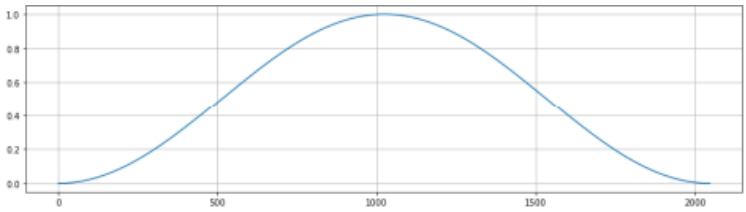
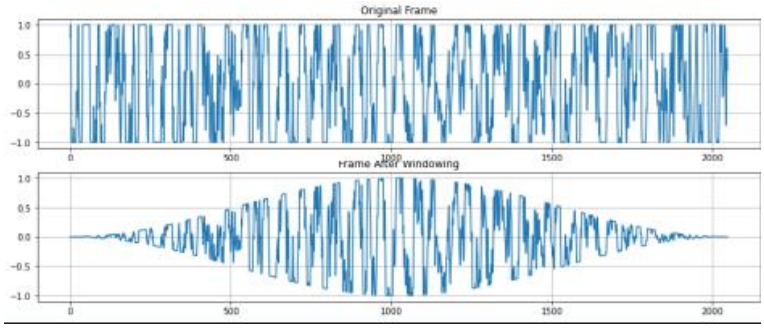


Figura 12. VENTANA HAMMING aplicado a una señal de audio



En la figura 11 se observa ambos extremos del marco que terminan en diferentes lugares en el eje y al aplicar la ventana de Hamming como se observa en la figura 12 se acercó los bordes de marco a cero.

Después de aplicar la ventana de Hamming se procede a realizar la FFT, se toma solo la parte positiva del espectro, como siguiente paso después de realizar estos procedimientos de calcula la potencia de la señal para luego ser utilizada en los bancos de filtro espaciados MEL.

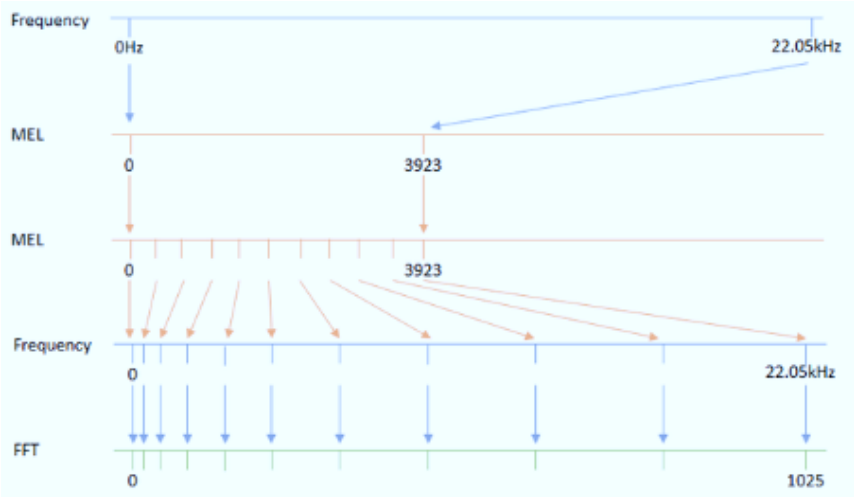
4.3.3 Banco de filtros espaciado MEL

En esta parte se calcula los bancos de filtros espaciado MEL, se pasa el audio enmarcado a través de los filtros, esto permite obtener información sobre la potencia en cada banda de frecuencia. Los filtros se consiguen para cualquier banda de frecuencia, en nuestro trabajo se fija en toda la banda muestreada.

El banco de filtros espaciados MEL le hace especial la parte en la que el espacio entre los filtros crece exponencialmente con la frecuencia. El banco de filtros se puede formar para cualquier banda sin restricciones.

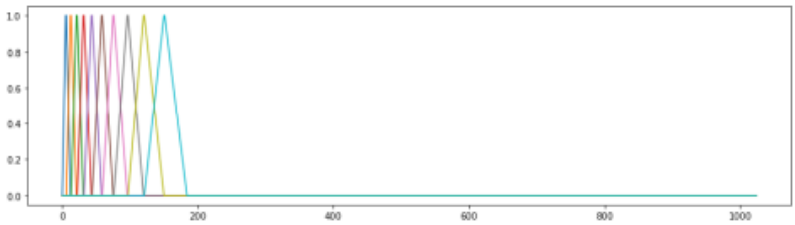
Para el cálculo de los puntos de filtro, primero se va a edificar puntos de filtro que van a establecer el inicio y la finalización de los filtros. Para realizar lo descrito anteriormente primero se convierten los dos bordes del banco de filtros al espacio de MEL. Después de realizar este procedimiento se convierte la matriz al espacio de frecuencia y finalmente se normaliza la matriz al tamaño de FFT y se prefiere los valores de FFT asociados.

Figura 13. Proceso del cálculo de puntos de filtro



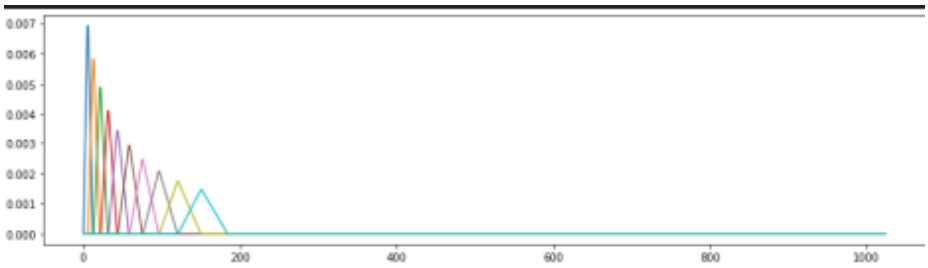
Después de obtener los puntos de filtro se procede con la construcción del banco de filtros.

Figura 14. Banco de filtros MEL



Como siguiente paso se divide los pesos MEL triangulares por el ancho de la banda MEL. En otras palabras de normaliza el área, si no se realiza la normalización va haber un aumento de ruido en la frecuencia debido al ancho del filtro.

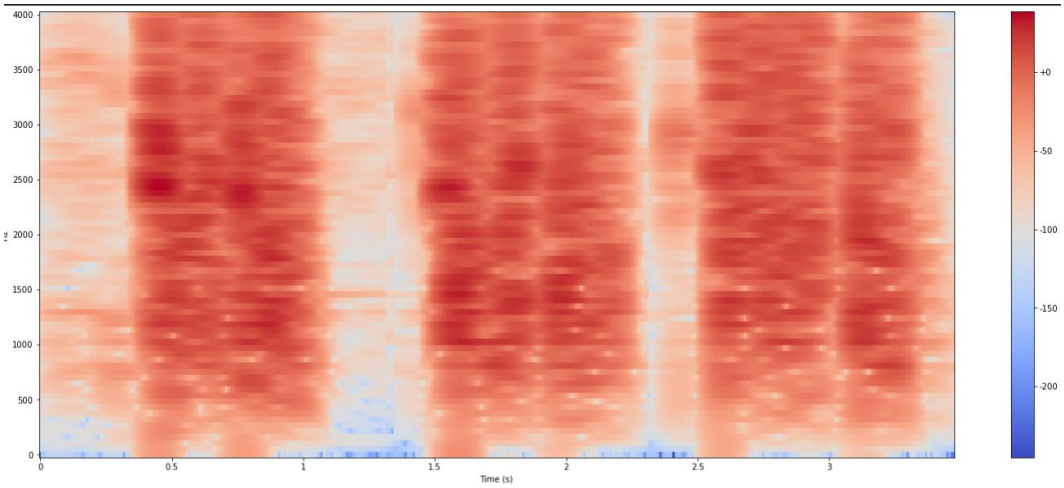
Figura 15 . Normalización del área



Al realizar este procedimiento se tiene una matriz que personifica la potencia de audio en los 20 filtros en desemejantes marcos de tiempo.

El paso final es la generación de los MFCCs, en este paso final se va a usar la transformada discreta del coseno (DCT), esta transformada extraerá cambios de alta y baja frecuencia en la señal, después se grafica el espectrograma de MEL figura 16.

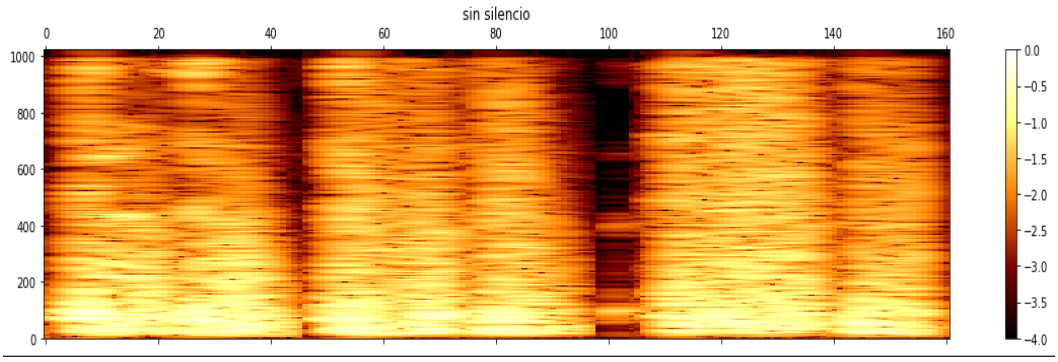
Figura 16 . Espectrograma de MEL de un audio de tos con asma



Después de obtener el espectrograma de MEL de los audios tanto de tos normal y tos con asma, se procede con la eliminación del silencio que tiene los audios, para eso se realiza un programa en Python, el cual elimina los silencios de los audios, para después mandarlos como entradas a las neuronas.

La figura 17 refleja el espectrograma final eliminado el silencio..

Figura 17 . Eliminación de silencio en los audios de la base de datos



4.4 Procesamiento con la Red Neuronal Convolucional (CNN)

4.4.1 Clasificadores

Para obtener el objetivo conclusivo de un método de aprendizaje automático, usualmente se establece una técnica entrenable para que se cumpla cierto espécimen de clasificación, como por ejemplo, la recuperación del contenido de los audios.

La primera tarea de clasificación es designada como localización de eventos de audio, esta detección está encaminada a descubrir señales de audio específicas dentro de un flujo de audio largo, la segunda tarea radica en la etiquetación de clase de un fragmento de audio, este se personifica por un vector de características.

Las orientaciones de clasificación son divididas en tres tipos: no supervisados, supervisados y semisupervisados. En el aprendizaje supervisado se utilizan etiquetas de clase predefinidas, esto ayuda en la creación de un modelo de clasificación, el cual va a determinar instintivamente una etiqueta de clase a informaciones desconocidas. Para el aprendizaje no supervisado no se tienen etiquetas predefinidas, esta metodología lo que busca es la exploración de los

datos y el poder detectar semejanzas entre las diferentes observaciones del espacio de características. Con respecto al aprendizaje semisupervisado se tiene el intermedio entre las dos metodologías, se basa en el etiquetado de una cantidad pequeña de datos y un gran conjunto de datos sin la etiquetación. El objetivo de esta metodología es el poder mejorar la ganancia del clasificador, así se logra superar la restringida disponibilidad de datos etiquetados (Carlos & Marta, 2001).

En este trabajo se utiliza la red neuronal convolucional que viene siendo un clasificador supervisado.

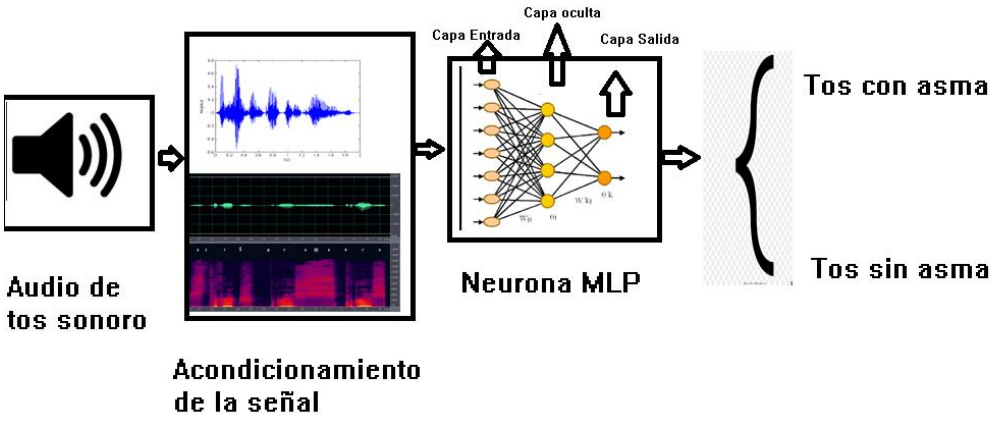
5. Resultados y discusión

En este fragmento de la tesis se observa los resultados obtenidos en donde la primera parte se entrenó una red neuronal perceptrón multicapa la cual es como línea base, por consiguiente, se entrena una red neuronal convolucional para detectar la tos si tiene asma o no, y se puede concluir cual es mejor dependiendo de los diferentes resultados que se obtiene en las matrices de confusión.

Para el análisis de los resultados se utilizará los valores de accuracy de la matriz de confusión y de manera más específica se analizarán los valores de especificidad y precisión para entender el comportamiento de cada algoritmo con cada clase de clasificación.

Para la Red Neuronal MLP se realiza el siguiente procedimiento figura 18.

Figura 18 . Diagrama de bloques del proceso aplicado para Red Neuronal MLP



En la programación de la Red Neuronal MLP utilizado en Python, se tiene n entradas, este valor depende de los datos, en este caso se utiliza 60300 neuronas de entradas. Se tiene también una activación Relu, esta función permite transformar los diferentes valores que se introducen cumpliendo con la anulación de los valores negativos y así se desiste de los valores positivos tal y como fueron

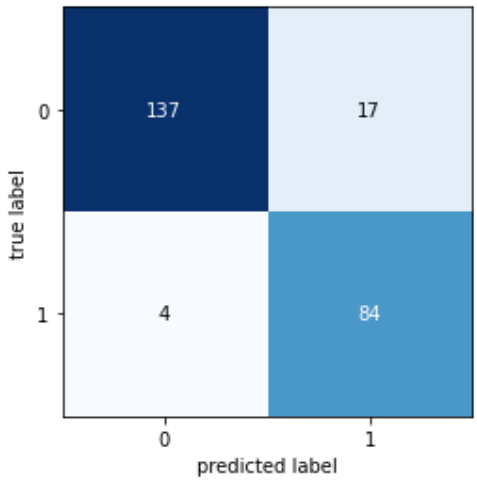
ingresados, posteriormente se tiene dos capas ocultas , las cuales tienen un valor de 1024 neuronas (Jaime, 2017).

Se configura con Dropout de 0.5 esto permite que se evite un oveefiting, esto quiere decir que es una de las técnicas que se utiliza para la regularización la cual se fundamenta en la exclusión de neuronas que se encuentran en las capas de la red neuronal, esta es aprovechada en base a la probabilidad conocida como la distribución de Bernoulli (Julio, R, Diez, & Carlos, 2020).

Al final se tiene dos neuronas de salida una por cada opción esto quiere decir una se tos con asma y una de tos sin asma, y se tiene la activación softmax, esta activación tiene la función de comprimir las salidas de cada una de las neuronas para que tengan valores de 0 y 1, de tal representación la suma de las salidas sea similar a uno, cabe recalcar que la función softmax causa salidas que tiene un semejante a probabilidades (Brownlee, 2020).

Después de la programación de la Red Neuronal MLP se consigue la matriz de confusión se observa un accuracy del 91,32 %, también se tiene una especificidad del 83,17 % de acierto lo que indica que el 83% de las muestras de la clase 1 serán clasificadas correctamente, finalmente los valores de precisión son de 97,16 % que sería el porcentaje de acierto al clasificar datos de la clase 0.

Figura 19 . Matriz de confusión red neuronal perceptrón multicapa

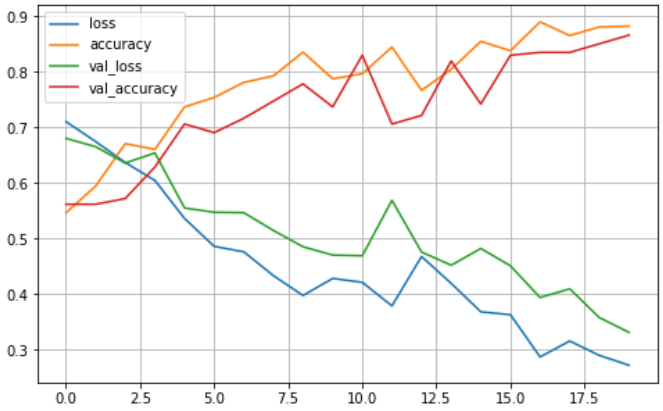


$$acc = \frac{137 + 84}{137 + 84 + 4 + 17} = 0.9132 \qquad Ec. 1$$

$$especificidad = \frac{84}{84 + 17} = 0.8317 \qquad Ec. 2$$

$$presicion = \frac{137}{141} = 0.9716 \qquad Ec. 3$$

Figura 20 . Métricas de entrenamiento de MPL



Para la Red Neuronal CNN se realiza el siguiente procedimiento figura 21.

Figura 21 . Diagrama de bloques del proceso aplicado para Red Neuronal Convolucional

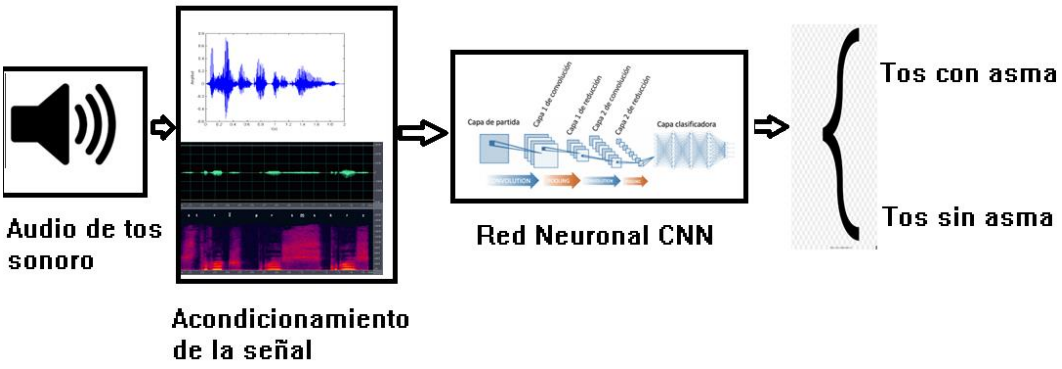
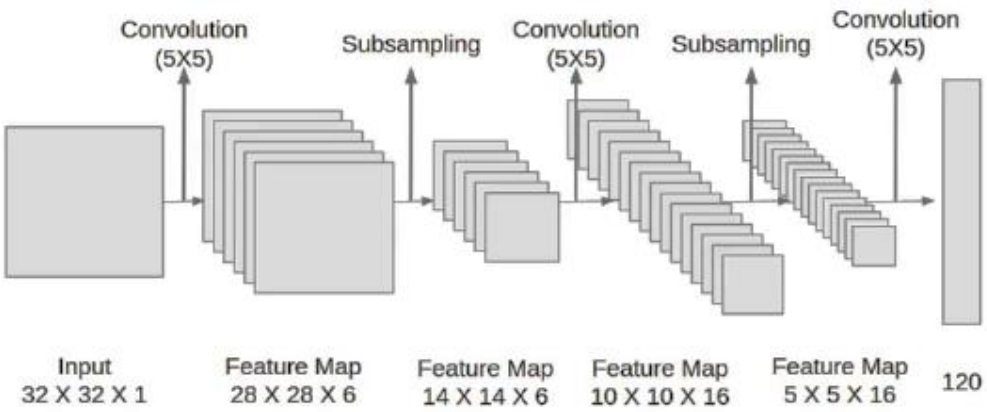


Figura 22 . Arquitectura de la Red Neuronal Convolucional



En el back-end, por otro lado, el espectrograma de Mel es analizado por una red neuronal convolucional (CNN), que puede determinar audio proviene de un sonido o de un evento de sonido. La arquitectura de red neuronal convolucional utilizada en este estudio se basa en la arquitectura LeNet-5, que es adecuada para conjuntos de datos pequeños. Además, según (Jaime, 2017) utilizando un pequeño conjunto de datos, la cifra de neuronas en cada capa varía.

De manera análoga a LeNet-5 (Diego, 2019), la CNN consta de cinco capas, divididas en dos capas convolucionales, dos capas completamente conectadas y una capa de clasificación binaria.

Se tiene un kernel de 3x3 el cual es un filtro que se emplea a la imagen en este caso a las de los espectrogramas para extraer algunas de las características significativos o patrones de la misma, que se utiliza para el proceso que va realizar la neurona en la determinación de si la imagen corresponde a un audio de tos normal o tos con asma (Jaime, 2017).

Las dos capas convolucionales tienen 64 unidades lineales rectificadas (ReLU) cada una. La primera capa convolucional tiene un filtro de tamaño 3x3 y toma los segmentos espectrales de Mel de 60300 como entradas. A continuación, se implementa una capa 2x2 de max-pooling, seguida de la segunda capa convolucional con un tamaño de filtro de 3x3 y otra capa de 2x2 de max-pooling.

Dos capas completamente conectadas siguen las capas convolucionales con 256 unidades para cada una, se tiene también una capa oculta de 32 neuronas.

Se utilizó la regularización de dropout con una tasa de 0,5 para reducir las probabilidades de sobreajuste. Finalmente, la última capa implementó la función ReLU para determinar si la señal de audio analizada es un sonido de tos u otro evento de sonido (Brownlee, 2020).

La función de pérdida utilizada fue la de entropía cruzada binaria. La relación entre el tamaño del conjunto de datos de entrenamiento y el lote para todos los modelos fue 8. Asimismo, se implementó una regularización de detección temprana con un monitor en la función de pérdida para el conjunto de datos de validación para reducir el sobreajuste. La arquitectura usada para todos los modelos se desarrolló en un entorno Python con la biblioteca Keras (Na8, 2018).

En el caso de la matriz de confusión de la red convolucional con una capa convolucional, se tienen los siguientes resultados:

$$acc = \frac{149 + 51}{149 + 51 + 3 + 39} = 0.8264 \qquad Ec. 4$$

$$especificidad = \frac{51}{51 + 3} = 0.9444 \qquad Ec. 5$$

$$presicion = \frac{149}{149 + 39} = 0.7925 \qquad Ec. 6$$

Donde se nota que el rendimiento en general de la red convolucional es del 82,64 %, el rendimiento con la clase 1 es del 94,44% y el rendimiento con la clase 0 es del 79,25 %.

Figura 23 . Matriz de confusión red convolucional (1 capa

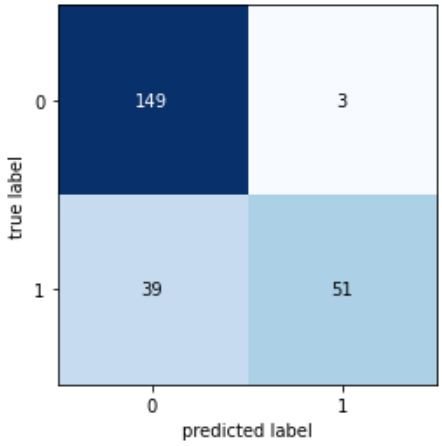
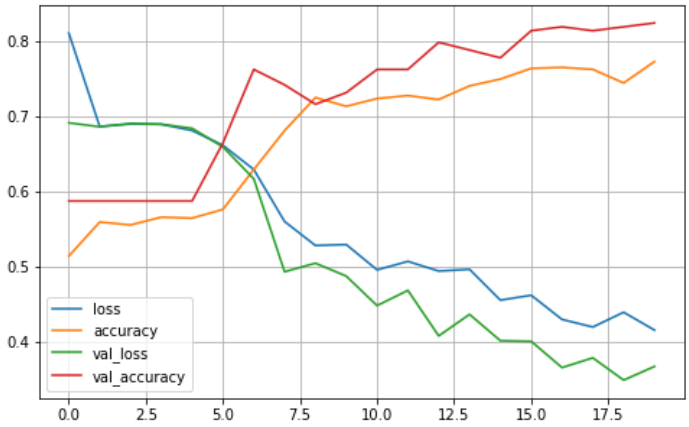


Figura 24 . Métricas de entrenamiento red convolucional (1 capa).



Por otro lado, al mejorar la red convolucional aumentado el número de capas convolucionales se tiene los siguientes resultados de la matriz de confusión:

$$acc = \frac{124 + 104}{124 + 104 + 0 + 14} = 0.9421 \qquad Ec. 7$$

$$especificidad = \frac{104}{104 + 14} = 0.8813 \qquad Ec. 8$$

$$presicion = \frac{124}{124 + 0} = 1 \qquad Ec. 9$$

Lo que da valores mucho mejores donde hay un accuracy del 94%, una especificidad del 88% y sorprendentemente una precisión del 100% lo que indica que la clase 0 clasifica correctamente todos los valores ingresados.

Figura 25 . Matriz de confusión red convolucional (2 capas)

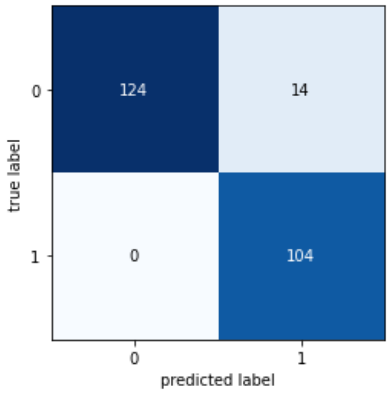


Figura 26 . Métricas de entrenamiento red convolucional (2 capas)

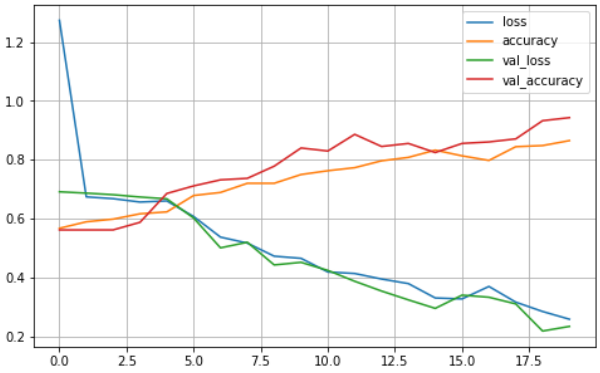
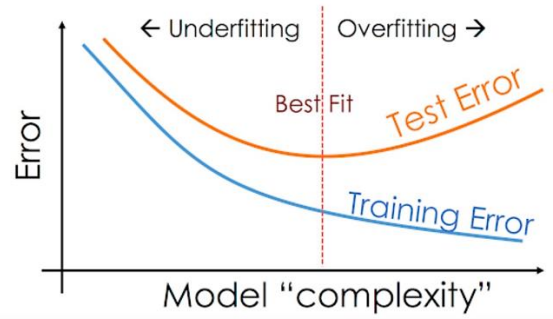


Tabla 2. Comparativa Resultados

	MPL	CNN 1 capa	CNN 2 capas
Accuracy	0,9132	0,8264	0,9421
Especificidad	0,8317	0,944	0,8813
Precisión	0,9716	0,7925	1

Figura 27 . Model Complexity y Overfitting in Machine Learning



Como se observa en la figura 27 se tiene la gráfica del Model Complexity esta misma se utiliza para el análisis en el machine learning, esto quiere decir que permite verificar si existe Overfitting (Kumar, 2022).

En la figura 24 y 26 se tiene la línea que representa Test error y Training error, se observa que las dos líneas tienden a una pendiente y no la apertura entre las dos líneas no es ancha, por lo cual se concluye que no hay Overfitting.

Figura 28 . Comparativa ACCURACY

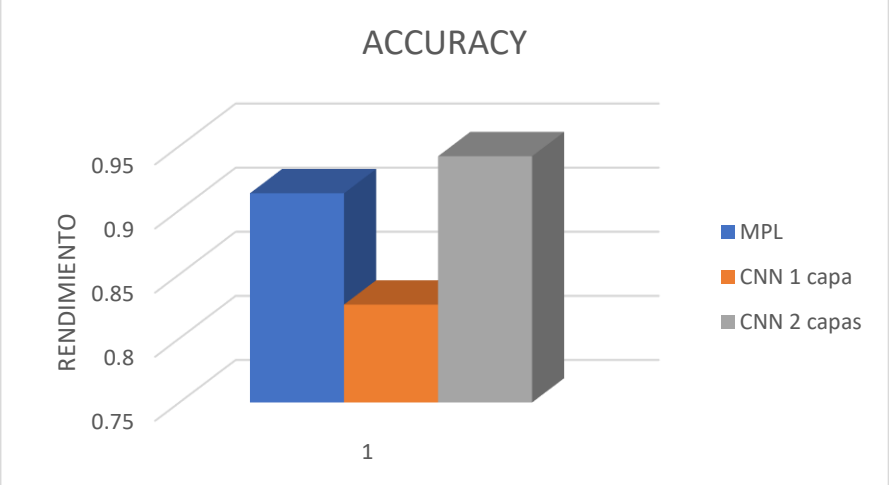


Figura 29 . Comparativa ESPECIFICIDAD

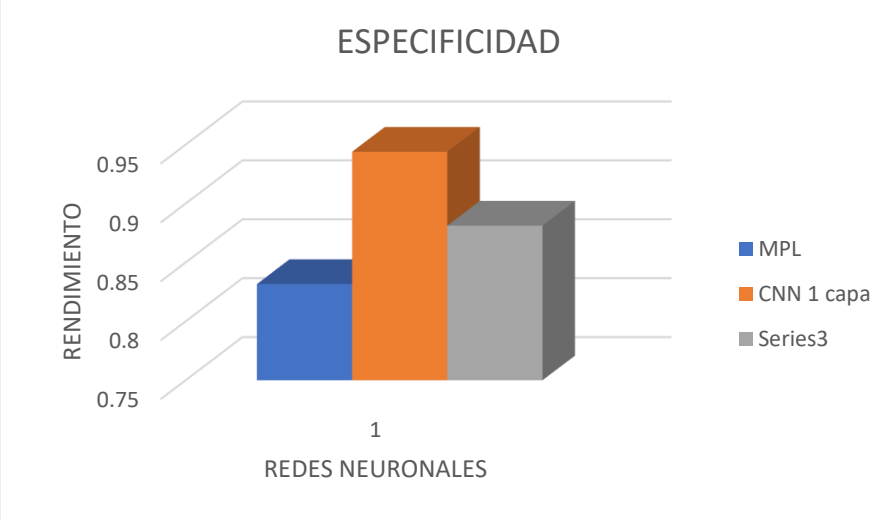
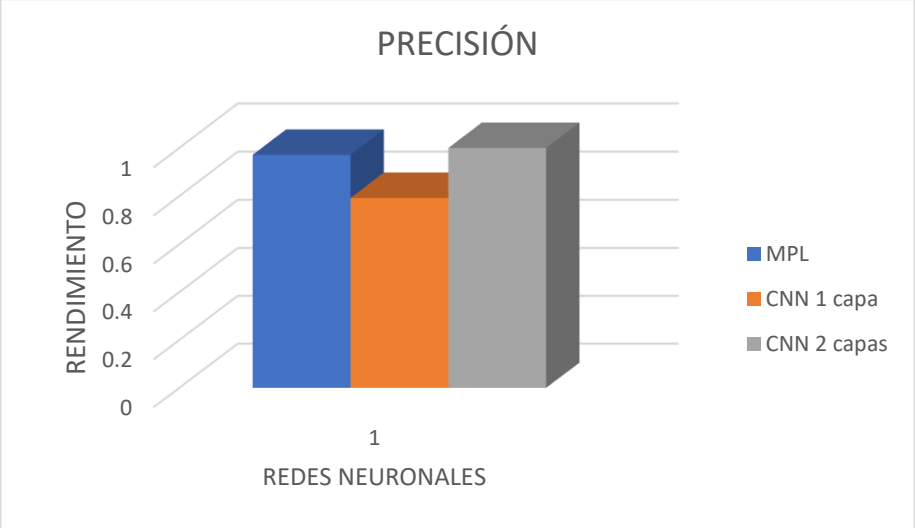


Figura 30 . Comparativa PRECISIÓN



6. Conclusiones

- La red MPL presenta un mejor desempeño en general en comparación a la red convolucional con una sola capa, pero al aumentar las capas vemos como esta tiene mejores resultados en todas las medidas obtenidas. Pese a que la red convolucional presento los mejores resultados hay que recalcar que la red MPL tubo valores muy altos pese a ser una red más simple a la convolucional, teniendo una diferencia máxima de 5%.

- Por otro lado, al momento de detectar si la tos es normal o con asma, la red neuronal convolucional (CNN) con una capa es la de peor rendimiento, pero si a esta misma se le aumenta una capa más pasa a ser la que mejor lo hace mejor, seguida muy de cerca por la MPL. La red neuronal convolucional mejorará tanto en precisión como en especificidad, tal como se observa en las figuras 29 y 30.

- Al basarse en la especificidad (Figura 29) se percata que una red convolucional (CNN) con una sola capa da excelentes resultados superando a la MPL e incluso a la CNN con dos capas por lo que la CNN con una capa es la mejor red neuronal artificial al momento de detectar tos con asma.

- Se concluye que las dos neuronas artificiales para los escenarios de tos, tienen una buena eficiencia, tanto la red neuronal MLP y la red neuronal convolucional (CNN) se puede utilizar para la detección de tos con asma con buenos resultados.

- La red neuronal convolucional, al extraer las principales características de cada imagen y al aprenderlas, es la que presenta mejores resultados en accuracy, precisión y especificidad. Dependiendo del número de capas convolucionales (en este caso 2) se ajusta mejor, siendo la red neuronal convolucional con dos capas la que obtiene mejores resultados en las tres métricas.

- La red neuronal artificial MPL demostró tener una mejor precisión en comparación a la red neuronal artificial CNN de una y dos capas, a pesar de ser la red neuronal base, tiene buenos resultados y se puede considerar para la detección de tos con asma.

Referencias

Ana, M. (2018). Aplicaciones de técnicas de inteligencia artificial basadas en aprendizaje profundo al análisis y mejora de la eficiencia de procesos industriales. Oviedo: Universidad de Oviedo.

Asensi, M. (2015). Educación en asma. Valencia: CS Serrería 1.

Barrios, K., López, J., Mendieta, S., & Benavides, R. S. (9 de Agosto de 2018). Sistema de reconocimiento de voz:un enlace en la comunicación hombre-máquina. *Revista De Iniciación Científica, 4*(4), 92-95.

Brownlee, J. (19 de Octubre de 2020). *Machine Learning Mastery*. Obtenido de https://machinelearningmastery.com/softmax-activation-function-with-python/

Carlos, R., & Marta, B. (2001). Redes Neuronales:Conceptos Básicos y Aplicaciones. Rosario: Universidad Tecnológica Nacional.

Ciudad, B. A. (14 de Febrero de 2022). *Buenos Aires Ciudad.* Recuperado el 17 de Septiembre de 2022, de https://www.buenosaires.gob.ar/jefaturadegabinete/innovacion/noticias/iatos-el-sistema-de-inteligencia-artificial-desarrollado-por

Clinic, M. (2020 de 06 de 2020). *Mayo Clinic*. (Mayo Foundation for Medical Education and Research (MFMER)) Recuperado el 30 de 08 de 2022, de https://www.mayoclinic.org/es-es/symptoms/cough/basics/definition/sym-20050846?p=1

Coronavirus, C. d. (14 de Febrero de 2022). *Buenos Aires*. Obtenido de Buenos Aires: https://buenosaires.gob.ar/jefaturadegabinete/innovacion/noticias/iatos-el-sistema-de-inteligencia-artificial-desarrollado-por#:~:text=Febrero%20de%202022-,Coronavirus%3A%20la%20Ciudad%20desarroll%C3%B3%20un%20sistema%20que%20analiza%20audios%20de,del%20m

Diego, A. (2019). ANÁLISIS DE SEÑALES DE TOS PARA DETECCIÓN TEMPRANA DE ENFERMEDADES RESPIRATORIAS. Valladolid: Escuela Técnica Superior de Ingenieros de Telecomunicación.

Droguett, Y. G. (4 de 12 de 2017). *scielo.* Recuperado el 30 de Agosto de 2022, de https://www.scielo.cl/scielo.php?pid=S0718-48162017000400474&script=sci_arttext&tlng=en

Edgar, S. (2017). *Desarrollo e Innovación en Ingeniería.* Medellín: Instituto Antioqueño de Investigación.

García, M., & Mora, G. (23 de Junio de 2013). *scielo.* Recuperado el 30 de Agosto de 2022, de https://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1139-76322013000300010

Ibrahim, N., Razak, Z., Tamil, E., Idna, M., & Yusoff, Z. (2008). Quranic Verse Recitation Feature Extraction using Mel-Frequency Cepstral Coefficient (MFCC). Malaya: Universidad de Malaya.

Jaime, D. (2017). Redes Neuronales Convolucionales en R. Sevilla: Escuela Técnica Superior de Ingeniería.

Juan, B. (11 de Junio de 2020). *Juan Barrios.* Recuperado el 10 de Septiembre de 2022, de https://www.juanbarrios.com/wp-content/uploads/2020/11/Instalaci%C3%B3n-Anaconda-Navigator-.pdf

Julio, C., R, H., Diez, G., & Carlos, M. (02 de Junio de 2020). Un estudio empírico del modelo de red neuronal MLP para problemas de predicción con salidas múltiples. *UCI*, págs. 1-14.

kagglle. (4 de Julio de 2018). *kaggle*. (kaggle) Recuperado el 10 de Septiembre de 2022, de https://www.kaggle.com/code/ilyamich/mfcc-implementation-and-tutorial

Kumar, A. (29 de Mayo de 2022). *Data Analytics.* Obtenido de https://vitalflux.com/model-complexity-overfitting-in-machine-learning/

Lima, A. (s.f.). *Acervo Lima*. (Acervo Lima) Recuperado el 11 de Septiembre de 2022, de https://es.acervolima.com/implementacion-de-ventana-de-hamming-y-rectangular-de-gibb-s-phenomenon/

M, R., D, A., J, B., & M, S. (2017). Actualización en asma. Madrid: Hospital Universitario Príncipe de Asturias.

Martín, V. (2004). *Farmacologia clinica y terapeutica medica.* España: Mc Graw Hill.

MedlinePlus. (16 de 07 de 2020). *MedlinePlus*. (MedlinePlus) Recuperado el 30 de 08 de 2022, de https://medlineplus.gov/spanish/cough.html#:~:text=La%20tos%20es%20un%20reflejo,de%202%20o%203%20semanas.

Medlineplus. (14 de Julio de 2021). *Medlineplus*. (team) Recuperado el 04 de Septiembre de 2022, de https://medlineplus.gov/spanish/ency/article/007535.htm

Na8. (29 de Mayo de 2018). *Aprende Machine Learning*. Obtenido de https://www.aprendemachinelearning.com/una-sencilla-red-neuronal-en-python-con-keras-y-tensorflow/

Rouhiainen, L. (2018). Inteligencia Artificial. Madrid: Alienta Editorial.