Universidad Politécnica Salesiana

TOS POR COVID-19

CARACTERIZACIÓN DESDE LA INTELIGENCIA ARTIFICIAL

Coordinador general
Christian Salamea Palacios - UPS

Coordinadores institucionales
Tarquino Sánchez Almeida - EPN
Xavier Calderón Hinojosa - EPN
Javier Guaña Moya - PUCE

Este trabajo es producto de la investigación del proyecto "Caracterización de la tos provocada por el COVID-19 en pacientes de diagnóstico positivo", financiado por CEDIA dentro su convocatoria a proyectos de investigación CEPRA XV. La publicación recoge la descripción de la propuesta; el proceso de diseño de la página web utilizada para la toma de muestras audibles de tos; la descripción de técnicas usadas para reconocer una señal de tos dentro de un audio utilizando aprendizaje automático; los sistemas de filtrado utilizados para aislar la señal de tos de cualquier sonido producido por circunstancias externas; y los modelos inteligentes pre-entrenados utilizados para la caracterización de la señal de tos como una tos COVID-19. Además consta información sobre las estrategias para reunir al equipo, generar la propuesta y conseguir su aprobación.

En síntesis, la obra presenta un caso exitoso de lo que es el desarrollo de un proyecto de investigación científica bajo la modalidad de financiamiento externo, con sus fases de planificación, ejecución y explotación de los resultados de investigación conseguidos.

ISBN-13: 978-9978-10-828-4
9 789978 108284

ABYA YALA | UPS
cedia
CORPORACIÓN ECUATORIANA PARA EL DESARROLLO DE LA INVESTIGACIÓN Y LA ACADEMIA
PUCE
ESCUELA POLITÉCNICA NACIONAL
DON BOSCO ECUADOR

Christian Salamea Palacios
Coordinador general

Tos por Covid-19

Caracterización desde la Inteligencia Artificial

ABYA YALA | UPS
cedia
CORPORACIÓN ECUATORIANA PARA EL DESARROLLO DE LA INVESTIGACIÓN Y LA ACADEMIA
PUCE
ESCUELA POLITÉCNICA NACIONAL

2023

Tos por Covid-19: caracterización desde la Inteligencia Artificial

Tarquino Sánchez Almeida - EPN; Xavier Calderón Hinojosa - EPN;
Javier Guaña Moya - PUCE (Coordinadores institucionales)

Autores: Christian Salamea Palacios, Tarquino Sánchez Almeida, Javier Guaña Moya, Xavier Calderón Hinojosa, Jessica Reina Trávez, David Romero Mogrovejo, Fernando Chica Ortiz, Paulo Castañeda Romero, David Naranjo, Santiago Luna.

1.ª edición:

© Universidad Politécnica Salesiana
Av. Turuhuayco 3-69 y Calle Vieja
Cuenca-Ecuador
P.B.X. (+593 7) 2050000
e-mail: publicaciones@ups.edu.ec
www.ups.edu.ec

CARRERA DE INGENIERÍA ELECTRÓNICA

ISBN UPS: 978-9978-10-828-4
ISBN digital: 978-9978-10-833-8

DOI: https://doi.org/10.17163/abyaups.16

Diseño, diagramación e impresión: Editorial Universitaria Abya-Yala
Quito-Ecuador

Tiraje: 300 ejemplares

Impreso en Quito-Ecuador, junio de 2023

Publicación arbitrada de la Universidad Politécnica Salesiana

Índice

Capítulo 4

Filtrado digital para señales audibles de tos

Capítulo 5

Modelado automático de una señal de tos COVID-19

Capítulo 6

Descripción del sistema integrado

Presentación

El 5 de mayo de 2023, la Organización Mundial de la Salud (OMS) da por finalizada la emergencia sanitaria internacional provocada por la COVID-19, esta decisión se toma tras 765 millones de diagnósticos y según la OMS calcula que ha cobrado 20 millones de vidas. La pandemia ha causado graves impactos económicos en todos los países del mundo, que ha dejado a su paso un gran conglomerado humano sumido en la pobreza, incluso llegó a cambiar el comportamiento humano, sin embargo, esta amenaza para la salud pública generada por la COVID-19 aún continúa. El director de la OMS Tedros Adhanom ha pedido a todos los países continuar con la vigilancia y la respuesta al SAR-CoV-2:

> Mientras hablamos, miles de personas en todo el mundo luchan por sus vidas en unidades de cuidados intensivos. Y millones más continúan viviendo con los efectos debilitantes de la COVID persistente. Este virus llegó para quedarse. Todavía está matando y todavía está cambiando. El riesgo sigue siendo que surjan nuevas variantes que provoquen nuevos aumentos en casos y muertes.

Por otro lado, el informe "Las Oportunidades de la Digitalización en América Latina frente al COVID-19", publicado por la Corporación Andina de Fomento y las Naciones Unidas en el 2020, destaca que el uso de las tecnologías de análisis avanzado como Big Data, Inteligencia Artificial (IA) y el poder de almacenamiento de la nube fueron muy útiles para mejorar el tratamiento de la enferme-

dad, reducir el tiempo requerido, el trabajo manual y prevenir más contagios mediante el monitoreo, lo que implica la importancia del uso cada vez más creciente del análisis de datos en el diagnóstico y tratamiento de enfermedades producidas por el SAR-COV-2. Así, tras el brote de COVID-19 y la contingencia sanitaria, los gobiernos del mundo, particularmente en América Latina, implementaron una serie de medidas para facilitar a la población el acceso a información oficial, servicios de educación a distancia y salud digital, a su vez los centros de investigación científica y universidades ponen a disposición de los investigadores bases de datos de acceso abierto con información provenientes de pacientes con diagnóstico positivo COVID-19.

En este contexto, la comunidad científica se encuentra uniendo esfuerzos para mitigar su alcance, la identificación temprana de la enfermedad, así como su oportuno tratamiento revisten especial interés en los académicos e investigadores. "Tos por Covid-19: caracterización desde la Inteligencia Artificial" recoge el procedimiento, metodología y resultados de investigaciones realizadas por el grupo de investigadores de la Universidad Politécnica Salesiana, la Escuela Politécnica Nacional y de la Pontificia Universidad Católica sede Quito, auspiciado por CEDIA, para la identificación temprana de la enfermedad de la COVID-19, mediante el uso de algoritmos de clasificación aplicando técnicas de Machine Learning, con el propósito de caracterizar a la señal de la tos provocada en pacientes con diagnóstico COVID-19 positivo, contribuyendo de esta manera al diagnóstico de estos pacientes, utilizando medios no invasivos como el hisopado.

El capítulo 1 comienza con el estudio de la COVID-19 a partir de señales audibles obtenidas de la tos de pacientes con diagnóstico positivo. Las muestras de tos se recogieron con la ayuda de una aplicación web, en la cual se despliega un cuestionario on-line para recolectar información manteniendo la confidencialidad del paciente voluntario, toda vez que ha sido llenado se procede a grabar la señal de tos, información que se presenta en detalle, su arquitectura y los componentes se detallan en el capítulo 2.

Los capítulos 3 y 4 hacen referencia al reconocimiento de la señal audible de tos y no tos y al filtrado de la señal respectivamente. Con relación al reconocimiento de la tos en una señal de audio y posteriormente separar las señales de tos de los ruidos externos, se utilizó un sistema de reconocimiento de patrones usando el modelo Redes Neuronales Convolucionales (CNN-Convolutional Neural Networks). Con relación al filtrado se utilizaron técnicas de filtrado digital y adaptativo.

En el capítulo 5 se presenta el Modelado Automático de la señal de tos COVID- 19, capítulo extenso que aborda en detalle las técnicas de caracterización de la señal de la tos de pacientes que poseen COVID-19 utilizando para ello los algoritmos de la red neuronal convolucional, tanto la nativa como la pre-entrenada. Los autores usan un modelo conformado de tres diferentes sistemas entrenados con "Transfer Learning" los cuales se emplean para decidir si la tos de entrada puede ser valorada o no como una tos-COVID, con el fin de reducir al mínimo un diagnóstico incorrecto. El modelo final conformado por tres arquitecturas o modelos; el primer modelo basado en una red neuronal convolucional nativa, el segundo modelo usa un algoritmo de Transfer Learning utilizando extractores de características basadas en la red pre-entrenada VGGIsh y la técnica de PCA para reducir la dimensionalidad de los datos y el tercer modelo consta de una red convolucional que usa como ingreso espectrogramas de Mel; se logra entonces, una precisión del 86,60 % y una exactitud del 88,76 % al diferenciar una persona sana de una que posee COVID-19, utilizando bases de datos modificadas con técnicas "data augmentation".

Finalmente, la descripción del Sistema Integrado utilizado se describe en el capítulo 6, el proyecto busca desarrollar un sistema que permita caracterizar y clasificar la señal de tos provocado por la COVID-19, por tanto, el sistema propuesto se compone de dos etapas: un Front-End para el pre-procesamiento de la señal y un Back-End para el modelado de la misma.

Los resultados finales son alentadores. Para la verificación del rendimiento del sistema se utiliza la base de datos utilizada por el grupo de investigación liderado por Cecilia Mascolo en la Universidad de Cambridge, para luego llevar a cabo una fase experimental con los audios recolectados en el Ecuador. Los resultados obtenidos experimentalmente mostraron que el enfoque de aprendizaje conjunto nos permite resolver las debilidades de algunos modelos con las fortalezas de otros. Se desarrolló un modelo robusto, con alta precisión de calificación, así como alta precisión y recuperación, en el sistema de reconocimiento de las señales de audio de tos.

Los autores

Introducción

En los últimos años, en el contexto de la pandemia del COVID-19, se puede describir que este proceso inesperado a nivel mundial ha afectado desde su brote inicial en Wuhan, China, en diciembre de 2019, y que ha generado una crisis sanitaria sin precedentes. Este nuevo virus denominado coronavirus, o también llamado SARS-CoV-2, ha causado millones de casos y ha cobrado la vida de un gran número de personas en todo el mundo. Ante esta situación, los gobiernos y las organizaciones de salud de todos los países han tomado medidas drásticas para contener la propagación del virus y proteger a la población.

Se puede decir que la pandemia del COVID-19 ha tenido un impacto significativo en todos los aspectos de la vida, salud, bienestar de las personas, economía global entre otras. También se han implementado medidas de distanciamiento social, cuarentenas y cierres de fronteras, lo que ha generado cambios en la forma en que interactuamos, trabajamos y nos relacionamos. Además, el desarrollo y la distribución de vacunas han sido cruciales en la lucha contra esta enfermedad.

En este contexto, la Organización Mundial de la Salud (OMS) ha desempeñado un papel fundamental en la coordinación de los esfuerzos internacionales para controlar la pandemia y brindar orientación a todos los países sobre las mejores prácticas en la prevención, detección y tratamiento del virus. Se han implementado medidas de monitoreo permanente de la población para localizar y controlar los focos de contagio, con el objetivo de reducir la propagación del virus.

Cabe destacar, que en muchos países, incluido el Ecuador, se ha enfrentado el desafío de realizar pruebas de detección masiva de

manera eficiente y efectiva, pero la falta de recursos, el alto costo de las pruebas y la falta de acceso a servicios de salud adecuados han limitado la capacidad de realizar pruebas de manera generalizada. Ante esta realidad se ha planteado la necesidad de explorar técnicas alternativas, que sean de bajo costo, de fácil acceso y que puedan brindar resultados confiables en la detección del virus.

En respuesta a esta necesidad ocasionada por la pandemia de la COVID-19, se han desarrollado y utilizado diversas aplicaciones tecnológicas para ayudar en la lucha contra la enfermedad. Estas aplicaciones han desempeñado un papel crucial en la detección de casos, el seguimiento de contactos, la educación pública y la gestión de recursos sanitarios.

A continuación se describen algunas de estas aplicaciones y su contribución durante la pandemia.

- Aplicaciones de rastreo de contactos: estas aplicaciones utilizan tecnologías como Bluetooth y GPS para rastrear los contactos cercanos de una persona infectada. Ayudan a identificar a las personas que podrían haber estado expuestas al virus y les proporcionan instrucciones sobre cómo proceder, como hacerse una prueba o ponerse en cuarentena. Estas aplicaciones han sido ampliamente utilizadas en muchos países para ayudar en la detección temprana de casos y frenar la propagación del virus.

- Aplicaciones de detección y diagnóstico: se han desarrollado aplicaciones que permiten a las personas realizar autoevaluaciones de síntomas relacionados con el COVID-19 y recibir recomendaciones personalizadas, como buscar atención médica o hacerse una prueba. Algunas de estas aplicaciones también ofrecen servicios de telemedicina, lo que permite a los pacientes comunicarse con profesionales de la salud de manera remota y recibir atención médica sin necesidad de acudir a un centro médico.

- Aplicaciones de educación y concienciación: estas aplicaciones brindan información actualizada sobre la pandemia, incluyendo consejos de prevención, estadísticas, noticias y recursos de salud. También pueden incluir funciones de seguimiento de síntomas y recordatorios para tomar medidas preventivas. Estas aplicaciones desempeñan un papel importante en la educación pública y la concienciación sobre la enfermedad, ayudando a las personas a tomar decisiones informadas y adoptar comportamientos seguros.

- Aplicaciones de gestión de recursos sanitarios: durante la pandemia, la gestión eficiente de los recursos sanitarios se ha vuelto crucial. Se han desarrollado aplicaciones para ayudar en la gestión de camas de hospital, suministros médicos y recursos humanos. Estas aplicaciones facilitan la coordinación entre los centros de salud, optimizan la asignación de recursos y ayudan a garantizar que los pacientes reciban la atención adecuada.

Entre los síntomas característicos de la enfermedad, la tos ha sido mencionada con frecuencia por el Ministerio de Salud Pública del Ecuador, por tal motivo, en este estudio, se plantea la hipótesis de que en el sonido de la tos se podría encontrar características distintivas que permitan identificar la presencia de la COVID-19 en un individuo. Cabe destacar que se ha tomado como base el análisis matemático y computacional de las señales de audio de la tos, ya que pueden ofrecer un enfoque técnico-científico para este propósito, sin requerir muestras biológicas.

Con el fin de investigar esta hipótesis se ha desarrollado un sitio web, en la cual se ha recopilado información voluntaria y anónima de señales de tos proporcionadas por la población. Estas señales serán analizadas utilizando técnicas de Aprendizaje Automático, Procesamiento Digital de Señales e Inteligencia Artificial para extraer características digitales específicas de la tos asociada con la COVID-19. Este estudio constituirá una base para futuras investigaciones que permitan

desarrollar un sistema de diagnóstico basado en estas características, en conjunto con otros síntomas relacionados.

Es importante destacar que los participantes en este estudio proporcionan información a través de un cuestionario en línea, indicando si padecen o no COVID-19 y si han sido diagnosticados mediante prueba PCR. Este enfoque descriptivo transversal exploratorio busca obtener datos relevantes y confiables que respalden los resultados obtenidos.

Este estudio se centró en el análisis de señales acústicas de tos donadas voluntariamente por la población a través de una aplicación web. El objetivo no es diagnosticar la COVID-19, sino determinar características propias de la tos asociada a la enfermedad mediante el uso de sistemas de aprendizaje automático. La metodología se basa en la recopilación de señales de tos mediante encuestas en línea, similar a otros proyectos en instituciones como la Universidad de Cambridge y el MIT. El enfoque se centra en extraer características digitales temporales y espaciales de la tos durante el proceso de expectoración.

Capítulo 1

Abordaje del estudio del COVID-19 a partir de señales audibles de tos

Introducción

En el contexto de la pandemia del COVID-19, la Organización Mundial de la Salud (OMS) ha recomendado monitorear permanentemente a toda la población para localizar focos de contagio y tratar de reducir la propagación del virus. Dada la realidad en el Ecuador, donde el número de pruebas realizadas por cada millón de habitantes no es de los más altos en la región, se ha considerado estudiar técnicas alternativas, que además de ser gratuitas o de bajo costo, sean de fácil acceso a la población y que se puedan realizar estos testeos con un buen margen de confianza y efectividad.

Según la información difundida por el Ministerio de Salud Pública del Ecuador, en torno a los cuidados que se deben tener en cuenta para el control de la pandemia, se indica frecuentemente que la tos puede ser uno de los síntomas característicos de la COVID-19. La hipótesis que sustenta este trabajo está enfocada en que, probablemente, en el sonido de la tos se podrían encontrar características que podrían identificar una tos que proviene de pacientes con COVID-19.

Una señal de tos es una señal de audio, que puede ser captada remotamente y analizada mediante procesos matemáticos y computacionales, y así, sin necesidad de muestras biológicas, se podría llevar a cabo un estudio técnico-científico de procesamiento de señales y buscar características propias de una tos COVID-19.

Para conseguirlo, por un lado, se ha recopilado información de señales de tos donadas por la ciudadanía de manera voluntaria y anónima por medio de una aplicación web desarrollada para dicho objetivo, y, por otro lado, con los patrones característicos de las señales obtenidas, se pretende extraer las características digitales de una tos COVID-19, utilizando técnicas de Aprendizaje Automático, y Procesamiento Digital de Señales e Inteligencia Artificial. Este estudio constituye una base para una futura investigación que permitirá realizar un diagnóstico de la enfermedad utilizando dichas características, junto con otros síntomas típicos relacionados. En este sentido, vale la pena aclarar que para este trabajo que corresponde a un estudio descriptivo transversal exploratorio, en ningún caso tiene como objetivo final el diagnóstico de las personas que tienen COVID-19, más bien, está enfocado a determinar, por medio de sistemas de aprendizaje automático, las características propias de una tos COVID-19, analizada a partir de la señal de audio proporcionada por los participantes, quienes, mediante un cuestionario on-line incluido en la misma página web donde pueden grabar la tos, indican si poseen o no COVID-19 y si han sido diagnosticados con una prueba PCR.

El presente trabajo ha tomado impulso gracias al hecho de que otros grupos de investigación alrededor del mundo se han abocado a esta tarea utilizando una metodología similar. Por ejemplo, las propuestas generadas en la Universidad de Cambridge (https://www.covid-19-sounds.org/es/) o el Massachusetts Institute of Technology (MIT) (https://opensigma.mit.edu/), donde toman las características de la señal de tos de los participantes a través de una encuesta y por medio de estimaciones probabilísticas y la aplicación de sistemas de aprendizaje automático, reconocen patrones característicos de una tos

COVID-19. Si bien, la metodología aplicada para el trabajo no incluye el contraste con datos certificados que confirmen la positividad o no de un participante, se confía en que el volumen de información que se utiliza para entrenar los sistemas será suficiente para garantizar la validez de los resultados. Este estudio conlleva un bajo y nulo riesgo para los participantes, tanto en cuanto, la donación de la señal de la tos se la realiza remotamente, desde cualquier lugar aislado y con el uso de dispositivos electrónicos particulares.

Para la realización de este estudio es necesaria una base de datos de señales de audio de la tos de personas sanas y de personas con un diagnóstico positivo de COVID-19. Para su obtención se ha implementado una página web (https://databasecovid19.ups.edu.ec/), donde las personas de forma libre y voluntaria pueden grabar la señal de su tos, la cual queda almacenada en una base de datos y sirve posteriormente para realizar los análisis correspondientes. En esta página web, mediante una encuesta, se recopila la señal digital de la tos y los metadatos requeridos para el estudio. Los metadatos corresponden a las preguntas orientadas a definir el tipo de sintomatología, el tiempo de contagio, la edad y el género de la persona. En ningún caso, la encuesta está orientada a identificar a los participantes, ya que se garantiza el derecho a la confidencialidad de la información conforme a la Ley Orgánica del Sistema Nacional de Registro de Datos Públicos.

Una vez cargados los datos en la correspondiente base, el primer paso que se lleva a cabo es el preprocesamiento de todos los audios recopilados, con el fin de extraer únicamente las señales de tos encontradas. Posteriormente se utilizará la información extraída para, mediante procesamiento digital de señales, obtener las características fundamentales de la señal de la tos provocada por la COVID-19, con las cuales se espera obtener patrones discriminativos que en un futuro puedan llevar a diferenciar una señal de audio de tos provocada por la COVID-19 de la tos provocada por otras sintomatologías.

Al finalizar este estudio se tendrá una visión clara de los parámetros que caracterizan la señal de la tos provocada por la CO-

VID-19, además de una idea de los alcances que presta la tecnología del aprendizaje automático y la inteligencia artificial para esta tarea en particular. Este trabajo constituye la base para estudios posteriores, que permitirán la implementación de un sistema de diagnóstico de la COVID-19 por medio de la caracterización de la señal de la tos.

En este punto, los principales beneficiarios serán la comunidad y las instituciones, dado que se contará con una herramienta confiable y de fácil acceso para la población.

Con esa visión se ha vislumbrado la idea de una aplicación informática que, a distancia y con base en la señal de la tos emitida por una persona en un dispositivo móvil, se determine si la misma corresponde a una tos COVID-19 o a una tos NO COVID-19.

Elaboración de la aplicación web

Este estudio está orientado al análisis de información de señales acústicas producidas por la tos, señales donadas de manera voluntaria por parte de la ciudadanía en general, tanto por personas con diagnóstico positivo de COVID-19, como de aquellas con diagnóstico negativo; por medio de una aplicación web a la que se podrá acceder, ya sea, desde un computador personal o desde un smartphone.

Como se ha mencionado previamente en ningún caso se evalúa la capacidad diagnóstica de los sistemas inteligentes obtenidos del trabajo desarrollado, sino que el estudio está enfocado a determinar, por medio de sistemas de aprendizaje automático, las características propias de una tos COVID-19 analizada a partir de la señal de audio proporcionada por los participantes, quienes, mediante un cuestionario on-line incluido en la misma aplicación web, pueden grabar la tos, indicando si poseen o no COVID-19. También se consulta si han sido diagnosticados con una prueba PCR. Así, la metodología de recolección de datos de las señales de tos, está basada en un concepto similar al realizado, por ejemplo, en la Universidad de Cambridge o en el MIT, donde toma la señal de tos de los participantes a modo de

encuesta, a través de aplicaciones web y por medio de estimaciones probabilísticas y la aplicación de sistemas de aprendizaje automático intentan reconocer patrones característicos de una tos COVID-19.

La información audible que se requiere para este estudio es la tos de las personas, la que puede aparecer, ya sea, de manera intencional o natural, puesto que al estudio le interesa determinar las características digitales temporales y espaciales que aparecen durante el proceso de expectoración (Sharma *et al.,* 2020). En este contexto, si bien se espera obtener el mayor número posible de señales de tos, y por ende de participantes del experimento, que lo llevaría a definirse como un experimento aleatorio; también es cierto que se requiere contar con una base de datos balanceada, donde las señales de tos de diagnóstico positivo sean comparables con las señales de tos de diagnóstico negativo, con lo que la investigación estaría en la categoría de "controlado" (Orlandic, 2020). En este sentido, dado que las técnicas a utilizarse para procesar la información no están basadas en estadística tradicional, no se ha definido una muestra de estudio, pues se espera trabajar con la mayor cantidad de datos disponibles para entrenar los modelos de aprendizaje automático y serán las métricas obtenidas las que nos indiquen si los datos utilizados para el entrenamiento han sido suficientes o no.

El objetivo principal de la página web es la recolección de grabaciones de tos de personas con un diagnóstico positivo o negativo de COVID-19. Los participantes, al momento de grabar su tos deben estar previamente informados sobre el propósito del estudio en el cual están participando. Por esta razón se considera eficiente y conveniente que la página web conste de una página principal informativa y una segunda página para la grabación de la tos. El diseño del aspecto visual y funcional de estas páginas se las realiza considerando los siguientes criterios:

a. Simplicidad y facilidad de manejo: en el diseño de la página web se busca que el manejo sea fácil e intuitivo para las personas,

con esto se trata de evitar dificultades en el manejo, lo que podría llevar a un abandono de la página web.

b. Tipo y cantidad de información a colocar: se busca que la información a colocar sea lo más corta y concisa posible para la persona participante, abarcando todos los aspectos importantes al estudio.

1. Página principal informativa: según los criterios descritos, a continuación se muestra en la figura 1 los bosquejos (*sketches*) de la página principal que se realiza con múltiples subsecciones.

Figura 1

Aspecto visual de la página principal del sitio web

Encabezado prrincipal
Información del proyecto
Video tutorial para la grabación de Tos
Infprmación del Grupo de Investigación y Entidades participantes

2. Página secundaria: esta página se crea con el objetivo de recopilar la información, tanto audible como de metatexto, para la base de datos. Una vez que la persona se haya informado sobre el proyecto en la página principal, la página principal pasa a la secundaria para realizar la grabación de la tos. Al ingresar a esta página, al usuario se le presenta una ventana donde deberá aceptar el "consentimiento informado" y posteriormente el botón para iniciar con la grabación de la tos. Posteriormente se realiza la recolección de metadatos por medio de un cuestionario on-line y luego se culmina el proceso con el envío de la información. Las preguntas están orientadas a definir el tipo de sintomatología que podría estar relacionada a la señal acústica de la tos emitida, así como para definir la posible etapa de contagio. La página secundaria dentro del aspecto visual contiene los elementos descritos en la figura 2.

Figura 2

Aspecto visual de la página secundaria del sitio web

Consentimiento informado
Grabación de la tos
Obtención de los metadatos
Envío de la información

Referencias bibliográficas

Orlandic, L., Teijero, T. y Atienza, D. (2020). *The COUGHVID crowdsourcing dataset: A corpus for the study of large-scale cough analysis algorithms* (Version 1.0) [Data set]. Zenodo. http://doi.org/10.5281/zenodo.4048312

Sharma, N., Krishnan, P., Kumar, R., Ramoji, S., Chetupalli, S. R., Ghosh, P. K. y Ganapathy, S. (2020). *Coswara-A Database of breathing, cough, and voice sounds for COVID-19 Diagnosis*. arXiv preprint arXiv:2005.10548.

Capítulo 2

Toma de muestras e interpretación de datos

Implementación de la aplicación web

En este capítulo se presentan los pasos que se siguieron para la implementación de la aplicación web y las pruebas de funcionamiento correspondientes, la que permite obtener las señales de tos de la ciudadanía.

Como punto de partida, la página web se puede ver en la figura 1 y la figura 2 (https://databasecovid19.ups.edu.ec/).

Figura 1

Página web para almacenamiento de audio de tos

Figura 2
Página web para almacenamiento de audio de tos

La estructura de la aplicación web parte de una arquitectura cliente-servidor, constituida por los siguientes componentes: la carga (Fat Client, Fat Server), el servicio que entrega (de BDD, ficheros, web, proxy, objetos y transacciones) y la distribución de funciones (presentación distribuida, presentación remota, acceso a datos remoto, BDD distribuidas, lógica o proceso distribuido) (Peterson, 2014; Duckett, 2011; Prettyman, 2018; Dyer, 2015).

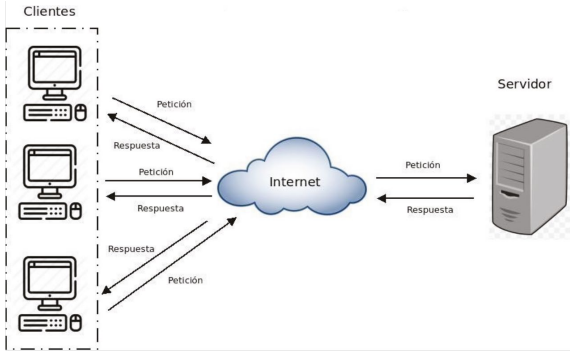
Para el tipo de aplicación que se ha implementado, la forma ideal de representar el sistema es una arquitectura cliente-servidor, que consiste en un modelo multicapas, que corresponde a una división de la clasificación por tamaño de componentes, tanto a nivel software (SW) como hardware (HW). En el primer caso, se refiere a servidores de aplicación distribuidos a lo largo de una red, pudiendo realizarse en dos y tres capas, según el modo de envío de mensajes desde el cliente, y la respuesta generada en relación con la devolución de información (Duckett, 2011; Dyer, 2015). Cada sistema presenta ventajas y desventajas dependiendo de ciertas variables, como el tráfico de información ocasionado o la simplicidad del lenguaje utilizado. En el segundo caso, el modelo se enfoca en la distribución de los procesos y elementos entre los componentes, donde la administración de la interfaz gráfica

se asocia a los clientes (Computador Personal-PC) y la seguridad e integridad de los datos se asocian a servidores locales y/o centrales.

En la figura 3 se observa igualmente en dos y tres capas según el modo de acceso a la base de datos (Peterson, 2014; Prettyman, 2018).

Figura 3

Arquitectura del sistema cliente-servidor



Tipos de arquitectura cliente-servidor

a. Por tamaño de componentes (Peterson, 2014)

Se basa en quién lleva la mayor carga de procesos. Se clasifica en dos tipos:

- Fat Client: el peso de la aplicación es ejecutada por el cliente.
- Fat Server: el peso de la aplicación es ejecutada por el servidor, el cliente solo tiene la interfaz de usuario.

b. Por naturaleza del servicio (Peterson, 2014; Frain, 2015)

- Servidores de Ficheros: con un servidor de archivos, el cliente es quien realiza el requerimiento de estos sobre una red.
- Servidores de Bases de Datos: permite que un proceso cliente solicite datos y servicios directamente a un servidor de bases de datos.

- Servidores de Transacciones: el proceso cliente llama a funciones, procedimientos o métodos que residen en el servidor.
- Servidores de Objetos: las aplicaciones cliente/servidor son escritas como un conjunto de objetos que se comunican.
- Servidores Web: este nuevo modelo consiste en clientes simples que hablan con servidores Web, mismos que devuelven documentos cuando el cliente pregunta por el nombre de estos.
- Servidores Proxy: permiten administrar el acceso a internet en una red de computadoras permitiendo o negando el acceso a diferentes sitios Web.

c. Por reparto de funciones entre cliente y servidor (Robbins, 2012)

Las distintas arquitecturas cliente-servidor varían en su forma de operar sobre la base de tres conceptos generales:

- La lógica de acceso a datos: funciones que gestionan todas las interacciones entre el SW y los almacenes de datos.
- La lógica de presentación: funciones que gestionan la interfaz entre los usuarios del sistema y el SW.
- La lógica de negocio o lógica de la aplicación: funciones que transforman entradas en salidas.

Según cómo se distribuyen estas tres funciones se puede realizar la siguiente clasificación (Duckett, 2011; He, 2015):

- Presentación distribuida: el cliente asume parte de las funciones de presentación de la aplicación, ya que siguen existiendo programas en el servidor dedicados a esta tarea. El resto de las funciones de la aplicación residen en el servidor.

- Presentación remota: toda la lógica de negocio y acceso a datos se ejecuta en el servidor. Todas las funciones de presentación son ejecutadas en el cliente.

- Lógica o proceso distribuido: la lógica de los procesos se divide entre los distintos componentes del cliente y del servidor. El diseñador de la aplicación debe definir los servicios y las interfaces del sistema de información de forma que los papeles de cliente y servidor sean intercambiables, excepto en el control de los datos que es responsabilidad exclusiva del servidor.

- Acceso a datos remoto: el cliente realiza tanto las funciones de presentación como los procesos. El servidor almacena y gestiona los datos que permanecen en una base de datos centralizada. En esta situación se dice que hay una gestión de datos remota.

- Bases de datos distribuidas: similar al modelo anterior, pero además el gestor de base de datos divide sus componentes entre el cliente y el servidor. Las interfaces entre ambos están dentro de las funciones del gestor de datos y, por lo tanto, no tienen impacto en el desarrollo de las aplicaciones. En este nivel se da lo que se conoce como bases de datos distribuidas.

d. Modelos Multicapas (Sklar, 2011; Abu-Naser, 2017)

Una de las clasificaciones mejor conocidas de las arquitecturas cliente-servidor se basa en la idea de capas, la cual es una variación sobre la división o clasificación por tamaño de componentes.

Dentro de esta categoría tenemos las aplicaciones en dos capas, tres capas y multicapas. Este término se utiliza indistintamente para referirse tanto a aspectos lógicos (software) como físicos (hardware).

A nivel de software: permite hablar de servidores de aplicación distribuidos a lo largo de una red (Sklar, 2011; Abu-Naser, 2017).

Dos capas: conexión directa entre el proceso cliente y un administrador de bases de datos.

Tres capas: el cliente envía mensajes directamente al servidor de aplicación, el cual debe administrar y responder todas las solicitudes. El servidor es quien accede y se conecta a la base de datos (He, 2015; Sklar, 2011).

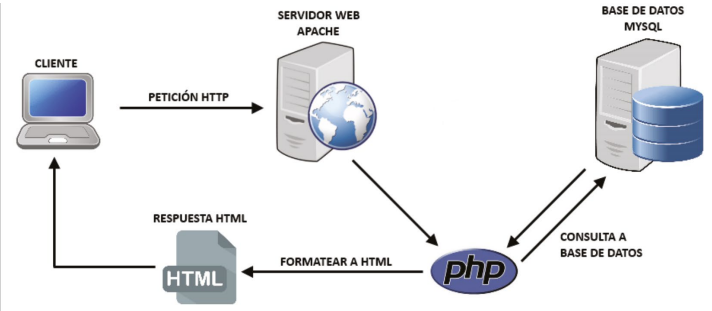
Para optimizar los sistemas no existe una estructura estática, por lo que muchas veces un servidor se comporta como cliente; este tipo de acciones son las que se denominan cooperaciones cliente servidor (He, 2015; Abu-Naser, 2017).

Procedimiento

El sistema propuesto es un servidor de tipo "Fat Server" de tres capas debido a sus ventajas de escalabilidad y fácil mantenimiento. En este contexto, los clientes web serían todos los dispositivos que realicen una consulta en el sitio web. El servidor web está basado en código HTML, el cual adquiere la información mediante código PHP y almacena esta información en una base de datos (SQL). La figura 4 muestra la arquitectura de este sistema (Duckett, 2011; Abu-Naser, 2017; Prettyman, 2018; Dyer, 2015).

Figura 4

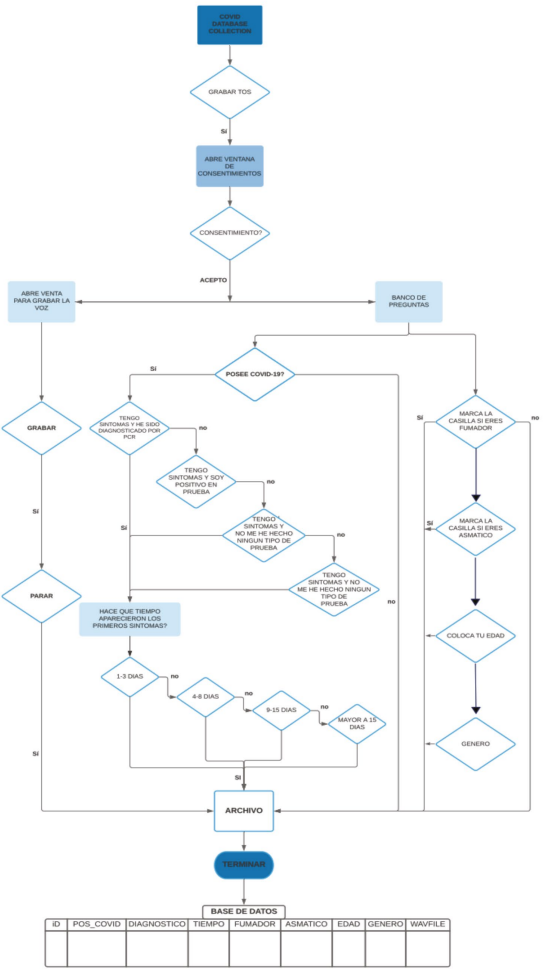
Arquitectura de sistema cliente-servidor para almacenamiento de audio de tos mediante una página web



Tanto el archivo de audio, como los archivos de datos del cliente que se reciben a través de la interfaz de usuario de la aplicación web desarrollada, son almacenados en una base de datos. La lógica del uso del sitio web puede ser observada en el diagrama de flujo de la figura 5.

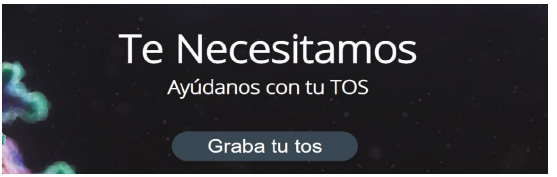
Figura 5

Diagrama de flujo sistema cliente servidor para almacenamiento de audio de tos mediante una página web



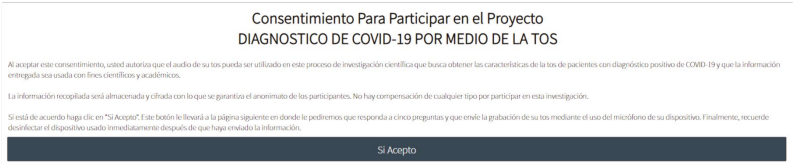
La lógica de acceso de este sistema inicia en la página principal con un botón que en la parte superior tiene un mensaje que dice "Te necesitamos Graba tu tos", como se puede ver en la figura 6.

Figura 6
Botón de inicio



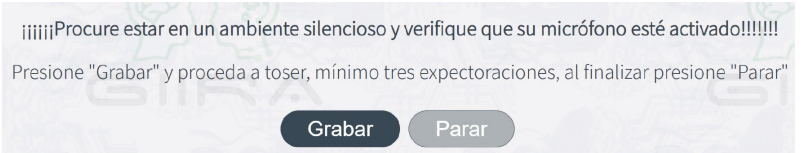
En la figura 7 se muestra al usuario un consentimiento informado, este tiene como principal objetivo informar al usuario sobre el proyecto y solicitar al usuario que autorice el uso de la grabación de su tos.

Figura 7
Consentimiento informado



El siguiente paso se ve en la figura 8, en el cual se informa al usuario cómo realizar la grabación de la tos. Para ello, se cuenta con dos botones, uno de *grabar* y otro de *parar*. Al momento de pulsar *grabar*, el sistema procede a registrar las grabaciones del micrófono del dispositivo cliente.

Figura 8
Grabación de tos



Dentro de este contexto y con el objetivo de diferenciar entre los diferentes tipos de tos, además de la señal audible de la tos, se solicitó la respuesta a una serie de preguntas de forma anónima, para proteger la privacidad de las personas (Alhawiti *et al.,* 2015; Deng, *et al.,* 2013). Las preguntas que se proponen en la encuesta dentro de la aplicación web son:

Pregunta 1: en esta pregunta se le consultará a la persona participante si es positivo o negativo para COVID-19.

Pregunta 2: si la persona participante ha seleccionado que es positiva para COVID-19, se le consulta si tiene síntomas, si se ha hecho una prueba PCR y si esta ha resultado positiva o negativa. El objetivo de esta pregunta es recabar información para determinar si la persona participante es asintomática o no, además de conocer si la muestra de su tos es validada con una prueba PCR positiva, teniendo de esta manera más conocimiento sobre la muestra de tos recolectada de las personas participantes.

Pregunta 3: esta pregunta tiene como objetivo saber el tiempo de aparición de síntomas en la persona participante, en caso de tenerlos. Si la persona es asintomática se pregunta hace cuántos días obtuvo su prueba PCR positiva. En esta pregunta se poseen diferentes intervalos de tiempo los cuales fueron seleccionados para tener un conocimiento de la fase en la que se puede encontrar la enfermedad. Como se expresa en Ullah *et al.* (2020), las muestras son útiles siempre y cuando los síntomas hayan aparecido hace un tiempo máximo de 14/15 días, dado que luego de ese tiempo la carga viral suele descender y las muestras pierden confiabilidad.

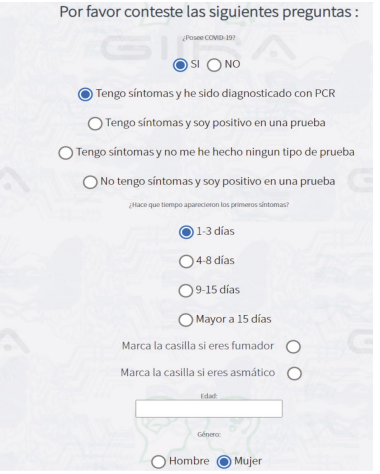
Pregunta 4: en esta pregunta se consulta a la persona participante si es fumador/a y asmático/a, esto debido a que según Deeks *et al.* (2020) esto podría afectar a la muestra de tos. Por ello se ha incluido este particular en la encuesta, dado que es de fundamental importancia para diferenciarlas de otras toses proporcionadas por personas que no cumplen con este perfil, pero que tienen un diagnóstico positivo a COVID-19.

Preguntas 5 y 6: las preguntas finales están destinadas a recabar información sobre la edad y el sexo de la persona participante.

Si la persona participante ha seleccionado en la pregunta 1 que no tiene COVID-19 solo se activarán las preguntas 4, 5 y 6 para recolectar información sobre la muestra de la señal acústica de la tos.

Las primeras tres preguntas están orientadas a definir el tipo de sintomatología a la que corresponde la señal acústica de la tos emitida y las siguientes preguntas buscan establecer la etapa de contagio con la que corresponde la señal acústica de tos enviada, entendiendo que la señal acústica de la tos puede variar con la evolución del contagio en cada persona. La lista de preguntas en la página web se muestra en la figura 9.

Figura 9
Preguntas en la página web



Caracterización de la muestra inicial de la base de datos

En concordancia con lo expuesto anteriormente, en este apartado se presenta la estructura de las bases de datos disponibles y la caracterización de los participantes que hasta la fecha de realización del presente informe han ingresado su información a través del sitio web del proyecto.

Bases de datos

Las bases de datos contienen la información recopilada a través del sitio web (https://databasecovid19.ups.edu.ec/). Actualmente se cuenta con dos bases de datos en formato de valores separados por comas, estructuradas en siete campos en los que se almacena la información declarada por los participantes en el sitio web mediante las siguientes preguntas:

- ¿Posee COVID-19?
- Verificación del diagnóstico.
- ¿Hace qué tiempo aparecieron los primeros síntomas?
- Edad.
- Sexo.

En los dos campos restantes se registra una ID ordinal para cada participante y la grabación realizada en el sitio web.

La primera base de datos (B1) cuenta con un total de 726 registros con fecha de recolección comprendida entre mayo y agosto de 2020. Por otra parte, la segunda base de datos (B2), cuenta con 197 registros con fecha de recolección comprendida entre septiembre de 2020 y febrero de 2021. La única diferencia en la estructura de estas bases de datos se encuentra en las categorías incluidas en el campo que describe la verificación del diagnóstico. En el caso de B1, las categorías únicamente contemplan si se verificó o no el diagnóstico mediante un test, mientras que en B2 se contemplan cuatro categorías distintas y es la estructura con la que se trabaja actualmente.

Además, a partir de marzo de 2021, se incluyeron dos campos adicionales: condición de fumador; y condición de asmático.

También se cuenta con la base de datos utilizada por Brown *et al.* (2020) como fuente para el estudio publicado en el congreso KDD'20. Esta base de datos contiene 459 muestras de tos y respira-

ción de 378 participantes recolectadas mediante aplicaciones web y Android hasta mayo de 2020.

Procedimiento

La información disponible en las bases B1 y B2 fue sometida a un proceso de validación de datos, estructurado a su vez en procedimientos de tipificación de datos, evaluación de rangos de datos, consistencia interna de los datos y limpieza de datos. Una vez concluido el proceso de validación se llevó a cabo un proceso de sistematización de la información, a partir de lo cual se construyeron las matrices depuradas y la tabla de resumen de la estructura de la base de datos.

Análisis de datos

A partir de la información depurada se procedió a caracterizar a los participantes del estudio a través de técnicas de visualización de datos. Todos los procedimientos se llevaron a cabo en el software especializado SPSS versión 22.

Resultados

En la tabla 1 se presentan los elementos de las bases de datos B1 y B2, con los campos adicionales que se han añadido hasta marzo de 2021.

Tabla 1

Estructura de la base de datos

Variable	Descripción	Tipo	Rango	Categorías	Validado por tipo	Validado en blanco
Id	Número de identificación asignado al participante	Nominal	N/A	N/A	Sí	Sí
Diagnóstico	Respuesta a la pregunta ¿Posee COVID-19?	Categórica	0-1	0. Negativo 1. Positivo	Sí	Sí

Verificación	Verificación del diagnóstico y sintomatología asociada al COVID-19	Categórica	0-2	0. Sintomático y diagnosticado por test 1. Sintomático sin diagnóstico 2. Asintomático	Sí	Sí
Temporalidad	Número de días desde que aparecieron los síntomas o se realizó la prueba.	Categórica	0-3	0. 1-3 días 1. 4-8 días 2. 9-15 días 3. Mayor a 15 días	Sí	Sí
Fumador	Condición de fumador del participante	Categórica	0-1	0. No 1. Sí	Sí	Sí
Asmático	Condición de asmático del participante	Categórica	0-1	0. No 1. Sí	Sí	Sí
Edad	Edad en años del participante	Continuo	0-71	N/A	No	No
Sexo	Sexo del participante	Categórica	0-1	0. Mujer 1. Hombre	Sí	Sí
Grabación	Grabación realizada por el participante	Continua	N/A	N/A	Sí	No

Se observa que la variable "Edad" no se encuentra validada por tipo, es decir, que recibe cualquier tipo de valores ingresado por los participantes. No se valida la variable si está en blanco, lo que implica que el participante puede cargar su información en la página web sin haber completado el campo correspondiente a esta variable.

Por otra parte se observa que la variable "Grabación" tampoco se encuentra validada en blanco, es decir, los participantes pueden enviar su información sin haber realizado la grabación.

Estas características de validación han ocasionado que se tenga un alto porcentaje de registros en blanco para la variable "Grabación",

como se muestra en la tabla 2. Sin embargo, esto no supone un problema significativo en el estudio ya que se han previsto acciones para depurar los datos y trabajar únicamente con información completa.

Tabla 2

Resultados de la depuración de la base de datos

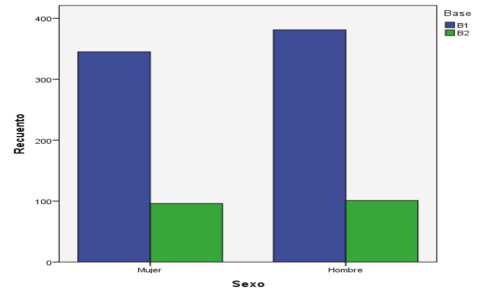
Variable	Base	Registros válidos	Registros no válidos	Total	Porcentaje de registros completos	Porcentaje de registros en blanco
Edad	B1	646	80	726	89 %	11 %
	B2	162	35	197	82 %	18 %
Grabación	B1	292	434	726	40 %	60 %
	B2	71	126	197	36 %	64 %

En lo que respecta a la base de datos de Brown *et al.* (2020), no se dispone de información de los participantes, pero se cuenta con grabaciones que cuentan con una garantía de calidad de datos.

En cuanto a los participantes del estudio, se observa en la figura 10 que en las dos bases existe una proporción similar entre hombres y mujeres, en general, el 48 % de participantes son mujeres y 52 % restante, hombres.

Figura 10

Participantes segmentados por género

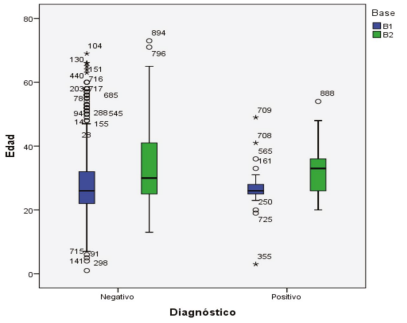


Se observa en la figura 11 que la edad de los participantes del estudio presenta una dispersión relativamente alta, sobre todo en los casos de

la base B1, en la que se observa una gran presencia de valores atípicos. En general, aquellos participantes con diagnóstico positivo en la base B1 tienen una edad media de 27 años, mientras que en la base B2 el valor correspondiente es de 33 años. Por otro lado, para los casos negativos el valor medio de la edad de los participantes en la base B1 es de 29 años, mientras que el valor correspondiente para los participantes de la base B2 es de 33 años.

Figura 11

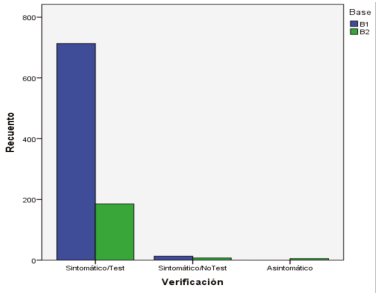
Distribución de la edad de los participantes del estudio en función del diagnóstico



En la figura 12 se observa que un equivalente al 95 % de los participantes han declarado estar diagnosticados con COVID-19, han presentado síntomas y han recibido la confirmación del resultado mediante un test. Mientras que el 5 % restante no ha presentado síntomas o habiendo presentado síntomas no han recibido ninguna confirmación del diagnóstico.

Figura 12

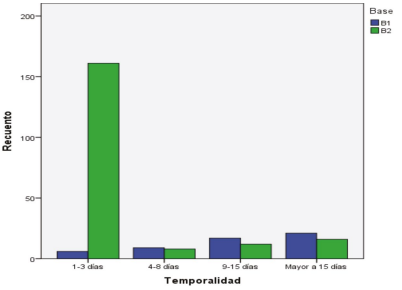
Medios de verificación del diagnóstico para cada base de datos



En la figura 13 se observa que en el caso de los participantes de la base B1 cerca del 40 % presentó síntomas después de 15 días, lo contrario sucede en los participantes de la base B2, en la cual el 81,7 % presentó síntomas entre el primer y tercer día.

Figura 13

Tiempo de aparición de los síntomas



Estrategias y metodologías utilizadas para recopilar la muestra inicial de la base de datos a procesar

Para la captación de las señales de tos, se ha implementado una aplicación web, donde los usuarios pueden donar la tos, guardándose así los registros en una base de datos, para posteriormente tener patrones clasificados en función de las características de una tos con COVID-19 y de otras que no estén asociadas a la enfermedad, lo que generará mayor información en cuanto a las investigaciones de cómo los patrones respiratorios analizados automáticamente podrían usarse como señales de preselección para ayudar al diagnóstico de COVID-19 (Brown, 2021).

Así, el objetivo es generar estrategias o métodos que ayuden a incrementar las muestras de tos dentro de la aplicación web, ya que se necesita la mayor cantidad de datos posibles para alimentar al modelo de Inteligencia Artificial (IA) y una muestra adicional con datos certificados de personas o usuarios que presenten diagnóstico positivo de COVID-19 para validar el modelo.

Participantes

Se considerarán como participantes del estudio a cualquier persona que haya ingresado al sitio web del proyecto y que haya aceptado participar en el estudio.

Mediciones

La publicidad para lograr obtener la mayor cantidad de muestras posibles se realizó con una campaña de difusión del uso de la plataforma, a través de la red social Facebook, entre las cuales se obtuvo gran aceptación por parte de los usuarios, teniendo inicialmente un aproximado de 300 reproducciones.

Procedimiento

En las reuniones de trabajo del equipo de investigación se establecieron las estrategias para la obtención y recopilación de las muestras iniciales, las cuales se enuncian a continuación:

- Desarrollo de un flyer y un poster que fueron compartidos en redes sociales con el objetivo de llegar a un mayor número de personas e incrementar la posibilidad de contar con una mayor cantidad de muestras de tos.
- Elaboración de un protocolo para la obtención de las muestras, en dónde se detallan las acciones a seguir por parte de los participantes, y las instrucciones que deben seguirse para cumplir a cabalidad la donación de la señal de tos, de manera eficaz y eficiente.

Análisis de datos

- Registro de muestras.
- Mayo-agosto 2020: 726.

- Agosto-octubre 2020: 47.
- Noviembre-diciembre 2020: N/A.
- Enero-febrero 2021: 156.

Resultados y discusión

Se han realizado una serie de estrategias y métodos para recopilar muestras de tos, con la ayuda de la campaña de difusión a través de la red social Facebook y así se pudo obtener datos de más participantes o usuarios. En particular, a partir del mes de enero de 2021, hasta finales de febrero, el conjunto de datos se incrementó a 929 registros, sin embargo, se deben analizar las grabaciones para confirmar que sean muestras validadas para el estudio.

Además, para incrementar el número de muestras se requiere que la estrategia de obtener los datos a través de la aplicación web sea difundida lo más pronto, y así recopilar más datos hasta la etapa de análisis y la caracterización propia de las señales de tos.

Descripción de las bases de datos de la plataforma

En la fase I del proyecto se llevó a cabo la caracterización de la tos provocada por la COVID-19, utilizando IA y de procesamiento de señales. Dicha caracterización de la tos fue propuesta con un enfoque sistémico, en el cual integraron tres componentes: la separación de señales de tos y de ruidos externos, el filtrado de la señal y el análisis en el tiempo y en la frecuencia de las señales de tos filtradas.

Para la captación de las señales de tos, mediante la aplicación web descrita en los apartados anteriores, donde los usuarios han donado la tos, se han almacenado los registros en la base de datos previamente descrita. Posteriormente se determinaron patrones y fueron clasificados en función de las características de una tos con COVID-19 y de otras que no están asociadas a la enfermedad, lo que generó más información en cuanto a cómo los patrones respiratorios analizados

automáticamente podrían usarse como señales de preselección para ayudar al diagnóstico de COVID-19 (Brown, 2021).

Metodología

a) Participantes

Se consideraron como participantes del estudio los casos siguientes:

- Cualquier persona que haya ingresado al sitio web del proyecto y que haya aceptado participar en el estudio.
- Cualquier persona referida por miembros del equipo debido a su condición de diagnóstico positivo de COVID-19.

b) Procedimiento

Para la obtención de la base de datos total se realizaron las estrategias mencionadas en las actividades anteriores como:

- La elaboración y difusión de un flyer, que se muestra a continuación en la Figura 2.13 con la información del proyecto para la obtención de las señales de tos a través de la red social Facebook.
- Elaboración de un protocolo para la obtención de las muestras, en dónde se detallan las acciones a seguir por parte de los participantes, y las instrucciones que el mismo debe realizar para cumplir a cabalidad la donación de la señal de tos, de manera eficaz y eficiente.

Una vez que se recopilaron los datos dentro de la aplicación web, se procedió a realizar la limpieza de los metadatos de la información generada por los participantes que donaron las señales de su tos.

Finalmente se procedió a realizar un análisis descriptivo con el fin de determinar y validar las respuestas dadas a los atributos de la estructura de la aplicación web.

Figura 14

Flyer promocional del proyecto para la captación de señales de tos

c) Análisis de datos

En la tabla 3 se muestra el seguimiento de los datos recopilados dentro de la aplicación web desarrollada dentro del proyecto.

Tabla 3

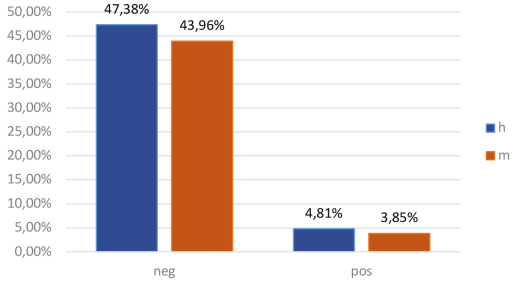
Seguimiento de la muestra total de datos-aplicación web

Registro de muestras		Observaciones
Mayo-agosto 2020	726	Categorías contemplan: únicamente si se verificó el diagnóstico mediante una prueba.
Agosto 2020-febrero 2021	197	Se contemplan cuatro categorías distintas.
Marzo 2021- mayo 2021	12	Se incluyeron dos campos adicionales: condición de fumador y condición de asmático.
Total	935	

En la figura 15 se muestra el análisis del diagnóstico de la enfermedad de COVID-19, en relación con el género de los participantes:

Figura 15

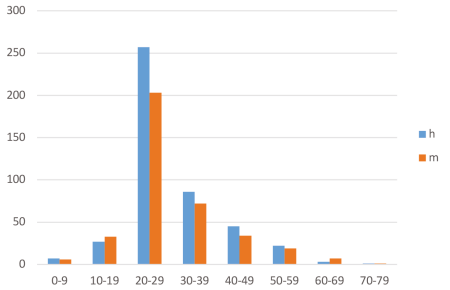
Clasificación del diagnóstico en relación con el género



En la figura 16 se muestran los datos recopilados en función de los rangos de edad de los participantes.

Figura 16

Distribución de datos de género por rangos de edad

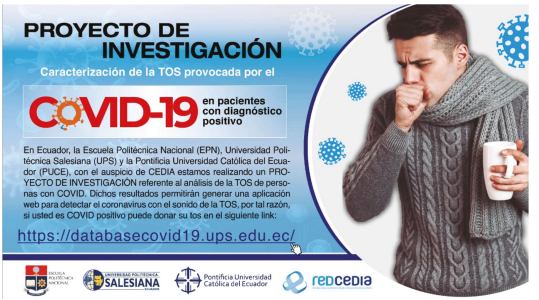


Resultados y discusión

Se han recopilado muestras de tos dentro de la aplicación web con la ayuda de una campaña mediática a través de las redes sociales: así se obtuvieron datos de las señales de tos de los participantes de forma voluntaria y anónima que aceptaron ser parte de la investigación. En particular, a partir del mes de enero de 2021, hasta finales de mayo de 2021, el conjunto de datos se incrementó a 935 registros, sin embargo,

se deben analizar las grabaciones para confirmar que sean muestras validadas para el estudio. El afiche promocional se muestra en la figura 17:

Figura 17
Poster de difusión para recolección de señales de tos



Del análisis de la base total de datos se obtuvo un registro de 823 muestras, donde los participantes colocaron sus edades. Sin embargo, el número de datos con edades registradas mayores de 20 años fue de 750 datos válidos. Obteniendo el mayor registro para el rango de edad entre 20 a 29 años, de 460 muestras, del cual 257 fueron hombres y 203 fueron mujeres.

Además, de la clasificación porcentual en relación con el diagnóstico de la enfermedad COVID-19, se obtuvieron mayores porcentajes con resultados negativos, con valores de 47,38 % de hombres y el 43,96 % de mujeres. Y existieron porcentajes bajos para muestras de tos con diagnóstico positivo, el cual fue el 4,81% para hombres y el 3,85 % para mujeres.

Descripción de las bases de datos totales

Bases de datos

a) Base de datos del proyecto

Se ha denominado *base de datos del proyecto* a aquella conformada por la información recopilada a través del sitio web (https://databasecovid19.ups.edu.ec/). En la base de datos del proyecto se recoge la información de las personas que decidieron participar en el estudio en ocho variables, tal como se presenta en la tabla 4.

A mediados del mes de octubre de 2021, que fue la fecha planificada para cerrar el proceso de toma de datos por medio de la página web, se contó con un total de 935 registros, de los cuales 368 tienen una grabación. Estas grabaciones fueron posteriormente analizadas por el sistema de detección de tos/no tos para que únicamente las señales de tos pasen por el proceso de filtrado y normalizado para su posterior caracterización. En la tabla 4 se observan los elementos que tienen las bases de datos del proyecto.

Tabla 4

Elementos de la base de datos del proyecto

Variable	Descripción	Tipo	Rango	Categorías
Id	Número de identificación asignado al participante	Nominal	N/A	N/A
Diagnóstico	Respuesta a la pregunta ¿Posee COVID-19?	Categórica	0-1	0. Negativo
				1. Positivo
Verificación	Verificación del diagnóstico y sintomatología asociada al COVID-19	Categórica	0-2	0. Sintomático y diagnosticado por test
				1. Sintomático sin diagnóstico
				2. Asintomático
Temporalidad	Número de días desde que aparecieron los síntomas o se realizó la prueba.	Categórica	0-3	0. 1-3 días
				1. 4-8 días
				2. 9-15 días
				3. Mayor a 15 días
Fumador	Condición de fumador del participante	Categórica	0-1	0. No
				1. Sí
Asmático	Condición de asmático del participante	Categórica	0-1	0. No
				1. Sí
Edad	Edad en años del participante	Continuo	0-71	N/A
Sexo	Sexo del participante	Categórica	0-1	0. Mujer
				1. Hombre
Grabación	Grabación realizada por el participante	Continua	N/A	N/A

b) Base de datos Cambridge

Se ha denominado *base de datos de Cambridge* a la base de datos utilizada por Brown (2020) para el estudio publicado en el congreso KDD'20. En esta base de datos no se dispone de información de los participantes, pero se cuenta con grabaciones con una garantía de calidad de datos.

En la *base de datos de Cambridge* se cuenta con 459 muestras de tos y respiración de 378 participantes recolectadas mediante aplicaciones web y Android hasta mayo de 2020. En la tabla 5 se presenta una descripción del tipo de grabaciones de tos que contiene esta base de datos.

Tabla 5

Características de las grabaciones de tos de la base de datos de Cambridge

Diagnóstico COVID-19	Tipo de tos	Medio de recolección	Total (de grabaciones)
Positivo	No sintomática	Aplicativo Android	64
Positivo	Sintomática	Aplicativo Android	46
Positivo	No sintomática	Página web	23
Positivo	Sintomática	Página web	8
Negativo	No sintomática	Aplicativo Android	137
Negativo	Sintomática	Aplicativo Android	8
Negativo	No sintomática	Página web	161
Negativo	Sintomática	Página web	24
Negativo	Asmática	Aplicativo Android	13
Negativo	Asmática	Página web	7

En esta base de datos se cuenta también con muestras artificiales producidas por la modificación de la frecuencia y la inyección aleatoria de ruido mediante la técnica *data augmentation*. De igual manera se cuenta con grabaciones de respiraciones de los participantes.

Las grabaciones de esta base de datos se están utilizando como referencia para entrenar al sistema de detección de tos/no, ya que, en este caso, no es necesario diferenciar si la tos proviene de un participante con COVID-19 o con otra patología pues lo que interesa en dicho sistema es discriminar a una señal de tos de cualquier otro evento sonoro.

c) Bases de datos de ruido ambiental

Se ha denominado *base de datos de ruido ambiental* a la base de datos utilizada por Piczak (2015) para desarrollar un estudio de clasificación de ruido ambiental. La *base de datos de ruido ambiental* contiene 2000 grabaciones ambientales etiquetadas y balanceadas en 50 clases (40 clips por clase). Por conveniencia se agrupan en cinco categorías principales (diez clases por categoría), en la tabla 6 se presenta una descripción completa de las categorías mencionadas:

- Animales.
- Paisajes sonoros y sonidos de agua.
- Sonidos humanos diferentes al habla.
- Sonidos interiores/domésticos.
- Sonidos exteriores/urbanos.

Tabla 6
Categorías de la base de datos de ruido ambiental

Animales	Paisajes sonoros y sonidos de agua	Sonidos humanos diferentes al habla	Sonidos interiores/ domésticos	Sonidos exteriores/ urbanos
Perro	Lluvia	Llanto de bebé	Aldaba	Helicóptero
Gallo	Olas de mar	Estornudo	Clic de ratón	Motosierra
Cerdo	Fuego crepitante	Aplauso	Tecleo	Sirena
Vaca	Grillos	Respiración	Madera que cruje	Bocina de carro
Rana	Trino de pájaros	Tos	Abrelatas	Motor
Gato	Gotas de agua	Pasos	Lavadora	Tren
Gallina	Viento	Risa	Aspiradora	Campanas
Insectos (volando)	Agua vertiéndose	Cepillado de dientes	Alarma	Avión
Oveja	Inodoro	Ronquido	Reloj	Fuegos artificiales
Cuervo	Tormenta	Beber, sorber	Vidrio rompiéndose	Sierra de mano

Aunque lo ideal sería que las grabaciones de los sonidos expuestos se mantuvieran en primer plano con un ruido de fondo limitado, en realidad, las grabaciones de campo están lejos de ser "puras", y algunas, inclusive, muestran una superposición auditiva en el fondo. No obstante, el conjunto de datos proporciona una exposición a una variedad de fuentes de sonido: algunas muy comunes (risas, maullidos de gatos, ladridos de perros, etc.), otras bastante distintas (rotura de cristales, cepillarse los dientes, etc.) y otras en las que las diferencias son más matizadas (ruido de helicópteros y aviones). Una de las posibles deficiencias de este conjunto de datos es el número limitado de clips disponibles por clase. Esto está relacionado con lo demandante que es el proceso de etiquetado y extracción manual, y la decisión de mantener un estricto equilibrio entre clases a pesar de la disponibilidad limitada de grabaciones para tipos más exóticos de eventos sonoros. Aun así, la calidad de las grabaciones ha hecho posible que se puedan

utilizar como clases de otros eventos sonoros para el sistema de detección tos/no tos, siempre y cuando se eliminen las 50 grabaciones que tienen señales de tos.

Con todo lo dicho se puede ver que la arquitectura cliente-servidor utilizada en este trabajo es flexible y adaptable al servicio implementado, lo que permite aumentar el rendimiento. Así mismo, la arquitectura cliente-servidor puede incorporar variadas plataformas, bases de datos, redes y sistemas operativos que pueden ser de diferentes fabricantes, en arquitecturas propietarias y no propietarias.

Referencias bibliográficas

Abu-Naser, M. W. A. a. S. S. (2017). CSS-Tutor: An intelligent tutoring system for CSS and HTML.

Alhawiti, K. M. (2015). Advances in artificial intelligence using speech recognition. *World Academy of Science, Engineering and Technology, International Journal of Computer, Electrical, Automation, Control and Information Engineering, 9*(6), 1351-1354.

Brown, C., Chauhan, J., Grammenos, A., Han, J., Hasthanasombat, A., Spathis, D. Xia, T., Cicuta, P. y Mascolo, C. (2020). *Exploring automatic diagnosis of COVID-19 from crowdsourced respiratory sound data.* arXiv preprint arXiv:2006.05919.

Deng, L. y Li, X. (2013). Machine learning paradigms for speech recognition: An overview. *IEEE Transactions on Audio, Speech, and Language Processing, 21*(5), 1060-1089.

Deeks, J., Dinnes, J., Takwoingi, Y., Davenport, C., Spijker, R., Taylor-Phillips, S., Adriano, A., Beese, S., Dretzke, J., Ruffano, L. F. di, Harris, I. M., Price, M. J., Dittrich, S., Emperador, D., Hooft, L., Leeflang, M. M., Bruel, A. V. den y Group, C. C.-19 D. T. A. (2020). Antibody tests for identification of current and past infection with SARS-CoV-2. *Cochrane Database of Systematic Reviews, 6.* https://doi.org/10.1002/14651858.CD013652

Duckett, J. (2011). *HTML & CSS: design and build websites*. Wiley Indianapolis. https://bit.ly/43oBYTG

Dyer, R. J. T. (2015). *Learning MySQL and MariaDB: Heading in the right direction with MySQL and MariaDB.* O'Reilly Media, Inc.

Frain, B. (2015). *Responsive web design with HTML5 and CSS3.* Packt Publishing Ltd.

He, R. Y. (2015). *Design and implementation of web based on Laravel framework.* International Conference on Computer Science and Electronic Technology (ICCSET 2014), pp. 301-304.

Peterson, C. (2014). *Learning responsive web design: a beginner's guide.* O'Reilly Media, Inc..

Piczak, K. J. (2015). *ESC: Dataset for environmental sound classification.* MM 2015- Proceedings of the 2015 ACM Multimedia Conference, pp. 1015-1018. https://doi.org/10.1145/2733373.2806390

Prettyman, A. W. W. a. S. (2018). *Practical PHP 7, MySQL 8, and MariaDB Website Databases.* Springer.

Robbins, J. N. (2012). *Learning web design: A beginner's guide to HTML, CSS, JavaScript, and web graphics*. O'Reilly Media, Inc.

Sklar, J. (2011). *Principles of web design: the web technologies series.* Cengage Learning.

Capítulo 3

Reconocimiento de tos no-tos en una señal audible

Introducción

En el presente capítulo se desarrolla un marco teórico respecto al reconocimiento de las señales de tos, luego se procede a señalar los principales componentes y procedimientos que se usaron para el reconocimiento de la tos en una señal de audio con base a una muestra inicial, además de separar las señales de tos de los ruidos externos por medio de un sistema de reconocimiento de patrones usando el modelo Redes Neuronales Convolucionales (CNN-Convolutional Neural Networks).

Finalmente se presentan los resultados obtenidos de la implementación del modelo aplicado con una muestra inicial y luego con aumento de datos, para el procesamiento de las señales de tos.

Procesamiento de señales de tos

Tradicionalmente, la detección de la tos se ha basado en la auscultación con el uso de un estetoscopio. La responsabilidad de este procedimiento para diagnosticar de manera efectiva diversas enfermedades

respiratorias recae principalmente en el personal médico, por lo que el diagnóstico correcto dependerá de lo bien capacitado que esté el profesional. Además, este procedimiento requiere principalmente que el paciente acuda a un hospital o programe una visita médica domiciliaria, lo que no siempre permite recopilar información cuando los episodios de tos ocurren en tiempo real (Spinou y Birring, 2014).

Se ha fomentado, por otro lado, la aparición gradual de sistemas informáticos para procesar la tos debido al desarrollo de las Tecnologías de la Información y Comunicación (TIC). Estos sistemas funcionan mediante el reconocimiento de patrones basados principalmente en características extraídas de los sonidos de la tos y señales complementarias, como, por ejemplo, en la electromiografía del movimiento del pecho.

Sin embargo, la mayoría de estos sistemas solo se han probado en entornos controlados donde los pacientes no participaron en ninguna acción o actividad física y no se les instó a utilizar sistemas de grabación complejos. Por tanto, una limitación de los métodos de procesamiento de audio implementados en estos sistemas (de monitoreo) es que no pueden hacer frente a entornos ruidosos (Spinou y Birring, 2014; Domingo y Sogo Sagardía, 2016).

Por el contrario, un sistema (llamado monitor) de tos portátil permitiría un seguimiento continuo en tiempo real de pacientes con enfermedades respiratorias. Esta forma de seguimiento de la tos podría permitir captar información fiable para los profesionales, que a la vez fortalecería la telemedicina en enfermedades respiratorias. Además, desde el punto de vista del paciente, este sistema de monitorización sería más cómodo y provocaría una mínima interrupción en sus actividades diarias. Sin embargo, el desarrollo de monitores portátiles implica la implementación de sistemas que puedan identificar y discriminar la tos de cualquier otro evento sonoro registrado en entornos potencialmente ruidosos.

El problema de diferenciar la tos de otros eventos sonoros se ha abordado desde diferentes perspectivas. Así, por ejemplo, en el campo

del aprendizaje automático, Khunarsal *et al.* (2013) implementaron un algoritmo para la clasificación del sonido ambiental utilizando espectrogramas y clasificadores K-Vecinos-más cercanos (KNN- K-Nearest Neighbors). Mitra y Wang (2008) propusieron una técnica de clasificación de audio basada en el análisis de contenido de audio utilizando perceptrones multicapa. Además, Costa *et al*. (2012) abordaron un problema similar utilizando máquinas de soporte vectorial.

Por otro lado, Matos *et al.* (2006) implementaron modelos ocultos de Markov para detectar señales de tos de grabaciones de audio continuas. Sin embargo, los resultados más prometedores se han logrado mediante la implementación de algoritmos de aprendizaje profundo, especialmente utilizando Redes Neuronales Convolucionales, como en el trabajo realizado por Amoh y Odame (2016) quienes implementaron efectivamente una CNN para identificar los sonidos de la tos.

Cuando se trabaja con CNN, los datos de entrada deben ser imágenes. En clasificación de sonidos, autores como Wang *et al.* (2015) o Barata *et al.* (2019) han sugerido técnicas como Transformada Espectral Relativa combinada con Predicción Lineal Perceptual o Coeficientes Cepstrales de Frecuencia Mel (MFCC- Mel Frequency Cepstral Coeffcients) para transformar el sonido en espectrogramas. De esta forma, la tarea de reconocimiento de audio se convierte en una tarea de reconocimiento visual. Aplicando estas técnicas han modelado CNN con exactitudes de clasificación superiores al 90 %, que se han implementado con éxito en monitores de tos portátiles.

Las CNN se utilizan para procesar imágenes que simulan el comportamiento de las neuronas en la corteza visual principal de un cerebro biológico. Pueden aprender las relaciones de entrada y salida y se basan en operaciones de convolución. Por lo tanto, convencionalmente, una CNN se compone de capas de convolución, capas agrupadas y capas completamente conectadas (Aloysius, 2017; Rawat y Wang, 2017).

Esta estructura involucra una gran cantidad de parámetros internos que se ajustan durante el aprendizaje. Esta gran cantidad de parámetros hace que las CNN sean propensas al sobreajuste, que ocurre cuando el modelo aprende patrones subyacentes al conjunto de datos de entrenamiento, pero no puede generalizar para reconocer nuevos datos y, por lo tanto, no realiza una clasificación eficiente (Aloysius y Geetha 2017; Amoh y Odame, 2016; Rawat y Wang, 2017).

Para evitar este problema se aplican técnicas como el *dropout*, que consiste en desconectar un grupo de neuronas en cada iteración de entrenamiento (Srivastava *et al.,* 2014). Sin embargo, una de las formas más eficientes de lidiar con el sobreajuste es aumentar el número de muestras en el conjunto de entrenamiento tanto como sea posible.

Procesamiento de señales de tos con CNN

La metodología de reconocimiento de una señal de tos que se aplica para la implementación del modelo CNN requiere de varios procesos y componentes que se describen a continuación.

Base de datos

La base de datos utilizada consta de 266 grabaciones de toses y 69 grabaciones de otros eventos sonoros como conversaciones, risas o ruido ambiental. Las grabaciones están en formato WAV y duran entre 1 y 112 segundos.

Se realizaron las grabaciones de forma telemática y anónima tras aceptar el consentimiento informado.

Arquitectura del sistema

La arquitectura del sistema consta de dos componentes: *Front-End* y *Back-End*, en el componente Front-End, la señal de audio se preprocesa y la detección de la tos se transforma en una tarea de reconocimiento visual mediante la generación de los correspondientes espectrogramas

de Mel. En este contexto, en primer lugar, la duración de las grabaciones se estandarizó al valor modal de 3 segundos. Así, las grabaciones de mayor duración se cortaron en segmentos de 3 segundos, mientras que las grabaciones de menor duración se replicaron y posteriormente concatenaron hasta llegar a los 3 segundos. Las muestras resultantes se volvieron a analizar para descartar todas aquellas con intervalos de silencio. Este procedimiento también permitió equilibrar los datos con una relación entre los sonidos de tos y otros eventos sonoros de alrededor de 1,08 segundos.

Posteriormente, las muestras de audio procesadas se transformaron en espectrogramas Mel (MFCC). Se eligió esta representación gráfica ya que ha sido ampliamente utilizada en los campos del reconocimiento de voz y la identificación de altavoces o eventos sonoros ambientales (García y Destéfanis, 2019). Los espectrogramas de Mel se generaron con un tamaño de ventana de 30 ms y una frecuencia de muestreo de 48000 Hz.

Por otro lado, en el componente Back-End, los espectrogramas de Mel se analizan a través de una Red Neuronal Convolucional (CNN) que permite determinar si una señal de audio proviene de un sonido de tos u otro evento sonoro. La arquitectura de Red Neuronal Convolucional utilizada en este estudio se basa en la arquitectura LeNet-5, que funciona bien con pequeños conjuntos de datos. Además, la cantidad de neuronas en cada capa se adaptó de acuerdo con Amoh y Odame (2016) que trabajaron con un conjunto de datos de tamaño reducido.

De manera análoga a LeNet-5 (Amoh y Odame, 2016; Lecun *et al.*, 1998), la CNN consta de cinco capas, divididas en dos capas convolucionales, dos capas completamente conectadas y una capa de clasificación binaria. Las dos capas convolucionales tienen 16 unidades lineales rectificadas (ReLU) cada una. La primera capa convolucional tiene un filtro de tamaño 9×3 y toma los segmentos espectrales de Mel de 67×300 como entradas. A continuación, se implementa una capa 2×1 de max-pooling, seguida de la segunda capa convolucional con un tamaño de filtro de 5x3 y otra capa de 2x1 de *max-pooling*. Dos

capas completamente conectadas siguen las capas convolucionales con 256 unidades para cada una.

En cada capa, además, se implementó la regularización batch y 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.

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. También 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.

Medidas de desempeño

La evaluación del desempeño del modelo se basó en las siguientes métricas:

- **Pérdida de entropía cruzada binaria:** describe la pérdida entre dos distribuciones de probabilidad, utilizando una penalización logarítmica que genera una puntuación grande para diferencias grandes y una puntuación pequeña para diferencias pequeñas. La pérdida de entropía cruzada binaria se utiliza al ajustar los pesos del modelo durante el entrenamiento. El objetivo es minimizar la pérdida; es decir, a menor pérdida, mejor modelo (Ramos *et al.,* 2018).
- **Exactitud:** tasa de éxito general del clasificador o la fracción de predicciones que el modelo hizo correctamente (Deng *et al.,* 2016; Juba y Le, 2019).
- **Precisión:** mide la relación entre los verdaderos positivos y las instancias clasificadas como positivas (Juba y Le, 2019).

- **Recall:** mide la relación de los casos positivos entre los que han sido correctamente clasificados (Juba y Le, 2019).
- **F1:** media armónica entre precisión y *recall*, también facilita la comparación del desempeño combinado de precisión y *recall* entre varias soluciones (Juba y Le, 2019; (Chicco y Jurman, 2020).

Resultados

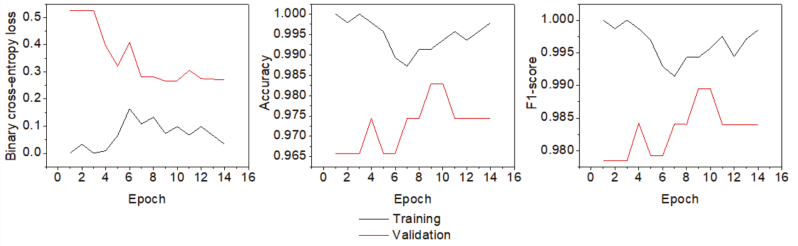
A continuación se presentan los resultados del reconocimiento de la tos en una señal de audio con base en una muestra inicial.

En la figura 1 se muestran los gráficos de las métricas de rendimiento del modelo entrenado. Se observa que la pérdida de entropía cruzada binaria en el momento de la detención temprana fue menor para el conjunto de datos de entrenamiento (training) que para el conjunto de datos de validación (*validation*).

Por otro lado, se observa que tanto la exactitud como el valor de F1 del conjunto de datos de entrenamiento fueron mayores en comparación con los valores correspondientes al conjunto de datos de validación.

Figura 1

Evolución de la pérdida de entropía cruzada binaria, exactitud y valor de F1 del modelo entrenado



La tabla 1 muestra los valores a los que convergieron las métricas de rendimiento hasta el momento de detención temprana.

Tabla 1
Valores de convergencia de la pérdida de entropía cruzada binaria, exactitud y valor de f1 del modelo entrenado

Medida de desempeño	Valor de convergencia en el conjunto de validación
Pérdida de entropía cruzada binaria	0,3955
Exactitud	0,9744
F1	0,9837

Procesamiento de señales de tos con CNN y aumento de datos

3.4.1. Técnica de aumento de muestras (datos)

En el subcapítulo (*Procesamiento de señales de tos con CNN*) se trató sobre el reconocimiento de tos no-tos en una señal de audio de una muestra inicial, se determinó las ventajas de abordar el reconocimiento de audio mediante el reconocimiento visual de los espectrogramas correspondientes a la señal de la tos. Dicha aproximación se la realizó mediante el modelamiento de una Red Neuronal Convolucional y se había establecido que este tipo de modelo suele presentar problemas de sobreajuste, para lo cual se aplican técnicas como el *dropout*, que consiste en desconectar un grupo de neuronas en cada iteración de entrenamiento (Srivastava *et al.,* 2014). Sin embargo, una de las formas más eficientes de lidiar con el sobreajuste es aumentar el número de muestras en el conjunto de entrenamiento tanto como sea posible.

Sin embargo, los datos utilizados para identificar la tos son relativamente difíciles de obtener porque implican registrar una acción refleja, es decir, un evento no planificado. Por lo tanto, las bases de datos disponibles a menudo no son lo suficientemente grandes para entrenar redes neuronales profundas (Shorten y Khoshgoftaar, 2019).

Una técnica utilizada para afrontar este problema es el aumento de datos, que consiste en crear nuevos datos mediante transformaciones sobre los datos originales. Cuando se usa CNN, estas transformaciones de aumento de datos se realizan en las imágenes de entrada. Se pueden lograr transformaciones de rotación, traslación, escalado, recorte, entre otras. De esta forma, al aumentar los datos, es posible evitar que el modelo aprenda patrones irrelevantes, optimizando la resolución de problemas de clasificación como en el caso del modelo AlexNet (Krizhevsky *et al.,* 2012) que implementa dos técnicas de aumento de datos para mejorar su desempeño (García, 2020; Shorten y Khoshgoftaar, 2019).

Aunque se puede realizar el aumento de datos en el conjunto de imágenes, estas transformaciones se pueden realizar directamente en las grabaciones de audio originales en el caso que se trate de identificación de tos.

En consecuencia, a continuación, se describe la metodología de reconocimiento de una señal de tos que se aplica para la implementación de un modelo de CNN para abordar el problema de identificar la tos como una tarea de reconocimiento visual, a la vez que se estudia el efecto sobre el rendimiento del modelo al aplicar técnicas de aumento de datos en las grabaciones de audio originales.

La metodología de reconocimiento de una señal de tos que se aplica para la implementación del modelo CNN y que utiliza el aumento de datos, requiere de varios procesos y componentes que se describen a continuación.

Bases de datos

Al igual que en el subcapítulo (*Base de datos*) se trabajó con las 266 grabaciones de toses y 69 grabaciones de otros eventos sonoros como conversaciones, risas o ruido ambiental. Las grabaciones están en formato WAV y duran entre 1 y 112 segundos.

Aumento de datos

Se definió una función en Python que toma como entrada un archivo de audio, la frecuencia de muestreo y el factor de variación de inyección de ruido o desplazamiento de frecuencia. Como salida, se obtiene una nueva muestra, ya sea con ruido inyectado o con un cambio de frecuencia.

La tasa máxima tanto de inyección de ruido como de desplazamiento de frecuencia se determinó mediante un proceso heurístico basado en las investigaciones de C. Shorten y TM (2019), García (2020) y Cui *et al.* (2015), quienes sugieren que la variación no debe ser mayor al 20 % ya que, de esta manera, el dominio del problema permanece invariante.

En este sentido, el proceso de inyección de ruido agrega valores aleatorios a la amplitud de la muestra de audio original. Asimismo, mediante el proceso de desplazamiento en frecuencia se varió el tono de la muestra de audio original aleatoriamente.

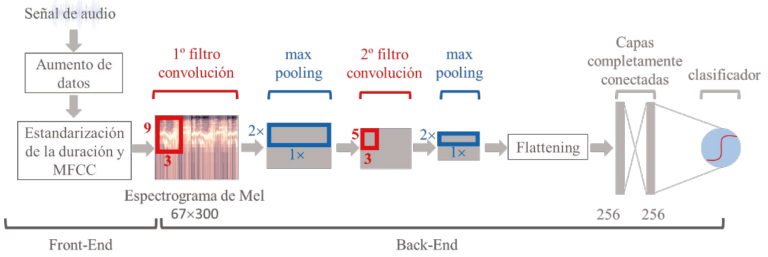
La aplicación de técnicas de aumento de datos hizo posible construir una base de datos cinco veces el tamaño de la base de datos original.

Arquitectura del sistema

Se trabajó con una arquitectura análoga a la del modelo CNN con una muestra inicial, salvo que en este caso se incluyó el proceso de aumento de datos. En resumen se partió de las señales originales de audio, a las cuales se las sometió a procesos de aumento de datos por inyección de ruido y desplazamiento en frecuencia. Después se llevó a cabo la estandarización de la duración de las grabaciones a 3 segundos, lo que a la par permitió balancear las clases del conjunto de datos con una relación entre los sonidos de tos y otros eventos sonoros de alrededor de 1,08. Posteriormente se generaron los espectrogramas de Mel de 67×300, los cuales constituyen la entrada a la primera capa convolu-

cional de la CNN y pasan por un filtro de tamaño 9×3, continuando a través de una capa 2×1 de *max-pooling*, seguida de la segunda capa convolucional con un tamaño de filtro de 5x3 y otra capa de 2x1 de *max-pooling*. Dos capas completamente conectadas siguen las capas convolucionales con 256 unidades para cada una. Cada capa utilizó regularización batch y regularización de *dropout* con una tasa de 0,5 para reducir las probabilidades de sobreajuste. Finalmente, en la última capa se implementó la función *sigmoide* para determinar si la señal de audio analizada es un sonido de tos. El esquema se presenta en la figura 2.

Figura 2
Arquitectura del Sistema



Se entrenó un primer modelo con el conjunto de datos original y los datos aumentados por desplazamiento en frecuencia, el segundo con el conjunto de datos original y los datos aumentados por inyección de ruido, y el tercero con los datos originales conjunto y los datos aumentados mediante inyección de ruido y desplazamiento de frecuencia.

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 detención temprana con un monitor en la función de pérdida para el conjunto de datos de validación para reducir el sobreajuste. Todos los modelos se desarrollaron en un entorno Python con la biblioteca Keras.

Medidas de desempeño

Similar a lo descrito en el subcapítulo (*Medidas de desempeño*), el desempeño de los modelos se midió a través de la pérdida de entropía cruzada binaria, la exactitud y el valor de F1.

Resultados

Las figuras 3, 4 y 5 muestran los gráficos de las métricas de rendimiento de los modelos entrenados. En todos los modelos, la pérdida de entropía cruzada binaria en el momento de la detención temprana fue menor para el conjunto de datos de entrenamiento (*training*) que para el conjunto de datos de validación (*validation*). Sin embargo, en todos los casos, excepto en el que utilizó el aumento de datos por inyección de ruido, la pérdida de entropía cruzada binaria para los datos de validación convergió a valores menores a 0,5.

Se observa, por otro lado, que tanto la exactitud como el valor de F1 del conjunto de datos de entrenamiento fueron mayores en comparación con los valores correspondientes al conjunto de datos de validación. En todos los casos, la exactitud y el valor de F1 para el conjunto de datos de validación alcanzaron valores superiores a 0,9.

La tabla 2 muestra las diferencias entre las métricas de rendimiento en el conjunto de datos de entrenamiento y las del conjunto de datos de validación en el momento de la detección temprana.

En todos los casos se observa que las diferencias para todas las métricas de rendimiento son menores en el modelo en el que se utilizó el aumento de datos, tanto por desplazamiento en frecuencia, como por inyección de ruido.

Figura 3

Evolución de la pérdida de entropía cruzada binaria, exactitud y valor de F1 del modelo entrenado con datos originales y aumentados por desplazamiento en frecuencia

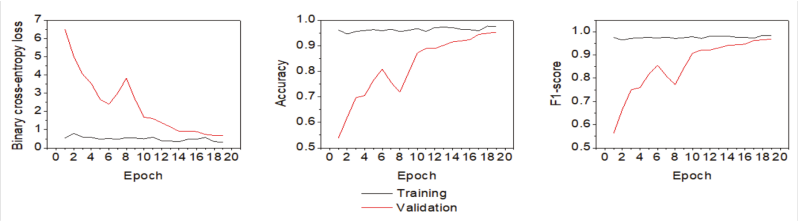


Figura 4

Evolución de la pérdida de entropía cruzada binaria, exactitud y valor de F1 del modelo entrenado con datos originales y aumentados por inyección de ruido

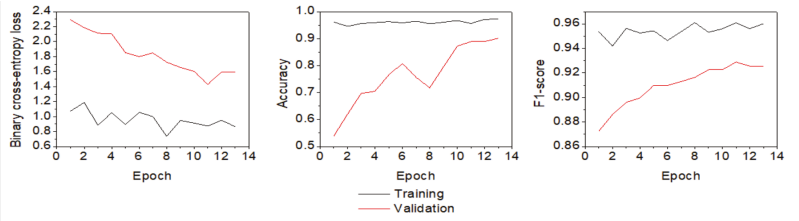
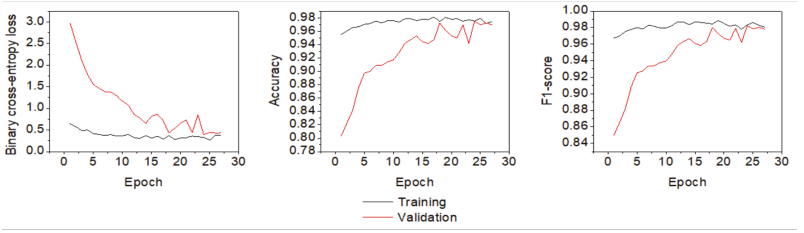


Figura 5

Evolución de la pérdida de entropía cruzada binaria, exactitud y valor de F1 del modelo entrenado con datos originales y aumentados por desplazamiento en frecuencia y por inyección de ruido



Además se observa que cuando se usa el aumento de datos solo por desplazamiento en frecuencia, también se obtienen diferencias más pequeñas entre las métricas de desempeño que cuando se usan solo los datos originales. Por el contrario, cuando el aumento de datos se aplica solo mediante inyección de ruido, se obtienen diferencias más pequeñas entre las métricas de rendimiento del modelo entrenado con los datos originales. Este comportamiento podría sugerir que la inyección de ruido aleatorio podría hacer que los modelos sean susceptibles de sobreajuste al identificar los sonidos de tos.

Tabla 2

Diferencia absoluta de las medidas de rendimiento entre los conjuntos de entrenamiento y validación de los modelos entrenados

Datos de entrenamiento	Pérdida de entropía cruzada binaria	Exactitud	F1
Or	0,27040	0,02560	0,01600
Or+FS	0,34020	0,02030	0,01400
Or+NI	0,58380	0,03720	0,03090
Or+FS+Ni	0,16740	0,01320	0,00920

Nota. Or: Datos originales, FS: datos aumentados por desplazamiento en frecuencia, NI: datos aumentados por inyección de ruido.

Discusión

Varios estudios han demostrado que la técnica de aumento de datos actúa como un regularizador, evitando así el sobreajuste en las redes neuronales al tiempo que permite obtener mejores resultados en situaciones en que las clases estén desequilibradas (Cui *et al*., 2015; García y Destéfanis, 2020; Krizhevsky *et al*., 2012; Shorten y Khoshgoftaar, 2019).

De hecho, cuantos menos datos estén disponibles, menos datos para entrenar y menos posibilidades existen de obtener predicciones precisas sobre datos que el modelo aún no ha visto. Por lo tanto, el aumento de datos adopta el enfoque de generar más datos de entre-

namiento a partir de los datos disponibles. En el caso de este trabajo, aunque la clasificación se abordó como una tarea de reconocimiento visual a partir de espectrogramas Mel, el aumento de datos se realizó en las señales de audio originales, mediante desplazamiento en frecuencia e inyección de ruido.

Como se ve en los resultados, la diferencia entre las métricas de rendimiento del modelo entre los datos de entrenamiento y validación es menor cuando se usa datos aumentados por desplazamiento en frecuencia que cuando se usan los audios originales.

Este resultado puede deberse a que los patrones invariantes que identifican la tos son independientes de la frecuencia. Esto concuerda con el hecho de que, en algunos estudios, se señala que los sonidos que dan un tono característico a la tos tienen su origen en la laringe. Sin embargo, no todas las personas producen estos sonidos con la misma intensidad. Incluso hay personas que no las producen en absoluto por factores genéticos o como consecuencia de intervenciones quirúrgicas (Gibson y Vertigan, 2009; Korpáš *et al.*, 1996). Asimismo, cuando hay un cambio de frecuencia en otros eventos sonoros como risas o conversaciones, aunque los cambios son notables, la naturaleza tímbrica no se altera en mayor medida. Por tanto, el modelo podría percibirlo como una grabación completamente diferente pero perteneciente a la misma clase. En consecuencia, el cambio de frecuencia en las grabaciones, en un factor de no más del 20 %, permitió que el modelo se enfocara en patrones que son independientes del tono de tos.

Por otro lado se observó que, al aplicar el aumento de datos mediante inyección de ruido, la diferencia entre las métricas de rendimiento del modelo entre los conjuntos de entrenamiento y validación es más significativa que cuando se utilizan las grabaciones originales.

En este contexto, según Van Hirtum y Berckmans (2002) y Korpáš *et al.* (1996), los sonidos que podrían caracterizarse como ruido pueden ser indicadores de que una persona fuma, de que padece determinadas enfermedades respiratorias o del nivel de moco en la

tráquea. En consecuencia, la adición aleatoria de ruido podría interferir con los patrones característicos que permiten la identificación de la tos.

Sin embargo, cuando se utilizan ambas técnicas de aumento de datos, la diferencia de métricas de rendimiento del modelo entre los datos de entrenamiento y validación es menor que cuando se utilizan las grabaciones originales. La aplicación de esta técnica permite una mejor generalización de la tarea de clasificación, como se puede apreciar.

Además, al entrenar la CNN con datos aumentados, se logró una precisión de clasificación del 97,50 %. Este resultado es equiparable con los reportados en la literatura, como Amoh *et al.* (2016), Wang *et al.* (2015), o Barata *et al.* (2019), quienes lograron exactitudes de clasificación superiores al 90 % en la identificación de los sonidos de la tos mediante la implementación de CNN. Este hecho corrobora que las redes CNN permiten abordar la tarea de reconocer la tos de manera eficiente (Monge-Alvarez *et al.,* 2019). Sin embargo, la aplicación del aumento de datos aumentó la precisión en el conjunto de validación en menos del 1 % en comparación con las grabaciones originales.

Esta leve mejora podría deberse a que, dado que las grabaciones originales son relativamente cortas y se registraron en diferentes condiciones, las modificaciones aplicadas debían ser muy pequeñas. Por lo tanto es probable que no tengan un efecto significativo en el aprendizaje de modelos. Además, el rendimiento del modelo sin utilizar el aumento de datos es bastante bueno (96,7 % de exactitud), lo que dificulta la tarea de mejorarlo.

En síntesis, se pudo observar que el sistema de reconocimiento de tos-no tos mediante una Red Neuronal Convolucional permite discriminar los sonidos de tos de otros eventos sonoros con una exactitud del 97,44 %, lo cual creó una base sólida para el desarrollo de los siguientes ensayos realizados sobre el sistema que ya se ha modelado. No obstante es necesario recalcar que el entrenamiento de dicho modelo se lo ha realizado sobre los datos originales, es decir, sin procesamiento

previo y con una cantidad de grabaciones relativamente limitada. Por ello, el trabajo futuro se centró, en primer lugar, en solventar la poca cantidad de grabaciones disponibles, también se requirió determinar los requerimientos de pre-procesamiento de las grabaciones para garantizar, en la medida de lo posible, un comportamiento estable del modelo entrenado.

Si bien los resultados del análisis realizado son alentadores, es fundamental enfatizar que las muestras artificiales introducidas a la red neuronal aún están altamente correlacionadas, lo cual es una limitación. Esto se debe a que los datos aumentados provienen de una pequeña cantidad de datos originales, así no se está produciendo información completamente nueva, sino que solo se ha mezclado la información actual.

Referencias bibliográficas

Aloysius, N. y Geetha, M. (2017). *A review on deep convolutional neural networks*. 2017 International Conference on Communication and Signal Processing (ICCSP).

Amoh, J. y Odame, K. (2016). Deep Neural Networks for Identifying Cough Sounds. *IEEE Transactions on Biomedical Circuits and Systems, [online] 10*(5), 1003-1011. https://bit.ly/43rPkPr

Barata, F., Kipfer, K., Weber, M., Tinschert, P., Fleisch, E. y Kowatsch, T. (2019). *Towards Device-Agnostic Mobile Cough Detection with Convolutional Neural Networks*. 2019 IEEE International Conference on Healthcare Informatics (ICHI), 1.

Chicco, D. y Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics, 21*(1).

Costa, Y. M. G., Oliveira, L. S., Koerich, A. L., Gouyon, F. y Martins, J. G. (2012). Music genre classification using LBP textural features. *Signal Processing, [online] 92*(11), 2723-2737. https://bit.ly/3C2K2Og

Deng, X., Liu, Q., Deng, Y. y Mahadevan, S. (2016). An improved method to construct basic probability assignment based on the confusion matrix for classification problem. *Information Sciences, 340-341*, 250-261. https://doi.org/10.1016/j.ins.2016.01.033

García, M. A. y Destéfanis, E. A. (2019). The Power Cepstrum Calculation with Convolutional Neural Networks. *Journal of Computer Science and Technology, 19*(2), p.e13.

García, M. A. y Destéfanis, E.A. (2020). Data Augmentation para la Clasificación Automática de la Calidad Vocal. *AJEA,* (5).

Gibson, P. G. y Vertigan, A. E. (2009). Speech pathology for chronic cough: A new approach. *Pulmonary Pharmacology & Therapeutics, 22*(2), 159-162.

Juba, B. y Le, H. S. (2019). Precision-recall versus accuracy and the role of large data sets. *Proceedings of the AAAI Conference on Artificial Intelligence, 33*, 4039-4048. https://doi.org/10.1609/aaai.v33i01.33014039

Khunarsal, P., Lursinsap, C. y Raicharoen, T. (2013). Very short time environmental sound classification based on spectrogram pattern matching. *Information Sciences, 243,* 57-74.

Korpáš, J., Sadloňová, J. y Vrabec, M. (1996). Analysis of the cough sound: an overview. *Pulmonary Pharmacology, 9*(5-6), 261-268.

Krizhevsky, A., Sutskever, I. y Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM, 60*(6), 84-90.

Matos, S., Birring, S. S., Pavord, I.D. y Evans, D. H. (2006). Detection of Cough Signals in Continuous Audio Recordings Using Hidden Markov Models. *IEEE Transactions on Biomedical Engineering, 53*(6), 1078-1083.

Mitra, V. y Wang, C.-J. (2007). Content based audio classification: a neural network approach. *Soft Computing, 12*(7), 639-646.

Monge-Alvarez, J., Hoyos-Barcelo, C., Lesso, P. y Casaseca-de-la-Higuera, P. (2019). robust detection of audio-cough events using local hu moments. *IEEE Journal of Biomedical and Health Informatics, 23*(1), 184-196.

Ramos, D., Franco-Pedroso, J., Lozano-Diez, A. y González-Rodríguez, J. (2018). Deconstructing cross-entropy for probabilistic binary classifiers. *Entropy, 20*(3), 208.

Rawat, W. y Wang, Z. (2017). Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. *Neural Computation, 29*(9), 2352-2449.

Shorten, C. y Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data, 6*(1).

Spinou, A. y Birring, S. S. (2014). An update on measurement and monitoring of cough: what are the important study endpoints? *Journal of Thoracic Disease, 6*(Suppl 7), S728-S734. https://bit.ly/3oDrot4

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. y Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research, 15*(56), 1929-1958. https://bit.ly/433Xdu1

Van Hirtum, A. y Berckmans, D. (2002). Assessing the sound of cough towards vocality. Medical Engineering & Physics, 24(7-8), pp.535–540.

Wang, H.-H., Liu, J.-M., You, M. y Li, G.-Z. (2015). *Audio signals encoding for cough classification using convolutional neural networks: A comparative study*. 2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM).

Capítulo 4

Filtrado digital para señales audibles de tos

Introducción

Este capítulo trata sobre el proceso realizado para filtrar y normalizar las señales de tos para aislar los sonidos indeseados, utilizando filtrado adaptativo y de la ejecución de las actividades para el filtrado y normalización de la señal de tos en base de los datos analizados. Se presenta una descripción general de algunos conceptos relacionados a la tos, de las técnicas de eliminación de ruido de la señal de tos, los filtros digitales clásicos y los filtros adaptativos. Además, se describen trabajos relacionados con los filtros adaptativos.

Se trata la problemática del ruido presente en las señales de audio de tos durante una grabación, así como también la implementación de técnicas de filtrado para minimizar el ruido presente en dicho audio.

Filtros para eliminación de ruido

La tos es un mecanismo de defensa del cuerpo para limpiar el tracto respiratorio de materiales extraños que se inhalan accidentalmente o se producen internamente por infecciones (Amrulloh *et al.*, 2015)

La tos es un sonido seco repentino y contundente para liberar aire y despejar una irritación en la garganta o las vías respiratorias. Las señales de tos transportan información vital ya que es responsable de diversas enfermedades respiratorias como bronquitis, asma, neumonía, etc. (Shankar *et al.*, 2020).

Aunque la tos es común en las enfermedades respiratorias y se considera un síntoma clínico importante, no existe un estándar único para evaluarla. En una sesión de consulta típica, los médicos pueden escuchar varios episodios de tos espontánea o voluntaria, para obtener información cualitativa como la "humedad" de una tos. Esta información cualitativa es extremadamente útil en el diagnóstico y el tratamiento de enfermedades respiratorias (Amrulloh *et al.*, 2015).

Las señales de tos también proporcionan ciertos datos esenciales de las vías respiratorias de los pulmones que ayudan en el diagnóstico de enfermedades relacionadas con los pulmones (Abeyratne *et al.*, 2013). Es evidente que siempre que se han analizado las señales de tos existe la posibilidad de contaminación acústica en las señales de tos, como el ruido ambiental cuando se registra una tos, los dientes y el sonido de la saliva cuando una persona abre la boca, etc. Por lo tanto, es necesario filtrar estas señales ruidosas para un análisis adecuado y preciso de las señales de tos (Shankar *et al.*, 2020). En los años anteriores se han utilizado diversas metodologías de filtrado para el particular, y las señales relacionadas con las vías respiratorias incluyen la eliminación de ruido generalmente con filtros de paso bajo (Aggarwal *et al.*, 2011). Se han propuesto diversas técnicas de modelado, como la descomposición en modo empírico (Liang *et al.,* 2005) (Blanco-Velasco *et al.,* 2008).

Pero las técnicas solo son útiles cuando las señales de sonido de la tos están contaminadas con un solo tipo de ruido. En los últimos años se han derivado diferentes enfoques para eliminar el ruido de las señales de sonido de la tos, como la Transformada de Fourier de Corta duración (Short-time Fourier Transform, STFT), la Transformada de Fourier (Fourier Transform, FT) y la transformada de Gabor. El inconveniente de FT es que muestra la información de las señales de

sonido de la tos solo en el dominio de la frecuencia. El STFT tiene análisis de tiempo-frecuencia, pero tiene una ventana deslizante fija (Tary *et al.*, 2018). Las principales restricciones de estos métodos son que generalmente se evita la continuidad del sonido de la tos. Recientemente, los investigadores han descubierto métodos basados en ondículas para la caracterización multiescalar y el análisis de señales. Estas herramientas son diferentes de la FT convencional, en la que la información está restringida en el plano de tiempo y frecuencia; básicamente, son capaces de intercambiar un tipo de resolución por otro, lo que los hace especialmente adecuados para el análisis de señales variables. El uso de las Transformadas Wavelets (WT) se ha implementado ampliamente (Poornachandra, 2008). En este artículo (Poornachandra, 2008), implementó una técnica basada en umbrales estrictos que utiliza WT continuas para eliminar el ruido de las señales de sonido de la tos (Shankar *et al.*, 2020).

Filtros digitales

El propósito de los filtros digitales es separar las señales que se han combinado y restaurar las señales que se han distorsionado de alguna manera (Mills *et al.*, 1978). La separación de la señal es necesaria cuando una señal se ha contaminado con interferencias, ruido u otras señales, mientras que la restauración se utiliza cuando una señal se ha distorsionado de alguna manera. En general, los filtros digitales se clasifican como filtros de Weiner y Kalman (Krishnan, s. f.; Dixit y Nagaria, 2017).

Filtros adaptativos

Un filtro adaptativo (Ram *et al.*, 2012) es un sistema con un filtro lineal que consiste en una función de transferencia restringida por parámetros variables y un medio para ajustar esos parámetros de acuerdo con un algoritmo de optimización. Los filtros lineales adaptativos (Park y Meher, 2013) son sistemas dinámicos lineales con estructura y parámetros variables o adaptables y tienen la propiedad de modificar los

valores de sus parámetros, es decir, su función de transferencia, durante el procesamiento de la señal de entrada, para generar una señal en la salida; que es sin componentes no deseados, ruido y degradación y también señales de interferencia (Dixit y Nagaria, 2017).

Los algoritmos adaptativos (Chandrakar y Kowar, 2012) se han estudiado ampliamente en las últimas décadas y los algoritmos adaptativos más populares son el Algoritmo de Mínimos Cuadrados medios (Least Mean Square, LMS) y el algoritmo de mínimos cuadrados recursivos (Recursive Least Squares, RLS). Lograr el mejor rendimiento de un filtro adaptativo requiere el uso del mejor algoritmo adaptativo con baja complejidad computacional y una tasa de convergencia rápida (Dixit y Nagaria, 2017).

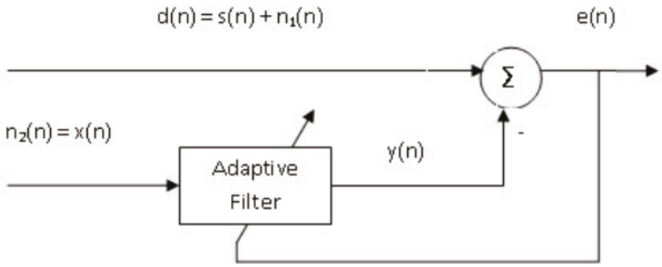
Dado que el ruido del entorno circundante reduce gravemente la calidad de las señales de voz y audio, es muy necesario suprimir el ruido y mejorar la calidad de la señal de voz y audio, por lo que las aplicaciones acústicas de la cancelación de ruido se han convertido en el área principal de investigación. El concepto básico de Cancelación de Ruido Adaptativo (Adaptive Noise Canceller, ANC), que elimina o suprime el ruido de una señal mediante filtros adaptativos, fue introducido por primera vez por Widrow (Yan *et al.*, 2010). Debido a las largas respuestas de impulso, los requisitos computacionales de los filtros adaptativos son muy altos, especialmente durante la implementación en procesadores de señales digitales. Donde, como en el caso de entornos no estacionarios y el ruido de fondo coloreado, la convergencia se vuelve muy lenta si el filtro adaptativo recibe una señal con un alto rango dinámico espectral (Sayadi y Shamsollahi, 2008) (CR. Para superar este problema se han propuesto numerosos enfoques en las últimas décadas (Aggarwal *et al.*, 2011).

La cancelación del ruido acústico es indispensable desde el punto de vista de la salud, ya que las exposiciones extensas a un alto nivel de ruido pueden causar graves riesgos para la salud de las personas. El método de cancelación de ruido convencional (Ram *et al.*, 2012) usa una señal de entrada de referencia (señal de ruido correlacionada)

que se pasa a través del filtro adaptativo para igualar el ruido que se agrega a la señal de soporte de información original.

Posteriormente, esta señal filtrada se resta de la señal de información alterada por ruido. Esto hace que la señal corrupta sea una señal libre de ruido. El concepto fundamental de la cancelación de ruido (Ram *et al.*, 2012) es producir una señal que sea igual a una señal de perturbación en amplitud y frecuencia, pero que tenga una fase opuesta. Estas dos señales dan como resultado la cancelación de la señal de ruido. La Cancelación de Ruido Adaptativa (ANC) original (Park y Meher, 2013) utiliza dos sensores para recibir la señal de ruido y la señal de destino por separado. La relación entre la referencia de ruido x (n) y el componente de este ruido que está contenido en la señal medida $d_{(n)}$ puede determinarse mediante la cancelación de ruido adaptativa que se muestra en la figura 1 (Aggarwal *et al.*, 2011).

Figura 1
Filtros adaptativos



Si varios ruidos no relacionados alteran la medición de interés, entonces se pueden implementar varios filtros adaptativos en paralelo siempre que las señales de referencia de ruido adecuadas estén disponibles dentro del sistema. En los sistemas de cancelación de ruido, el objetivo es producir una salida del sistema $e_{(n)} = [s_{(n)} + n1] - y_{(n)}$ que se ajusta mejor en el sentido de mínimos cuadrados a la señal $s_{(n)}$. Este objetivo se logra ajustando el filtro a través de un algoritmo adaptativo y retroalimentan-

do la salida del sistema al filtro adaptativo y minimizando la potencia de salida total del sistema (Park and Meher, 2013). En un sistema de Cancelación de Ruido Adaptativo, la salida del sistema sirve como una señal de error para el proceso adaptativo (Aggarwal *et al.*, 2011).

Metodologías de eliminación de ruido (técnicas)

Algunas metodologías (técnicas) que no minimizan las interferencias pueden ser: la primera utilizando filtros adaptativos y la segunda mediante auto-encoders.

La metodología de filtrado adaptativo operativo se presenta en la figura 2. El sistema consta de cuatro etapas:

1. La primera lleva la recolección de datos de tos, para la adquisición de datos, se seleccionarán registros históricos de conjuntos de datos de tos de diferentes bases de datos publicadas y ubicadas en diferentes países para garantizar la capacidad de generalización de la técnica propuesta.

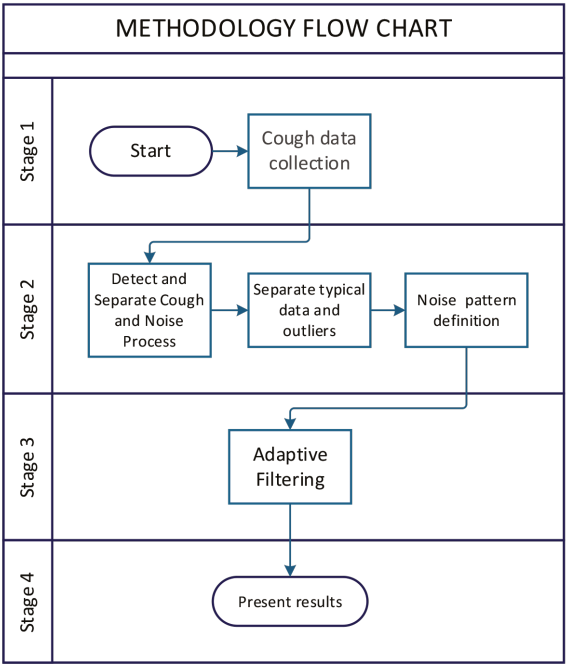
2. La segunda fase realiza tres procesos correspondientes a la detección y separación de tos y ruido de un archivo de audio, la separación de datos típicos y atípicos en dichos grupos, y finalmente la definición de un patrón de ruidos que representan las interferencias en un archivo de registro de tos.

3. La tercera fase corresponde al uso de una técnica de filtrado adaptativo para minimizar las interferencias en un archivo de registro de tos.

4. Finalmente, en la fase 4 se presentan los resultados.

Figura 2

Diagrama de flujo de la metodología propuesta

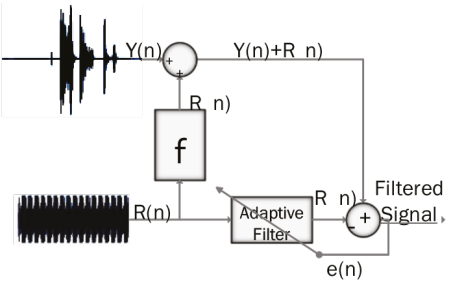


El filtro adaptativo necesita una señal de referencia de ruido base que esté asociada con el ruido original presente en el registro de tos que pretende filtrar. Para probar la funcionalidad del filtrado adaptativo, se presenta la figura 3, en la cual, la entrada del sistema corresponde a una señal ideal sin ruido representada como $Y_{(n)}$. Esta señal se perturba con una señal de ruido representada como $R'_{(n)}$ que es el resultado de una aplicación de proceso "f " de una señal de ruido de referencia $R_{(n)}$. Por tanto, la suma de las señales $Y_{(n)}$ y $R'_{(n)}$ es la señal que debe corregirse mediante filtrado adaptativo. La señal $R'_{(n)}$ es la señal de entrada para el filtro adaptativo también, en consecuencia, el filtro adaptativo replicará el proceso "f" entre las señales $R_{(n)}$ y $R'_{(n)}$ para minimizar el ruido al final del proceso. El filtro adaptativo usa la

señal $e_{(n)}$ como orientación para optimizarse usando un algoritmo de optimización. En el trabajo (Amrulloh *et al.*, 2015) se han analizado diferentes tipos de algoritmos y en función de la convergencia y la velocidad se recomienda utilizar el algoritmo de Mínimos Cuadrados Recursivos (RLS) (Shankar *et al.*, 2020).

Figura 3

Diagrama de la metodología de filtrado adaptativo



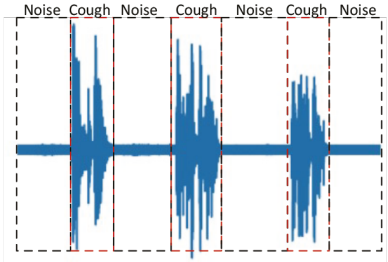
La clave de este enfoque es obtener la señal de referencia de ruido R (n). Este trabajo propone una metodología estadística para crear esta referencia de ruido que involucra una definición de patrón de ruido utilizando el conjunto de datos de tos. Las siguientes secciones describen esta metodología en detalle.

Detectar y separar el ruido de los registros de tos

Originalmente en los archivos de datos de grabación de audio se tienen dos señales importantes, la primera corresponde a la señal de tos y la segunda corresponde al ruido. Se supone que el ruido está presente antes de una expectoración de la tos como se puede ver en la figura 4. Para separar este ruido y la señal de tos es necesario analizar la energía de la señal, por lo que utilizando un umbral de decisión se pueden separar las señales. Utilizando el proceso mencionado en todos los conjuntos de datos de registros de tos, se crea un nuevo conjunto de datos de ruidos presentes en los registros de tos.

Figura 4

Tos y ruido en una señal de audio



Separación de datos típicos y atípicos

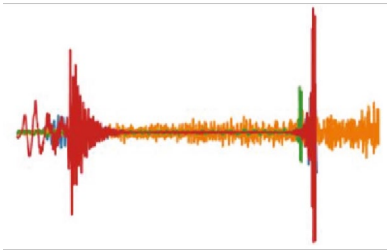
Para el conjunto de datos descrito en la sección anterior es necesario procesar los valores de los valores atípicos, por lo que un valor se considera un valor atípico cuando está fuera del intervalo de confianza del 95 % en una función de distribución normal y está definido por la media más o menos dos desviaciones estándar aproximadamente. Este procedimiento se ha realizado y se han obtenido el valor medio y la desviación estándar de todos los conjuntos de datos.

Creación del patrón de ruido

Finalmente, el patrón de ruido se denomina "señales de patrón final" y está formado por el conjunto de señales que pertenecen al conjunto de datos de ruidos típicos. Para ajustarse a la señal de referencia para el filtro adaptativo se calcula la media del patrón de ruido, este patrón se puede observar en la figura 5.

Figura 5

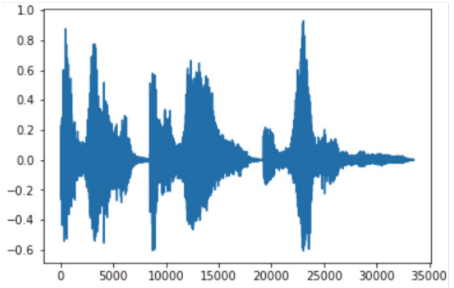
Definición de patrón de ruido



Filtrado de las señales de audio de entrada

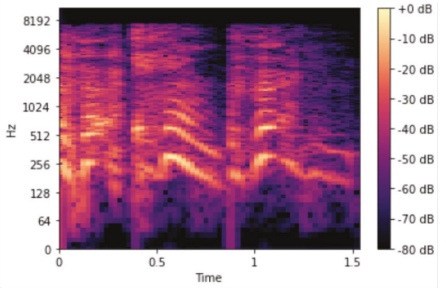
Una señal de audio de tos es similar a la observada en la figura 6, en la cual claramente se pueden observar dos componentes, una correspondiente a las expectoraciones de tos caracterizadas por tener mayor energía, y otra componente correspondiente al ruido presente, la cual es distinguible antes de cada expectoración.

Figura 6
Señal de audio de tos



Otra forma de analizar una señal es por medio de su espectrograma (figura 7), el cual es una distribución en el tiempo de cada una de las componentes frecuenciales de la señal de audio.

Figura 7
Espectrograma



Filtro digital técnica tradicional

A continuación, se presentan los resultados de aplicar un filtro digital clásico en una señal de audio de tos, la señal de audio es posible observar en la figura 8, y la señal filtrada en la figura 9.

Figura 8
Señal de audio sin filtrar

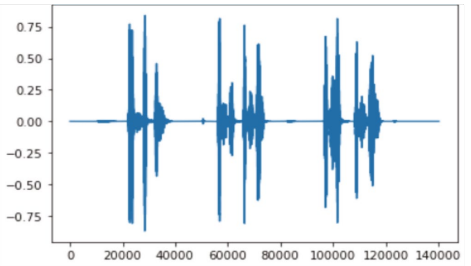
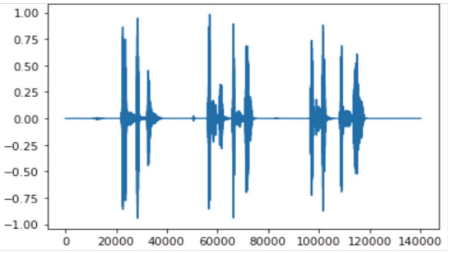


Figura 9
Señal de audio filtrada



En las figuras 10 y 11 se presenta el espectrograma de la señal de audio antes de filtrar y después del proceso de filtrado.

Figura 10

Espectrograma de señal de audio sin filtrar

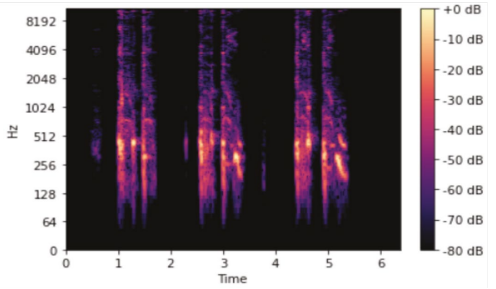
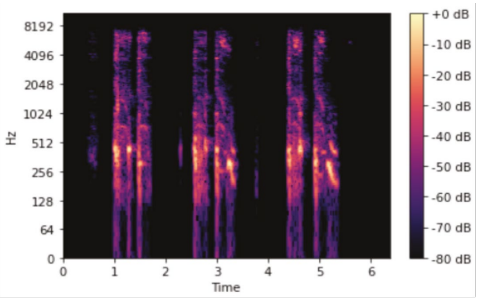


Figura 11

Espectrograma de señal de audio después del proceso de filtrado

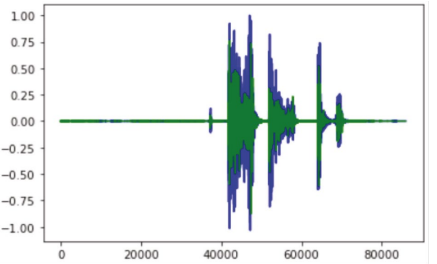


Filtrado adaptativo

En la figura 12 se presenta una señal de audio sin filtrar (en color azul) y una señal de audio luego de aplicar un filtrado adaptativo.

Figura 12

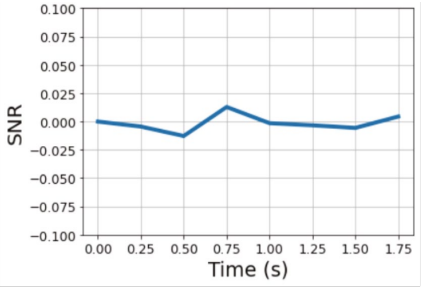
Señal de audio antes y después de aplicar filtrado adaptativo



En la figura 13 se presenta la relación señal a ruido de la señal filtrada con técnica adaptativa a lo largo del tiempo.

Figura 13

Relación señal a ruido de un audio de tos después de aplicar filtrado adaptativo



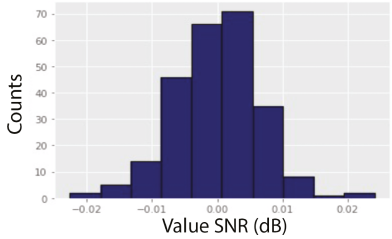
Aplicación en la base de datos inicial y final

Para demostrar la generalidad de la metodología propuesta se utiliza la técnica adaptativa con la misma señal de referencia que se obtuvo en todos los conjuntos de datos de tos y se captura el valor SNR de cada registro. En la figura 14 se presenta un histograma en el que el eje de los contenedores corresponde al número de señales de tos que contienen valores similares de SNR. Para el análisis se parte de la idea de que en un caso ideal todas las señales comparten el mismo valor de SNR, pero en un caso real, el valor de SNR no tiene una variación relativa, como se puede ver en la figura 14; porque la mayoría de los registros de tos comparten el mismo valor centrado en aproximadamente 0,0001 dB en ambas bases de datos.

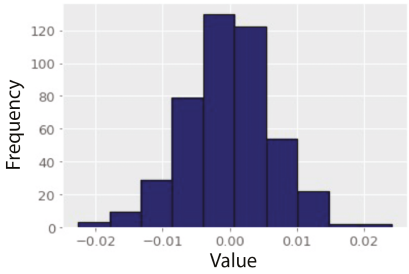
Figura 14

Histogramas en el que el eje de los contenedores corresponde al número de señales de tos que contienen valores similares de SNR

Base de datos iniciales



Base de datos finales



De esta manera se ha definido un "patrón de ruido" que puede contener la información de todo tipo de ruidos que contaminan una señal de tos por medio de filtros adaptativos. Como se ha indicado, se puede minimizar el ruido en un archivo de registro de tos en un factor cercano a 0 dB, resultando en una técnica de preprocesamiento óptima en un sistema inteligente de caracterización y clasificación de la tos.

Referencias bibliográficas

Abeyratne, U. R., Swarnkar, V., Setyati, A. y Triasih, R. (2013). Cough sound analysis can rapidly diagnose childhood pneumonia. *Annals of Biomedical Engineering, 41*(11), 2448-2462. https://doi.org/10.1007/S10439-013-0836-0

Aggarwal, R., Singh, J. K., Gupta, V. K., Rathore, S., Tiwari, M. y Khare, A. (2011). Noise reduction of speech signal using wavelet transform

with Modified Universal Threshold. *International Journal of Computer Applications, 20*(5), 14-19. https://doi.org/1010.5120/2431-3269

Amrulloh, Y. A., Abeyratne, U. R., Swarnkar, V., Triasih, R. y Setyati, A. (2015). Automatic cough segmentation from non-contact sound recordings in pediatric wards. Biomedical *Signal Processing and Control, 21*, 126-136. https://doi.org/10.1016/j.bspc.2015.05.001

Blanco-Velasco, M., Weng, B. y Barner, K. E. (2008). ECG signal denoising and baseline wander correction based on the empirical mode decomposition. *Computers in Biology and Medicine, 38*(1), 1-13. https://doi.org/10.1016/J.COMPBIOMED.2007.06.003

Chandrakar, Ch. y Kowar. M. K. (2012). Denoising ECG signals using adaptive filter algorithm 121. *International Journal of Soft Computing and Engineering (IJSCE), 2*(1). https://bit.ly/43BUZ58

Dixit, S. y Nagaria, D. (2017). LMS adaptive filters for noise cancellation: a review. *International Journal of Electrical and Computer Engineering (IJECE), 7*(5), 2520-2529. https://doi.org/10.11591/IJECE.V7I5.PP2520-2529.

Krishnan, V. (s. f.) *Probability and random processes.* Wiley. https://bit.ly/3MI3Viu

Liang, H., Lin, Q. H. y Chen, J. D. Z. (2004). *Application of the Empirical Mode Decomposition to the analysis of esophageal manometric data in gastroesophageal reflux disease.* Conference proceedings. Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference, 2006, 620-623. https://doi.org/10.1109/IEMBS.2004.1403234

Liang, H., Lin, Z. y Yin, F. (2005). Removal of ECG contamination from diaphragmatic EMG by nonlinear filtering. *Nonlinear Analysis, Theory, Methods and Applications, 63*(5-7), 745-753. https://doi.org/10.1016/J.NA.2004.09.018

Mills, W. L., Mullis, C. T. and Roberts, R. A. (1978). Digital filter realizations without overflow oscillations. *IEEE Transactions on Acoustics, Speech, and Signal Processing, 26*(4), 334-338. https://doi.org/10.1109/TASSP.1978.1163114

Park, S. Y. y Meher, P. K. (2013). Low-Power, High-Throughput, and Low-Area Adaptive FIR Filter based on Distributed Arithmetic. *IEEE Transactions on Circuits and Systems II: Express Briefs, 6*(60), 346-350. https://doi.org/10.1109/TCSII.2013.2251968

Poornachandra, S. (2008). Wavelet-based denoising using subband dependent threshold for ECG signals. *Digital Signal Processing, 18*(1), 49-55. https://doi.org/10.1016/J.DSP.2007.09.006

Ram, M. R., Venu Madhav, K., Hari Krishna, E., Komalla, N, R. y Ashoka Reddy, K. (2012). A novel approach for motion artifact reduction in PPG signals based on AS-LMS adaptive filter. *IEEE Transactions on Instrumentation and Measurement, 61*(5), 1445-1457. https://doi.org/10.1109/TIM.2011.2175832

Sayadi, O. y Shamsollahi, M. B. (2008). ECG denoising and compression using a modified extended Kalman filter structure. *IEEE transactions on bio-medical engineering, 55*(9), 2240-2248. https://doi.org/10.1109/TBME.2008.921150

Shankar, A., Bhateja, V., Srivastava, A. y Taquee, A. (2020). Continuous wavelets for pre-processing and analysis of cough signals. *Smart Innovation, Systems and Technologies, 159,* 711-718. https://doi.org/10.1007/978-981-13-9282-5_68

Tary, J. B., Herrera, R. H. y Van Der Baan, M. (2018). Analysis of time-varying signals using continuous wavelet and synchrosqueezed transforms. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 376*(2126). https://doi.org/10.1098/RSTA.2017.0254

Yan, J., Lu, Y., Liu, J., Wu, X. y Xu, Y. (2010). Self-adaptive model-based ECG denoising using features extracted by mean shift algorithm. *Biomedical Signal Processing and Control, 5*(2), 103-113. https://doi.org/10.1016/J.BSPC.2010.01.003

Capítulo 5

Modelado automático de una señal de tos COVID-19

Introducción

La caracterización de la tos provocada por el COVID-19 parte del uso del audio de tos de personas con un diagnóstico positivo o negativo de esta enfermedad. Estos audios contienen una secuencia de tos que está conformada de múltiples acciones, cada una formada de una inspiración seguido de una o múltiples acciones de "tos". Sin embargo, una secuencia de tos también posee segmentos de silencio. Con la eliminación de los silencios, se busca optimizar y simplificar la información presentada en estos audios, los cuales serán usados como ingreso a los algoritmos encargados de caracterizar dicha tos. El método escogido para el desarrollo de este algoritmo es el propuesto por Pramono Renard *et al.* (2016), donde se realiza una detección de los silencios en secuencias de tos de la enfermedad de pertussis.

Por otro lado, el uso de la inteligencia artificial para ayudar a combatir la pandemia del COVID-19 ha ido incrementando progresivamente, con el uso de las señales de tos provocadas por esta enfermedad. Las señales de audio generadas por el cuerpo humano

han sido usadas para detectar la presencia de ciertas enfermedades o el progreso de estas a través de la aplicación de algoritmos capaces de encontrar patrones característicos de una cierta patología. Recientes investigaciones han aplicado esta tecnología para realizar un diagnóstico de la COVID-19 a través de la voz y la tos.

Los autores Brown Chloë *et al.* (2020) presentan un algoritmo capaz de distinguir a una persona positiva de COVID-19 de una persona sana, así como de toses de personas que tienen asma. Este algoritmo aplica "transfer learning" con "VGG net" además de otro tipo de fuentes de información como Mel Frecuency Cepstral Coefficient (MFCCs), Energía, Cruce por cero, entre otras. Con lo cual logran un 0,82 de AUC, 0,80 de *precisión* y un 0,72 de *recall* en diferenciar una persona sana de una infectada de COVID-19. Laguarta *et al.* (2020) utilizan grabaciones de tos que son convertidas a MFCCs para usarlas como ingreso a tres modelos en paralelo ResNet-50 pre entrenados que están basados en redes neuronales convolucionales (CNN), logrando un modelo con un 98,5 % de sensibilidad y una especificidad del 94,2 % y en personas asintomáticas se logra un 100 % de sensibilidad y un 83,2 % de especificidad.

Imran *et al.* (2020) estudian las alteraciones dadas en el sistema respiratorio producidas por la enfermedad del COVID-19 y lo compara con otras enfermedades con el objetivo de observar si realmente existen características que diferencian a esta enfermedad de otras. Los autores usan un modelo conformado de tres diferentes sistemas entrenados con "Transfer Learning" los cuales son usados para combinar sus salidas y decidir con base en las tres decisiones el diagnóstico final del modelo, esto con el fin de reducir al mínimo un diagnóstico incorrecto. El primer modelo está basado en una red neuronal convolucional que utiliza como entrada espectrogramas de Mel junto con un clasificador con cuatro clases en las que se encuentran COVID-19 y otro tipo de enfermedades similares; el segundo modelo usa un algoritmo de Machine Learning utilizando extractores de características basadas en MFCC y la técnica de PCA, además de usar un clasificador "Super Vector Machine"

(SVM). El tercer modelo consta de una red convolucional que usa como ingreso espectrogramas de Mel, siendo una arquitectura muy similar a la primera diferenciándose en que posee un clasificador binario a su salida. El modelo final conformado de las tres arquitecturas logran una precisión del 86,60 % y una exactitud del 88,76 % al diferenciar una persona sana de una que posee COVID-19. Todos estos artículos utilizan bases de datos obtenidas individualmente recolectados en su mayoría en línea a través de internet para la recolección masiva de grabaciones de tos. Por esta razón, en la sección 5.2 se analizan varias técnicas para caracterizar la señal de tos de pacientes con COVID-19. De la misma manera, en la sección 5.3 se continúa con el análisis para determinar las características en señales de audios de tos de pacientes que poseen COVID-19, utilizando bases de datos modificadas con técnicas "data augmentation".

Como parte de la ejecución de este proyecto, cuyo objetivo general es: "Caracterizar la tos provocada por el COVID-19 en pacientes con diagnóstico positivo, utilizando técnicas de Inteligencia Artificial y Procesamiento Digital de Señales, con el fin de contribuir científicamente en la identificación de la enfermedad y tratamiento oportuno, así como en la economía social por la época de crisis del país"; se presenta el diseño final del sistema integrado de reconocimiento de una tos COVID-19, donde se integran todas las fases del proyecto.

Finalmente, se presentan los resultados, tras la implementación del sistema integrado de reconocimiento, con base en las metodologías y técnicas analizadas.

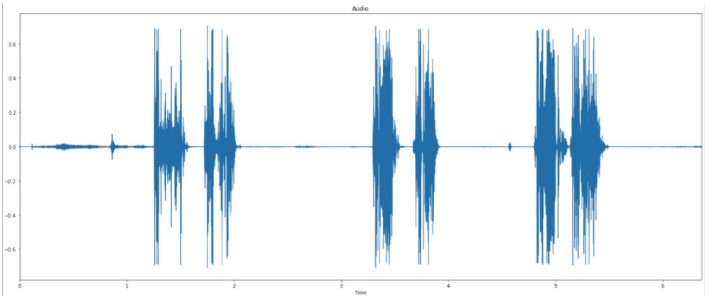
Eliminación de silencios en una secuencia de la señal de la tos

Algoritmo

Una señal de un audio de una secuencia de tos se presenta en la figura 1.

Figura 1

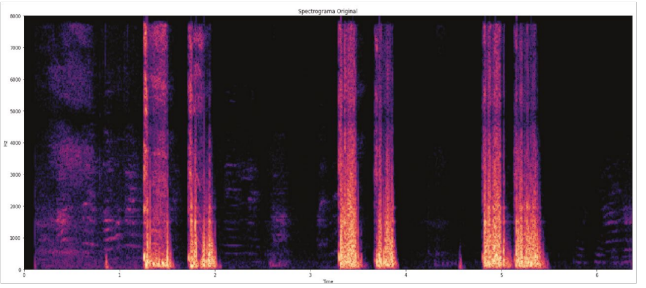
Señal de audio de secuencia de tos



En esta se observa que una secuencia de tos está conformada de múltiples acciones de inspiración y de toses. Esta señal puede ser visualizada a través de un espectrograma que muestra las componentes de frecuencia de dicha señal, esta se presenta en la figura 2.

Figura 2

Espectrograma de señal de tos



Para la obtención de estas señales se utilizan los parámetros presentados en la tabla 1.

Tabla 1

Parámetros

Parámetros	
Frecuencia de Muestreo	16Khz
N_fft (Transformada de Fourier - Espectrograma)	512
Hop_Lenght (Transformada de Fourier - Espectrograma)	128

La detección de los sonidos en el audio de tos es realizada mediante la comparación de la desviación estándar de cada *frame* (ventana pequeña de milisegundos) a la media de la desviación estándar de cada grabación de tos. Con lo cual se fija en un límite mínimo para la desviación estándar de cada *frame* con el objetivo de decidir si cierto *frame* posee silencio o no. Los límites para realizar estas detecciones son fijados de manera empírica hasta la obtención de resultados aceptables. Los *frames* descritos son las mismas ventanas obtenidas de la extracción del espectrograma con la transformada de Fourier. Los pasos del algoritmo diseñado son:

- Primeramente, se fija un límite máximo de desviación estándar para que se detecte un sonido, al detectar uno se fija un límite mínimo para que el algoritmo ahora detecte una bajada de energía y por lo tanto el fin del sonido.
- Posteriormente, se fijó un límite máximo para el ancho mínimo de detección del sonido, si un sonido detectado es muy fino y este no supera el límite fijado este no se mantiene.
- Finalmente, el último límite fijado es la cantidad de energía existente dentro del sonido, este se coloca con el objetivo de evitar detectar sonidos con picos pequeños de energía.

Este procedimiento se muestra en un diagrama de flujo presentado en la Figura 3. Todos los límites descritos tienen como unidad la cantidad de desviación estándar de cada trama analizada. Estos parámetros se presentan en la tabla 2.

Tabla 2
Límites

Límites-Desviación Estándar (SD)	
Presencia de sonido	1
Ausencia de sonido	0,01
Ancho de sonido	0,14
Cantidad de energía en la detección	0,1

Una vez realizada las detecciones de los sonidos se obtiene el siguiente gráfico, que se muestra en la figura 3.

Figura 3
Proceso del algoritmo

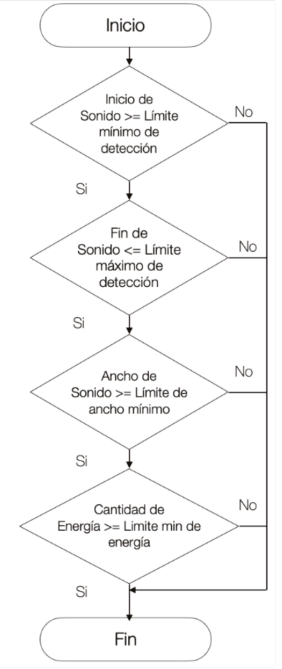
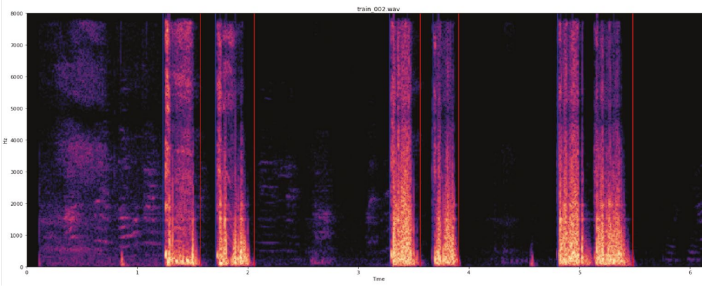


Figura 4
Detección de sonidos



Donde las líneas azules son el inicio del sonido mientras las líneas rojas indican la terminación de este. Luego de realizar estas detecciones se elimina todo lo demás del sonido y se obtiene un sonido libre de silencios como se presenta en la figura 5 y su correspondiente espectrograma en la figura 6.

Figura 5
Sonido de silencios

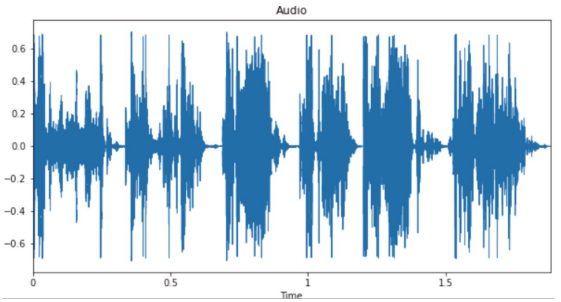
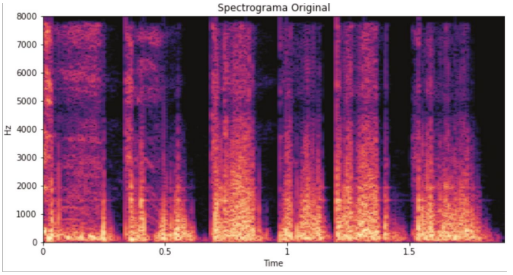


Figura 6
Espectrogramas sin silencios



Técnicas para caracterizar la señal de tos de pacientes que poseen COVID-19

Algoritmo: red neuronal convolucional simple y entrenada

La base de datos inicial a usar se presenta en la tabla 3.

Tabla 3
Base de datos

	N. Archivos	N. Positivos	N. Negativos	D. Promedio	D. Máxima	D. Mínima
Train	286	71	215	6,47s	17,07s	2,48s
Devel	231	48	183	6,10s	16,11s	2,04s

Con esta base de datos inicial se procedió a usarla para entrenar un modelo inicial básico que está conformado de una red neuronal convolucional simple, la cual toma como ingreso imágenes de los espectrogramas de cada uno de los audios de los cuales se conforma la base de datos. Los datos para la obtención de la imagen y el espectrograma se presentan a continuación:

Espectrograma: N_fft = 512, Hop_Lenght = 128

Tamaño de Imagen: 64 x 64 pixeles

La arquitectura de la red neuronal convolucional simple se presenta en la tabla 4.

Tabla 4
Parámetros

Parámetros	
Rescaling	1./255
Conv2D / Relu	16 Filters, 3x3 Kernel
MaxPooling2D	
Conv2D / Relu	32 Filters, 3x3 Kernel
MaxPooling2D	
Conv2D / Relu	64 Filters, 3x3 Kernel
MaxPooling2D	
Flatten	
Dense	64 Neurons / Relu
Dropout	0,4
Dense	2 Neurons / Relu

Esta arquitectura fue usada para analizar los resultados obtenidos con una red simple y también analizar aspectos de la base de datos inicial usada en estos experimentos. Los resultados usando esta red neuronal se presentan en la figura 7 y tabla 5.

Figura 7
Pérdida y exactitud del modelo

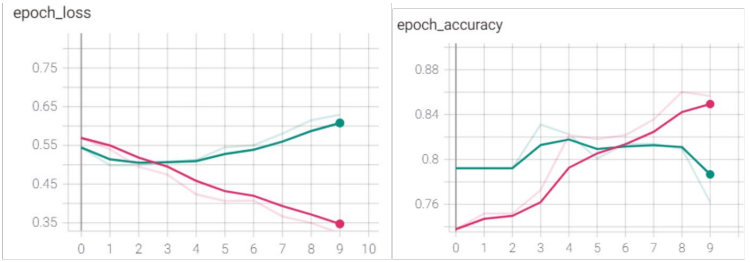


Tabla 5
Resultados modelo inicial

	Resultado						
	Train		**Devel**				
	Loss	Accuracy	Loss	Accuracy	Precision	F1-Score	Recall
Métricas	0,47	77,27 %	0,50	83,12 %	18,75 %	31,57 %	100 %

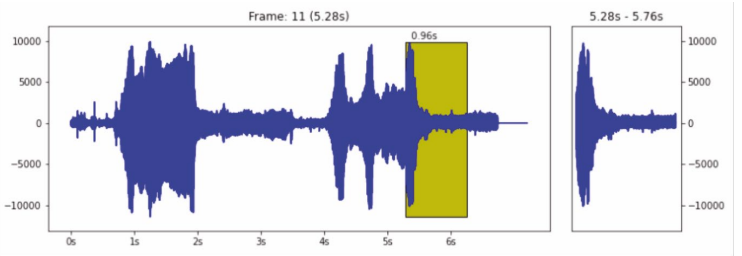
Estas imágenes presentan las curvas de pérdida y exactitud del modelo siendo la curva roja la curva de entrenamiento y la verde la curva de evaluación. En estas se puede observar que el modelo presenta un gran sobre entrenamiento debido a la poca cantidad de audios y el desequilibrio de la base de datos. Además, en la tabla 5 se puede observar en las métricas de *precisión* y *recall* que el modelo tiene un buen rendimiento en reconocer audios de personas negativas en COVID-19 teniendo un *recall* muy alto, pero tiene muchos fallos en la detección de positivos lo cual se presenta en la precisión del modelo. Siendo el *F1-Score* la métrica que muestra el rendimiento del modelo con la combinación de la precisión y el *recall*.

Entrenar una red neuronal desde cero requiere una gran base de datos para ayudar al aprendizaje de la red y evitar un sobre entrenamiento en los datos lo cual produce una mala generalización de la red neuronal. Una alternativa tomada por varias investigaciones en estos casos es el uso de una red neuronal que ya haya sido entrenada en otro tipo de datos y usarla para usar ese conocimiento y resolver otra tarea, siendo este método llamado "Transfer-Learning".

Actualmente en esta investigación se está analizando este método, las pruebas iniciales realizadas utiliza una red neuronal llamada "Yamnet" que es una red neuronal entrenada para clasificar 521 clases de sonidos en los que se encuentran desde animales, vehículos hasta el sonido de una tos. Debido a que esta red ya tiene conocimiento de que es una tos, el objetivo principal es usarla para entrenarla de nuevo y analizar su rendimiento en esta tarea. Esta red neuronal convolucional utiliza ventanas pequeñas de 0.96s las cuales se desplazan por la imagen obteniendo un vector de características por cada una de ellas, las cuales son usadas posteriormente para realizar una clasificación del sonido, esto se presenta en la figura 8.

Figura 8

Yamnet



En los experimentos iniciales se ha utilizado esta red para entrenarla en la base de datos mostrada en la tabla 3, obteniendo los resultados presentados en la figura 9 y tabla 6.

Figura 9

Curvas de pérdida y exactitud

Tabla 6

Resultados

Resultado 4 EPOCH							
	Train		Devel				
	Loss	Accuracy	Loss	Accuracy	Precision	Recall	F1-Score
Métricas	0,59	75,17 %	0,65	74,45%	33,33 %	22,91 %	27,15 %

En la figura 9 se puede observar que ya no se presenta un sobre entrenamiento de la red, esto debido a que no se inició un modelo desde cero. Sin embargo, se pudo observar que el desequilibrio de la base de datos, afecta en el aprendizaje del modelo. En la tabla 6 se puede observar también que a pesar que la red tiene un buen resultado de *Accuracy* en las predicciones, este dato no representa totalmente los resultados, ya que la base de datos al tener más datos negativos que positivos la red tiende a predecir todos los negativos pero los positivos no los detecta, lo cual produce un buen resultado en *Accuracy*, pero resultados muy malos de *precisión* y *recall* que expresa los resultados tomando en cuenta los falsos positivos como negativos de las predicciones. La matriz de confusión se presenta en la tabla 7.

Tabla 7
Matriz de confusión

	Matriz de confusión		
Labels	**0**	161	22
	1	37	11
		0	**1**
		Predicción	

En esta se puede observar que el modelo es muy bueno prediciendo casos negativos, pero muy malo en predecir los positivos lo cual es producido por el desequilibrio de la base de datos.

Para resolver este desequilibrio se está analizando diferentes técnicas, una de ellas es la creación de nuevos datos, la primera que se ha analizado es el aumento de la amplitud del sonido de las muestras positivas hasta 20 dB, con lo cual incrementamos la cantidad de muestra y tratamos de equilibrar la base de datos. Los resultados obtenidos luego de la realización de esta técnica se presentan en la tabla 8 y la matriz de confusión en la tabla 9.

Tabla 8
Resultados- incremento de amplitud

Resultado 4 EPOCH							
	Train		**Devel**				
	Loss	Accuracy	Loss	Accuracy	Precision	Recall	F1-Score
Métricas	0,72	59,94 %	0,70	46,75 %	20,93 %	56,25 %	30,50 %

Tabla 9
Matriz de confusión

	Matriz de confusión		
Labels	**0**	81	102
	1	21	27
		0	**1**
		Predicción	

Análisis de características en señales de audios de tos de pacientes que poseen COVID-19 utilizando bases de datos con técnicas "Data Augmentation"

Red Neuronal Básica

La base de datos inicial a usar se presenta en la tabla 10.

Tabla 10
Base de datos

	N. Archivos	N. Positivos	N. Negativos	D. Promedio	D. Máxima	D. Mínima
Train	286	71	215	6,47s	17,07s	2,48s
Devel	231	48	183	6,10s	16,11s	2,04s

Con esta base de datos inicial se procedió a usarla para entrenar un modelo inicial básico que está conformado de una red neuronal convolucional simple la cual toma como ingreso imágenes de los espectrogramas de cada uno de los audios de los cuales se conforma la base de datos. Los datos para la obtención de la imagen y el espectrograma se presentan a continuación:

Espectrograma: N_fft = 512, Hop_Lenght = 128

Tamaño de Imagen: 64 x 64 pixeles

La arquitectura de la red neuronal convolucional simple se presenta en la tabla 11.

Tabla 11
Parámetros

Parámetros	
Rescaling	1./255
Conv2D / Relu	16 Filters, 3x3 Kernel
MaxPooling2D	
Conv2D / Relu	32 Filters, 3x3 Kernel
MaxPooling2D	
Conv2D / Relu	64 Filters, 3x3 Kernel
MaxPooling2D	
Flatten	
Dense	64 Neurons / Relu
Dropout	0,4
Dense	2 Neurons / Relu

Esta arquitectura fue usada para analizar los resultados obtenidos con una red simple y también analizar aspectos de la base de datos inicial usada en estos experimentos. Los resultados utilizando esta red neuronal se presentan en la figura 10 y tabla 12.

Figura 10
Pérdida y exactitud del modelo

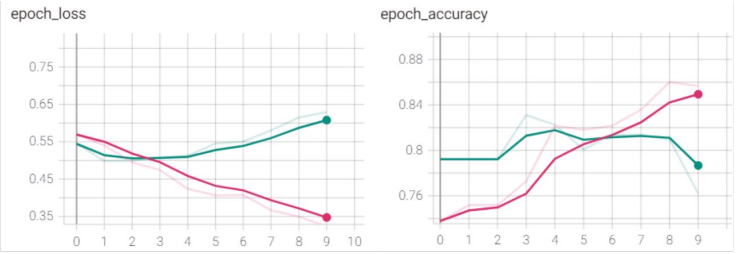


Tabla 12
Resultados modelo inicial

	Resultado						
	Train		**Devel**				
	Loss	Accuracy	Loss	Accuracy	Precision	Recallre	F1-Score
Métricas	0,47	77,27 %	0,50	83,12 %	18,75 %	31,57 %	100 %

Estas imágenes presentan las curvas de pérdida y exactitud del modelo siendo la curva roja la curva de entrenamiento y la verde la curva de evaluación. En estas se puede observar que el modelo presenta un gran sobre entrenamiento debido a la poca cantidad de audios y el desequilibrio de la base de datos. Además, en la tabla 12 se puede observar en las métricas de *precisión* y *recall* que el modelo tiene un buen rendimiento en reconocer audios de personas negativas en COVID-19 teniendo un *recall* muy alto, pero tiene muchos fallos en la detección de positivos lo cual se presenta en la precisión del modelo. Siendo el *F1-Score* la métrica que muestra el rendimiento del modelo con la combinación de la *precisión* y el *recall*.

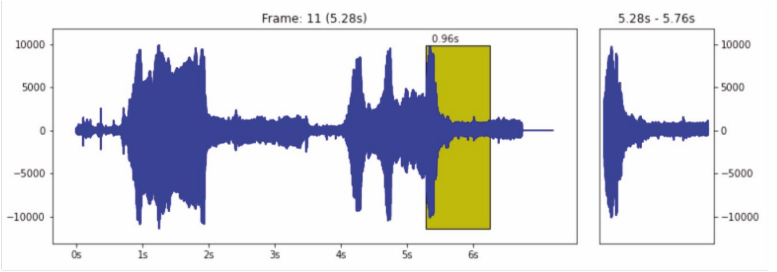
Transfer Learning-YAMNET

Entrenar una red neuronal desde cero requiere una gran base de datos para ayudar al aprendizaje de la red y evitar un sobre entrenamiento en los datos lo cual produce una mala generalización de la red neuronal. Una alternativa tomada por varias investigaciones en estos casos es el uso de una red neuronal que ya haya sido entrenada en otro tipo de datos y usarla para usar ese conocimiento y resolver otra tarea, siendo este método llamado "Transfer-Learning".

Las pruebas iniciales realizadas utilizan una red neuronal llamada "Yamnet" que es una red neuronal entrenada para clasificar 521 clases de sonidos en los que se encuentran desde animales, vehículos hasta el sonido de una tos. Debido a que esta red ya tiene conocimiento de que es una tos, el objetivo principal es usarla para entrenarla

de nuevo y analizar su rendimiento en esta tarea. Esta red neuronal convolucional utiliza ventanas pequeñas de 0,96s las cuales se desplazan por la imagen obteniendo un vector de características por cada una de ellas, las cuales son usadas posteriormente para realizar una clasificación del sonido, esto se presenta en la figura 11.

Figura 11
Yamnet



En los experimentos iniciales se ha utilizado esta red para entrenarla en la base de datos, obteniendo los resultados presentados en la figura 12 y en la tabla 13.

Figura 12
Curvas de pérdida y exactitud

Tabla 13
Resultados

Resultado 4 EPOCH							
	Train		**Devel**				
	Loss	Accuracy	Loss	Accuracy	Precision	Recall	F1-Score
Métricas	0,59	75,17 %	0,65	74,45 %	33,33 %	22,91 %	27,15 %

En la figura 12 se puede observar que ya no se presenta un sobre entrenamiento de la red, esto debido a que no se inició un modelo desde cero. Sin embargo, se pudo observar que el desequilibrio de la base de datos, afecta en el aprendizaje del modelo. En la tabla 13 se puede observar también que a pesar que la red tiene un buen resultado de a*ccuracy* en las predicciones, este dato no representa totalmente los resultados, ya que la base de datos al tener más datos negativos que positivos la red tiende a predecir todos los negativos pero los positivos no los detecta, lo cual produce un buen resultado en *accuracy*,pero resultados muy malos de *precisión* y *recall* que nos expresa los resultados tomando en cuenta los falsos positivos como negativos de las predicciones. La matriz de confusión se presenta en la tabla 14.

Tabla 14
Matriz de confusión

	Matriz de confusión		
Labels	**0**	161	22
	1	37	11
		0	**1**
		Prediction	

En esta matriz se puede observar que el modelo es muy bueno prediciendo casos negativos, pero muy malo en predecir los positivos lo cual es producido por el desequilibrio de la base de datos.

Para resolver este desequilibrio se ha analizado diferentes técnicas, una de ellas es la creación de nuevos datos, la que mejor ha resultado ha sido el aumento de la amplitud del sonido de las muestras positivas hasta 20 dB, con lo cual incrementamos la cantidad de muestra y tratamos de equilibrar la base de datos. Los resultados obtenidos luego de la realización de esta técnica se presentan en la tabla 15 y la matriz de confusión en la tabla 16.

Tabla 15
Resultados-incremento de amplitud

Resultado 4 EPOCH							
	Train		**Devel**				
	Loss	Accuracy	Loss	Accuracy	Precision	Recall	F1-Score
Métricas	0,72	59,94 %	0,70	46,75 %	20,93 %	56,25 %	30,50 %

Tabla 16
Matriz de confusión

	Matriz de confusión		
Labels	**0**	81	102
	1	21	27
		0	**1**
		Prediction	

Transfer Learning VGGish

Luego de realizar pruebas usando la red de "YAMNET", se pudo observar que la red no proporcionaba una gran flexibilidad en su código para realizar las pruebas respectivas; por lo cual se utilizó otra red neuronal llamada VGGish, esta red neuronal está entrenada en una base de datos llamada AUDIOSET la cual posee una clase de tos, por lo que es adecuada para realizar la técnica de *Transfer Learning*, en estas pruebas se han realizado con los parámetros colocados en

la tabla 17. Primero se realizaron las pruebas con la base de datos original y se obtuvo los siguientes resultados presentados en la figura 13 y en las tablas 18 y 19.

Tabla 17
Parámetros

Etapa	Size
VGGish	(batch_size, 128)
Dense_1	64
Dense_2	2

Figura 13
Curvas de métricas

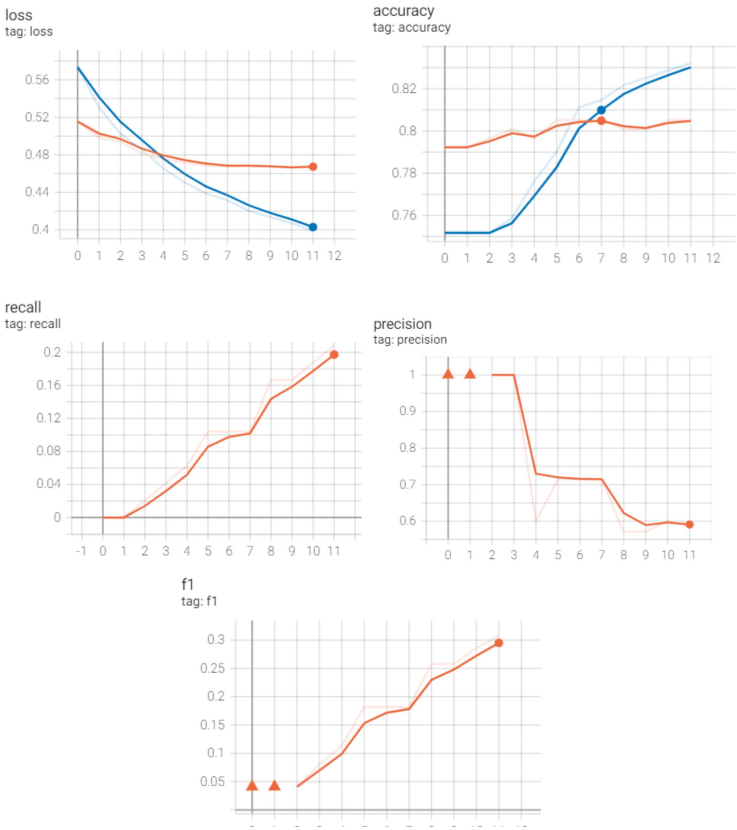


Tabla 18
Métricas

Resultado 12 EPOCH							
	Train		Devel				
	Loss	Accuracy	Loss	Accuracy	Precision	Recall	F1-Score
Métricas	0,39	83,21 %	0,46	80,51 %	58,82 %	20,83 %	30,76 %

Tabla 19
Matriz de confusión

	Matriz de confusión		
Labels	**0**	176	7
	1	38	10
		0	**1**
		Predicción	

En la tabla 19 se puede ver unos resultados mucho más estables y equilibrados que con YAMNET. Luego se procedió a probar técnicas de *Data Augmentation*, para incrementar la clase positiva (casos positivos de COVID-19) y tener una base de datos más equilibrada, estas se muestran a continuación:

a) Base de datos modificada con incremento de amplitud

Se incrementó la amplitud de la clase positiva de 5 a 20 dB, obteniendo un aumento de la clase:

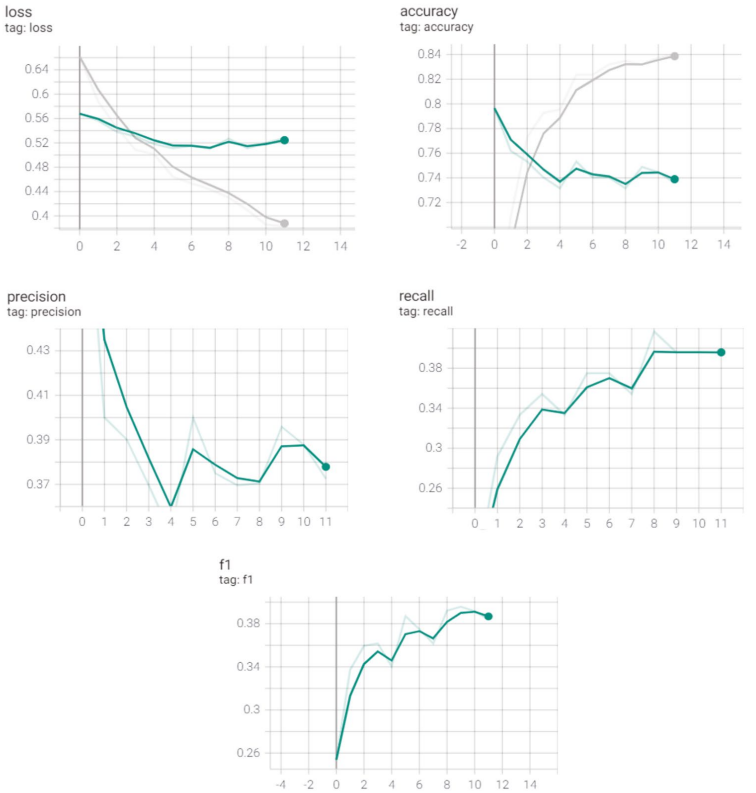
Cantidad de archivos de train: 357, positivos: 142, negativos: 215.

Los intervalos se colocaron empíricamente, 5dB se vio que sea un aumento considerable para que no sea imperceptible para el modelo y 20db dado que se escuchó que a 25dB ya comenzaba a existir un poco de distorsión, entonces se tomó un valor aleatorio de 5 a 20 dB y se incrementó la clase positiva, esto se presenta en la tabla 20, los resultados se presentan en la figura 14 y en las tablas 21 y 22.

Tabla 20
Base de datos

	N. Archivos	N. Positivos	N. Negativos	N. Archivos	N. Positivos	N. Negativos
Train	286	71	215	357	142	215
Devel	231	48	183	231	48	183

Figura 14
Curvas de métricas



La comparación de *F1 score* con resultado previo realizado con la base original se muestra en la figura 15.

Figura 15
Comparación de F1-Score

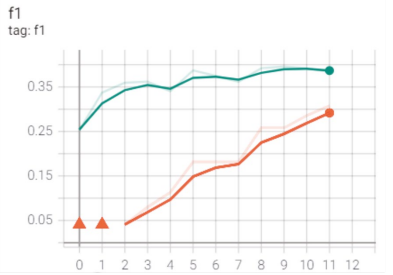


Tabla 21
Métricas

Resultado 10 EPOCH							
	Train		Devel				
	Loss	Accuracy	Loss	Accuracy	Precision	Recall	F1-Score
Métricas	0,40	83,19 %	0,51	74,89 %	39,58 %	39,58 %	39,58 %

Tabla 22
Matriz de confusión

	Matriz de confusión		
Labels	**0**	154	29
	1	29	19
		0	**1**
		Predicción	

En estos resultados se puede observar que se obtiene un muy buen equilibrio el cual se lo expresa en la métrica del *F1-Score*, es decir no se presenta un desbalance en los resultados de las métricas, con lo cual se tienen resultados mejores a los obtenidos con la base de datos original.

b) Base de datos modificada con ruido

En esta técnica se añade ruido blanco a los audios de la clase positiva con el objetivo de crear nuevos audios e incrementar la base de datos original (tabla 20), los resultados se presentan en la figura 16 y en las tablas 23 y 24.

Figura 16
Curvas de métricas

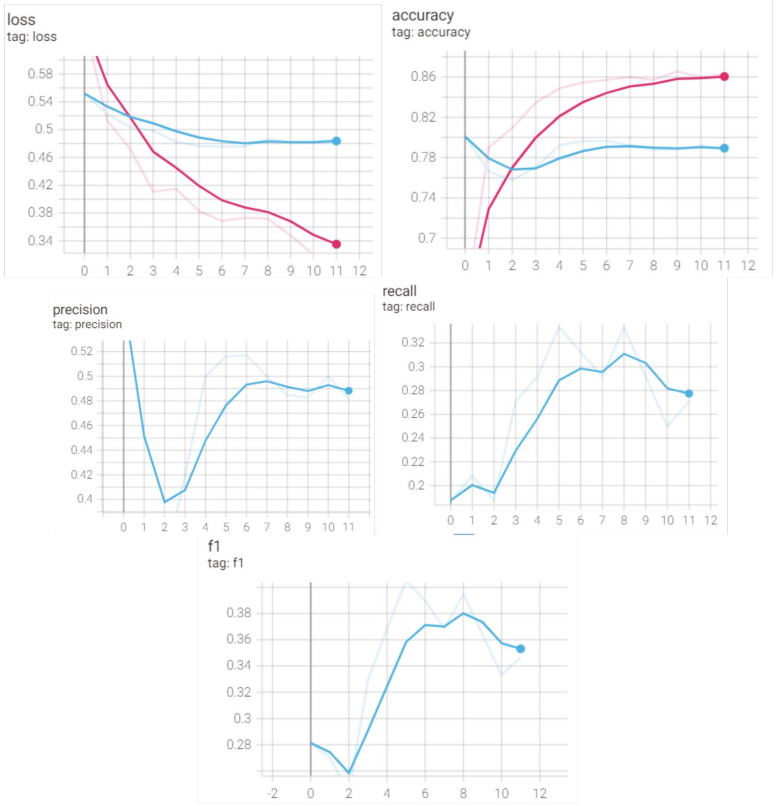


Tabla 23
Métricas

Resultado 9 EPOCH							
	Train		Devel				
	Loss	Accuracy	Loss	Accuracy	Precision	Recall	F1-Score
Métricas	0,37	85,71 %	0,48	78,78 %	48,48 %	33,33 %	39,50 %

Tabla 24
Matriz de confusión

	Matriz de confusión		
Labels	**0**	166	17
	1	32	16
		0	**1**
		Predicción	

Luego de aplicar la técnica de añadir ruido blanco a los audios se presentan resultados muy parecidos a la técnica anterior, pero con un desequilibrio más grande, es decir se presentan resultados mejores de precisión, pero el rendimiento en *recall* es muy bajo.

c) Base de datos modificada con más velocidad

En esta técnica se incrementa la velocidad de los audios de la clase positiva y se analiza si mejora los resultados obtenidos con las técnicas anteriores. Partiendo de los datos originales de la tabla 20, los resultados se presentan en la figura 17 y las tablas 25 y 26.

Figura 17
Curvas de métricas

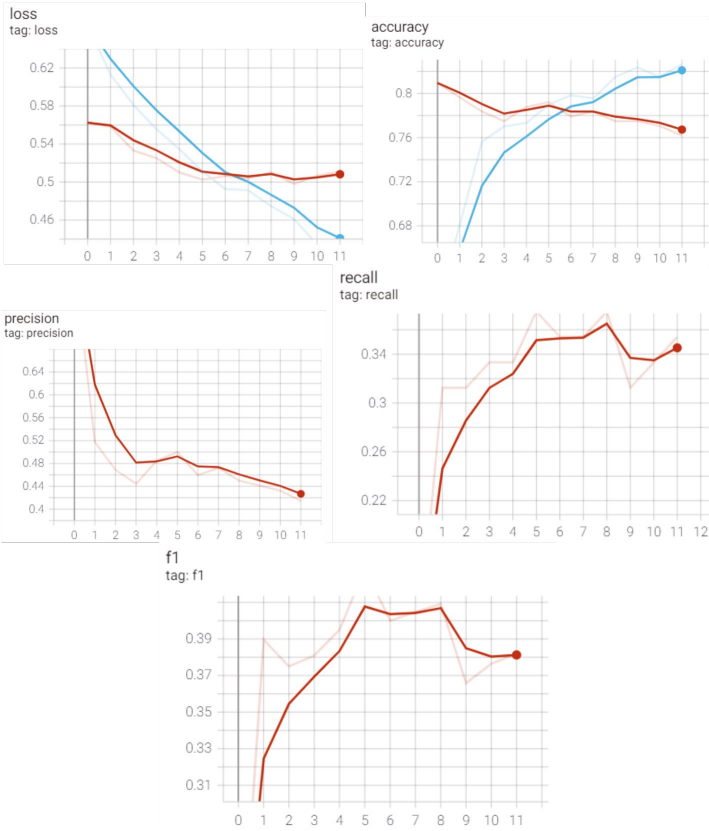


Tabla 25
Métricas

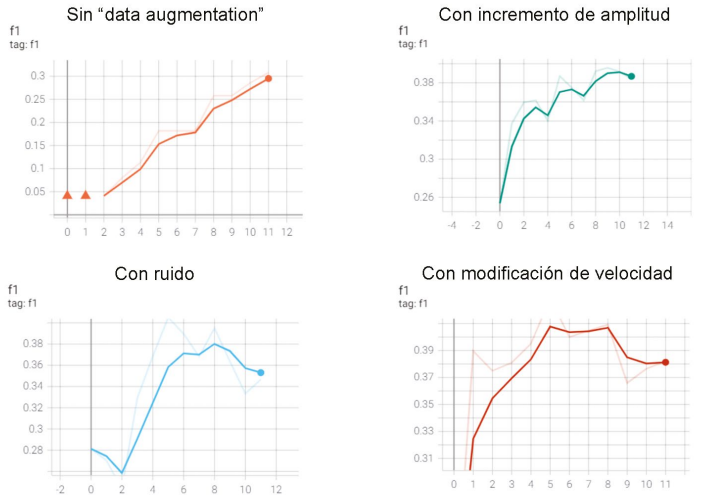
Resultado 9 EPOCH							
	Train		Devel				
	Loss	Accuracy	Loss	Accuracy	Precision	Recall	F1-Score
Métricas	0,47	81,51 %	0,51	77,48 %	45 %	37,5 %	40,9 %

Tabla 26
Matriz de confusión

	Matriz de confusión		
Labels	**0**	161	22
	1	30	18
		0	**1**
		Predicción	

Como se puede observar con esta técnica de igual manera se obtienen rendimientos similares a los resultados obtenidos con las técnicas anteriores, y de igual manera se presenta un mayor desbalance en *precisión* y *recall*. Una vez probado diferentes técnicas se procede a comparar en la tabla 27.

Tabla 27
Matriz de comparación



Como se puede observar todas las técnicas de "*Data Augmentation*" ayudan a mejorar el rendimiento del sistema al compararlo con el sistema inicial que utiliza la base de datos original. Sin embargo, al

compararlos entre ellos se observa que estos llegan a un rendimiento muy similar. Por último, se combinan técnicas para analizar si complementan la información.

d) Base de datos modificada con ruido e incremento de amplitud

Se procedió a unir dos técnicas y analizar el rendimiento del modelo para observar si llegan a complementarse, los resultados se presentan en la figura 18 y en las tablas 28 y 29.

Figura 18
Curvas de métricas

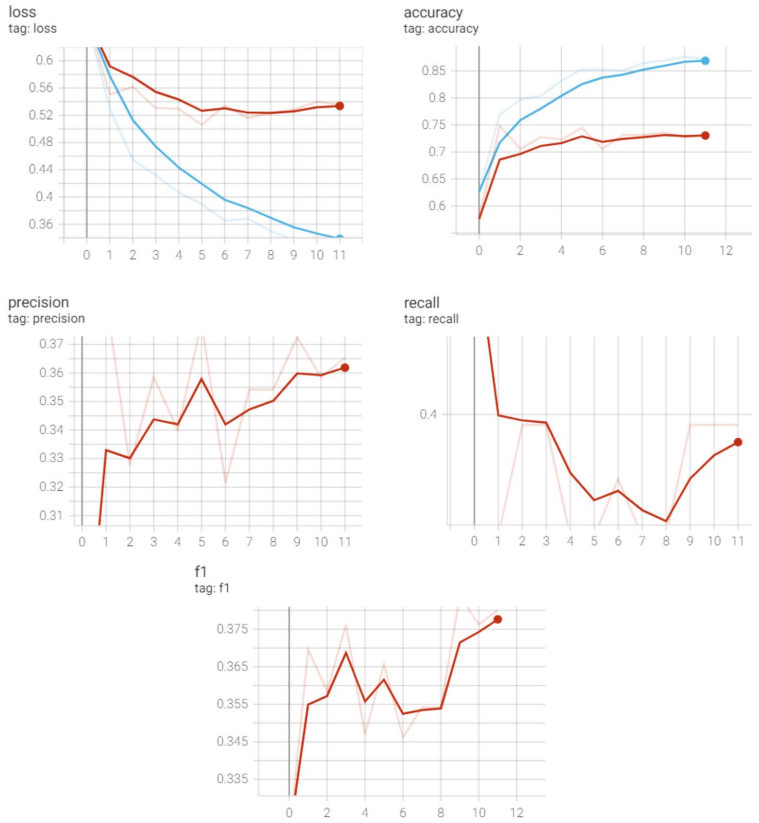


Tabla 28
Métricas

Resultado 10 EPOCH							
	Train		Devel				
	Loss	Accuracy	Loss	Accuracy	Precision	Recall	F1-Score
Métricas	0,33	86,91 %	0,52	73,59 %	37,25 %	39,58 %	38,38 %

Tabla 29
Matriz de confusión

	Matriz de confusión		
Labels	**0**	151	32
	1	29	19
		0	**1**
		Predicción	

e) Base de datos modificada con incremento y menos velocidad

Los resultados de esta combinación se presentan en la figura 19 y en las tablas 30 y 31.

Figura 19

Curvas de métricas

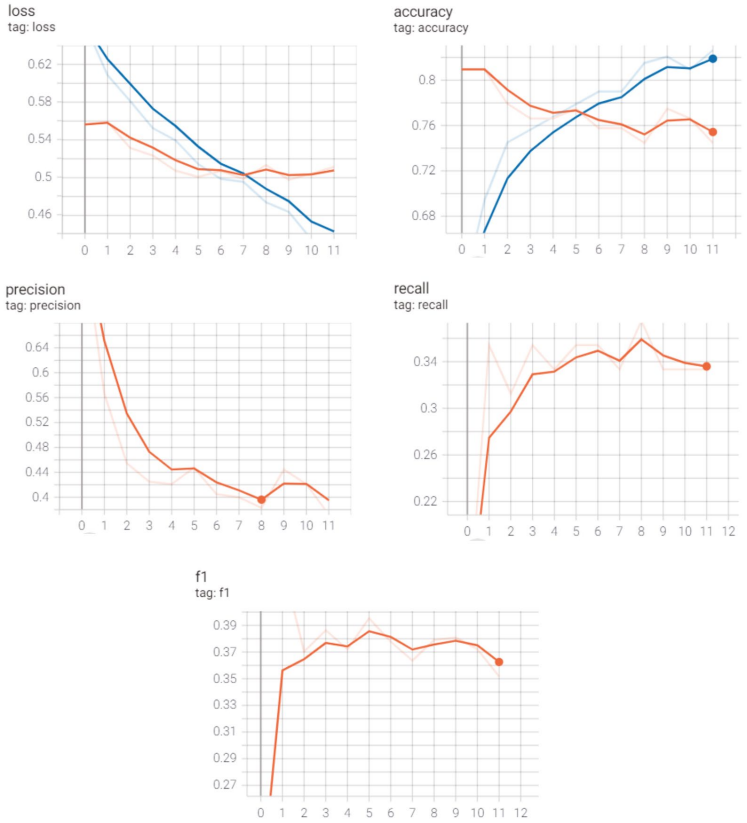


Tabla 30

Métricas

Resultado 4 EPOCH							
	Train		Devel				
	Loss	Accuracy	Loss	Accuracy	Precision	Recall	F1-Score
Métricas	0,55	74,78 %	0,52	77,92 %	46,15 %	37,5 %	41,37 %

Tabla 31
Matriz de confusión

	Matriz de confusión		
Labels	**0**	162	21
	1	30	18
		0	**1**
		Predicción	

f) Base de datos modificada con incremento de Amplitud y Modificación de velocidad

Los resultados de esta combinación se muestran en la figura 20 y en las tablas 32 y 33.

Figura 20
Curvas de métricas

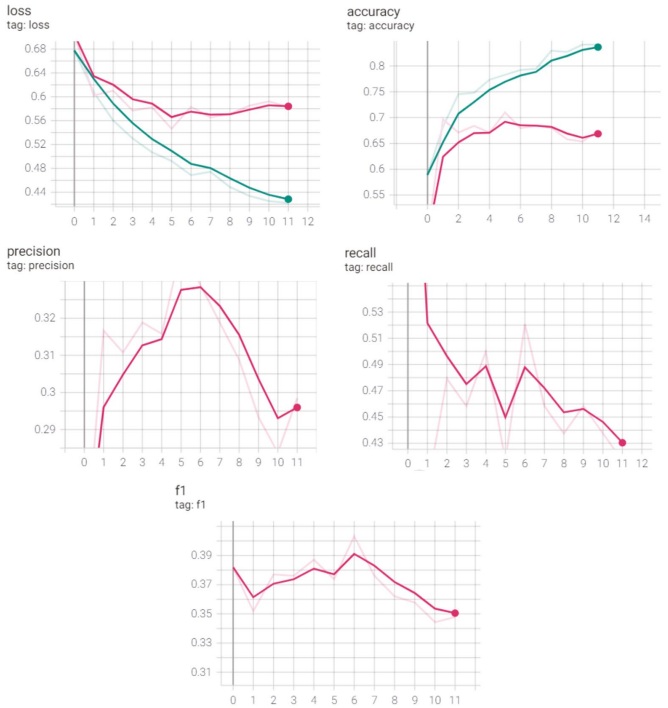


Tabla 32
Métricas

Resultado 2 EPOCH							
	Train		Devel				
	Loss	Accuracy	Loss	Accuracy	Precision	Recall	F1-Score
Métricas	0,60	68,22 %	0,60	69,69 %	31,66 %	38,58 %	35,18 %

Tabla 33
Matriz de confusión

	Matriz de confusión		
Labels	**0**	142	41
	1	29	19
		0	**1**
		Predicción	

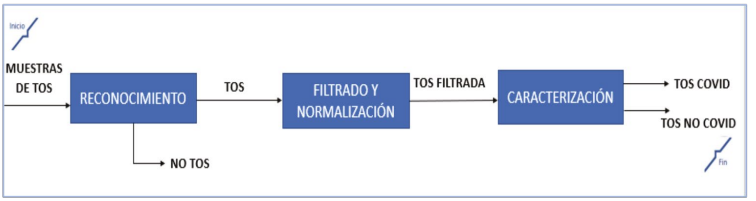
Juntando las técnicas de *Data Augmentation,* estos muestran que se presentan prácticamente los mismos resultados que las técnicas individuales, es decir, no se muestra que exista un complemento de información entre ellas, al contrario, esto muestra que hay un límite de rendimiento el cual parece dejar de ser producido por la cantidad de data y es provocado por un límite de conocimiento del modelo.

Diseño del sistema integrado de reconocimiento de una señal de tos COVID-19

Procedimiento

En la figura 21 se muestra el Sistema Integrado de reconocimiento de la tos en las señales de audio, desde la recopilación de las muestras de tos, hasta la caracterización de la tos COVID-19 y no COVID-19.

Figura 21
Sistema Integrado de Reconocimiento



A continuación, se detalla el procedimiento del sistema integrado de reconocimiento y caracterización de la tos en las señales de audio:

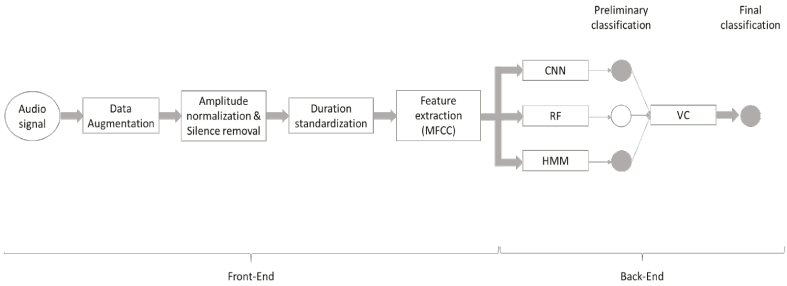
Inicialmente, en la fase del reconocimiento, se ha abordado la problemática sobre la detección de tos en señales de audios, donde se ha encontrado que un monitor de tos portátil permitiría un seguimiento continuo en tiempo real de pacientes con enfermedades respiratorias. Esta forma de seguimiento de la tos podría permitir captar información fiable para los profesionales, que a la vez potenciaría la telemedicina en enfermedades respiratorias. Además, desde el punto de vista del paciente, este sistema de monitorización sería más cómodo y provocaría una mínima interrupción en sus actividades diarias. Sin embargo, el desarrollo de monitores portátiles implica la implementación de sistemas que puedan identificar y discriminar la tos de cualquier otro evento sonoro registrado en entornos potencialmente ruidosos.

Con base en la literatura, los investigadores han encontrado técnicas, que se han implementado con éxito en monitores de tos portátiles. Así, por ejemplo, en el campo del aprendizaje automático, Khunarsal *et al.* (2013) implementaron un algoritmo para la clasificación del sonido ambiental utilizando espectrogramas y clasificadores K-Vecinos-más cercanos (KNN). Por otro lado, Matos *et al.* (2006) implementaron modelos ocultos de Markov para detectar señales de tos de grabaciones de audio continuas. Sin embargo, los resultados más prometedores se han logrado mediante la implementación de algoritmos de aprendizaje profundo, especialmente utilizando redes

neuronales convolucionales (CNN), como en el trabajo realizado por Amoh y Odame (2016) quienes implementaron efectivamente una CNN para identificar los sonidos de la tos.

Figura 22

Arquitectura del Sistema de Reconocimiento



Dentro del proyecto, luego de la implementación de clasificadores como las CNN, Random Forest, y cadenas ocultas de Markov, en esta fase, se aplicó la técnica del Clasificador de Votación, como se muestra en la figura 22, el cual permitió crear un clasificador con una mayor precisión. Para el cual, el modelo agrega las predicciones de cada clasificador y predice la clase que obtiene la mayor cantidad de votos. Este clasificador de votos por mayoría se llama clasificador de votación dura, logrando así, una precisión mayor que el mejor clasificador del conjunto.

Por tal motivo se decidió ensamblar tres modelos de aprendizaje automático para diferenciar los sonidos de tos de entre otros sonidos ambientales. Previamente se simplificó el pre-procesamiento de las señales de audio para que, en lugar de tomar 67 componentes de los espectrogramas de Mel, se consideren únicamente 16. Las características de los tres modelos entrenados fueron: una red neuronal convolucional con cinco capas, divididas en dos capas convolucionales, dos capas completamente conectadas y una capa de clasificación binaria, implementada con la biblioteca Keras; un modelo de Random Forest con los parámetros por defecto del clasificador de la biblioteca Sklearn; y un clasificador con base en modelos ocultos de Markov gaussianos

de la biblioteca Hmmlearn. Los tres modelos se ensamblaron en un algoritmo *hard voting classifier*, cuyos resultados se basan en el valor de la moda de las predicciones de cada uno de los modelos que lo constituyen. Los modelos se entrenaron con la base de datos de sonidos de tos del proyecto, los sonidos de tos obtenidos mediante *data augmentation* de la base de datos previamente mencionada, y con las grabaciones de tos de la base de datos de Cambridge. Los sonidos de eventos sonoros diferentes a la tos fueron tomados de la base de datos de ruidos ambientales. En cada uno de los modelos, se determinó la *exactitud de clasificación*, la *precisión*, el *recall* y el valor de *F1-score*.

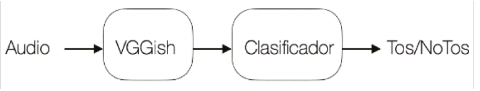
Posteriormente, en la fase de filtrado y normalización, se abordó la problemática del ruido presente en las señales de audio de tos durante una grabación, así como también la implementación de técnicas de filtrado para minimizar el ruido presente en dicho audio.

Por lo tanto, es necesario filtrar estas señales ruidosas para un análisis adecuado y preciso de las señales de tos Shankar *et al.* (2020). En los años anteriores se han utilizado diversas metodologías de filtrado para el particular, y las señales relacionadas con las vías respiratorias incluyen la eliminación de ruido generalmente con filtros de paso bajo Aggarwal *et al.* (2011). Se han propuesto diversas técnicas de modelado, como la descomposición en modo empírico Liang *et al.* (2005), y Blanco-Velasco *et al.* (2008) y vectores de estado con retardo de tiempo Liang *et al.* (2005), filtros de Kalman *et al.* (2010), algoritmo de desplazamiento medio, etc. (Shamsollahi, 2008).

La fase final tiene que ver con la caracterización de la señal de tos separada y filtrada, para ello se evaluaron tres modelos: una red neuronal convolucional nativa, el modelo Yamnet y el modelo VGGish.

El modelo VGGish obtiene las características de los audios de tos en forma de vectores, los cuales son luego usados como entrada a una red que clasifica los mismos. En la figura 23 se presenta un esquema general del sistema.

Figura 23
Arquitectura del Sistema de Caracterización



La red VGG es una red neuronal convolucional que fue entrenada en la base de datos de AUDIOSET, la cual posee aproximadamente 640 clases donde una de ellas es una clase que contiene audios de tos, lo cual la hace adecuada para utilizarla en la tarea de la detección del COVID-19 mediante la técnica de Transfer Learning.

Con base en la programación y desarrollo del algoritmo, y a los mejores resultados obtenidos, luego de cargar el conjunto de datos, se aplicó la técnica de data aumenting, incrementando solo el parámetro de la amplitud de las muestras de tos.

Finalmente se muestran en cada época los resultados de: pérdida, exactitud tanto de entrenamiento como del conjunto de datos de ajuste, la matriz de confusión y las métricas de rendimiento como la precisión, recall y f1-score Todas se definen a continuación para una mejor comprensión:

Métricas de desempeño del Sistema Integrado

La evaluación del desempeño del modelo se basó en las siguientes métricas:

Pérdida de entropía cruzada binaria: describe la pérdida entre dos distribuciones de probabilidad, utilizando una penalización logarítmica que genera una puntuación grande para diferencias grandes y una puntuación pequeña para diferencias pequeñas. La pérdida de entropía cruzada binaria se utiliza al ajustar los pesos del modelo durante el entrenamiento. El objetivo es minimizar la pérdida; es decir, a menor pérdida, mejor modelo (Ramos *et al.,* 2018).

Matriz de confusión y sus métricas

La matriz de confusión es una herramienta que nos permite visualizar el desempeño de un algoritmo que se emplea en el aprendizaje supervisado. Uno de los beneficios de las matrices de confusión es que facilitan ver si el sistema está confundiendo dos clases, como se muestra en la figura 24:

Figura 24
Estructura de la Matriz de confusión

		Predicción	
		Positivos	**Negativos**
Observación	**Positivos**	Verdaderos Positivos (VP)	Falsos Negativos (FN)
	Negativos	Falsos Positivos (FP)	Verdaderos Negativos (VN)

- **VP** es la cantidad de *positivos* que fueron *clasificados correctamente* como positivos por el modelo.
- **VN** es la cantidad de *negativos* que fueron *clasificados correctamente* como negativos por el modelo.
- **FN** es la cantidad de *positivos* que fueron *clasificados incorrectamente* como negativos.
- **FP** es la cantidad de *negativos* que fueron *clasificados incorrectamente* como positivos.

Accuracy (Exactitud): se usa como una medida estadística, para saber que tan bien una prueba de clasificación binaria identifica o excluye correctamente una condición. También se puede interpretar como la proporción de los resultados verdaderos, tanto positivos verdaderos como negativos verdaderos, entre el número total de casos examinados (Deng *et al.*, 2016; Juba y Le, 2019).

$$Exactitud = \frac{VP + VN}{Total}$$

Recall: debe ser lo más alto posible, se interpreta como la proporción del pronóstico de todas las clases positivas, es decir cuánto se pronosticó correctamente, se la interpreta como la sensibilidad, exhaustividad, o simplemente tasa de verdaderos positivos (Juba y Le 2019).

$$\text{Sensibilidad} = \frac{VP}{\text{Total Positivos}}$$

Especificidad: indica la tasa de verdaderos negativos, es decir, el porcentaje de clasificación, cuando la clase es negativa.

$$\text{Especificidad} = \frac{VN}{\text{Total Negativos}}$$

Precisión: la proporción de todas las clases, es decir, cuanto se pronostica correctamente, de igual manera debe ser lo más alto posible (Juba y Le 2019).

$$\text{Precisión} = \frac{VP}{\text{Total clasificados positivos}}$$

F-measure (F1 score): es difícil comparar dos modelos con alta recall y baja precisión o viceversa, resumiendo la precisión y sensibilidad en una sola métrica. Es aquí donde se usa el F1-score ya que ayuda a medir el recall y precisión al mismo tiempo. Utiliza la media armónica en vez de la media aritmética para castigar más los valores extremos. Por ello es de gran utilidad cuando la distribución de las clases es desigual, por ejemplo, cuando el número de pacientes con una condición es del 15 % y el otro es 85 %, lo que en el campo de la salud es bastante común (Juba y Le, 2019; Chicco y Jurman, 2020).

$$\text{F1} = 2x\,\frac{Precision*Recall}{Precision+Recall}$$

Por lo cual, las métricas a determinar sirven para seleccionar el mejor modelo en función del sistema dinámico a analizar. De las características expuestas, conforme a las métricas mencionadas, se puede obtener cuatro casos posibles para cada clase (Barrios, 2019):

- Alta precisión y alto recall: el modelo de Machine Learning escogido maneja perfectamente esa clase.
- Alta precisión y bajo recall: el modelo de Machine Learning escogido no detecta la clase muy bien, pero cuando lo hace es altamente confiable.

- Baja precisión y alto recall: El modelo de Machine Learning escogido detecta bien la clase, pero también incluye muestras de la otra clase.
- Baja precisión y bajo recall: El modelo de Machine Learning escogido no logra clasificar la clase correctamente.

Resultados

La tabla 34 muestra una comparación de las métricas de desempeño de los modelos entrenados. Observamos que los modelos CNN y RF lograron los valores de precisión más altos y el modelo HMM presenta el valor de *recall* o *sensibilidad* más alto. Mientras que el modelo VC, al estar ensamblado con los tres anteriores, logró las métricas de rendimiento más altas, por lo que seleccionó el modelo más adecuado para identificar los sonidos de la tos, dentro del sistema integrado de reconocimiento y caracterización de la tos.

Tabla 34
Resultados del desempeño de los modelos entrenados

Métrica	CNN	RF	HMM	VC
Exactitud	0,942	0,946	0,813	0,950
Precisión	0,964	1,000	0,574	1,000
Recall-Sensibilidad	0,794	0,779	0,911	0,794
F1- score	0,871	0,876	0,705	0,885

Con la metodología implementada del uso del filtrado adaptativo, se creó un sistema que puede minimizar el ruido en un archivo de registro de tos. Se definió un "patrón de ruido" que puede contener la información de todo tipo de ruidos que contaminan una señal de tos récord. Para eso se propuso considerar el ruido a la información del registro previo a una expectoración de tos, además, se recopilaron todos los ruidos de la base de datos, se dividieron las muestras de los extraños y finalmente definimos el patrón de ruido.

Como resultado de la fase del filtrado se logró minimizar el ruido en un archivo de registro de tos cercano a 0 dB, y para verificar la generalidad de la metodología propuesta, se probó en todos los conjuntos de datos, y se verificaron que todos los archivos de registros comparten el mismo valor SNR cercano a 0 dB en la base de datos testeada.

En la fase de la caracterización, una vez que ingresaron los datos filtrados al modelo seleccionado, por la arquitectura VGGish, con un batch_size de 128, una Dense_1 de 64 y dense_2 de 2.

Los resultados en función de la obtención del mejor funcionamiento, y con las métricas más altas en el rendimiento del modelo fueron los siguientes:

Se incrementó la amplitud de la clase positiva de 5 a 20 dB., obteniendo un aumento de la clase: cantidad de archivos de datos de entrenamiento: 357, positivos: 142, negativos: 215.

Los intervalos se colocaron empíricamente, 5 dB se vio que sea un aumento considerable para que no sea imperceptible para el modelo y 20 dB, dado que se escuchó que a 25d B ya comenzaba a existir un poco de distorsión, entonces se tomó un valor aleatorio de 5 a 20 y se incrementó la clase positiva.

Obteniendo así, el resultado del sistema integrado de reconocimiento, en la iteración o "Epoch" 10, con una *pérdida* de 0,40, *exactitud* de 83,19 %, *perdida de prueba*: 0,51, y *exactitud de prueba*: 74,0260, p*recisión*: 0,3958, *recall*: 0,3958, *f1*: 0,3958. Así, las métricas de rendimiento se muestran en la tabla 35.

Tabla 35
Métricas del modelo de clasificación

Resultado 10 EPOCH				
	Entrenamiento		**Prueba**	
	Perdida	**Exactitud**	**Perdida**	**Exactitud**
Métricas de rendimiento	0,40	83,19 %	0,51	74,89 %

Y, los resultados en cuanto a la predicción, se muestra la matriz de confusión en la tabla 36.

Tabla 36
Matriz de confusión

	Matriz de confusión		
Labels	**0**	154	29
	1	29	29
		0	**1**
		Predicción	

En estos resultados se puede observar que se obtiene un muy buen equilibrio el cual se lo expresa en la métrica del F1-Score, es decir no se presenta un desbalance en los resultados de las métricas, con lo cual se tienen resultados mejores a los obtenidos con la base de datos original.

Finalmente, el sistema integrado de caracterización de la tos, inicialmente se desarrolló con la arquitectura LeNet-5, que funciona bien en pequeños conjuntos de datos y una red neuronal que fue entrenada especialmente para distinguir la tos, Posteriormente, la señal de audio analizado es cortado en trozos de 1 segundo y estos fragmentos son evaluados en esta etapa de detección, para descartar los fragmentos que no tienen presencia de tos. Logrando una precisión del 100 % determinando qué fragmentos tienen tos y cuáles tienen ruido, para pasar a la siguiente etapa, que es evaluar los fragmentos restantes, para determinar si el audio coincide con el patrón de un individuo que es potencialmente COVID-19 positivo o no, con una arquitectura VGGish, obteniendo una exactitud del 74,89 %.

Referencias bibliográficas

Amoh, J. y Odame, K. (2016). Deep neural networks for identifying cough sounds. *IEEE Transactions on Biomedical Circuits and Systems, 10*(5), 1003-1011. https://doi.org/10.1109/TBCAS.2016.2598794

Aggarwal, R., Singh, J. K., Gupta, V. K., Rathore, S., Tiwari, M. y Khare, A. (2011). Noise reduction of speech signal using wavelet transform with modified universal threshold. *International Journal of Computer Applications, 20*(5), 14-19.

Amrulloh, Y. A., Abeyratne, U. R., Swarnkar, V., Triasih, R. y Setyati, A. (2015). Automatic cough segmentation from non-contact sound recordings in pediatric wards. *Biomedical Signal Processing and Control, 21,* 126-136.

Brown, C., Chauhan, J., Grammenos, A., Han, J., Hasthanasombat, A., Spathis, D., Xia, T., Cicuta, P. y Mascolo, C. (2020). Exploring automatic diagnosis of COVID-19 from crowdsourced respiratory sound data. arXiv preprint arXiv:2006.05919.

Barrios, J. I. (26 julio, 2019). BIG DATA, Ciencia de datos, Informática, Médica, Inteligencia Artificial, machine learning,

Blanco-Velasco, M., Weng, B. y Barner, K. E. (2008). ECG signal denoising and baseline wander correction based on the empirical mode decomposition. *Computers in Biology and Medicine, 38*(1), 1-13.

Chicco, D. y Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics, 21*(1), 1-14. https://doi.org/10.1186/s12864-019-6413-7

Deng, X., Liu, Q., Deng, Y. y Mahadevan, S. (2016). An improved method to construct basic probability assignment based on the confusion matrix for classification problem. *Inf. Sci., 340-341*, 250-261. https://doi.org/10.1016/j.ins.2016.01.033

Imran, A., Posokhova, I., Qureshi, H. N., Masood, U., Riaz, M. S., Ali, K., John, Ch., Hussain, I. y Nabeel, M. (2020). AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app. *Informatics in Medicine Unlocked*, *20*, 100378.

Juba, B. y Le, H. S. (2019). Precision-recall versus accuracy and the role of large data sets. *Proc. AAAI Conf. Artif. Intell., 33,* 4039-4048. https://doi.org/10.1609/aaai.v33i01.33014039

Khunarsal, P., Lursinsap, C. y Raicharoen, T. (2013). Very short time environmental sound classification based on spectrogram pattern matching. *Information Sciences, 243*, 57-74. https://doi.org/10.1016/j.ins.2013.04.014

Laguarta, J., Hueto, F. y Subirana, B. (2020). COVID-19 artificial intelligence diagnosis using only cough recordings. *IEEE Open Journal of Engineering in Medicine and Biology, 1*, 275-281.

Liang, H., Lin, Z. y Yin, F. (2005a). Removal of ECG contamination from diaphragmatic EMG by nonlinear filtering. *Nonlinear Analysis: Theory, Methods & Applications, 63*(5-7), 745-753.

Liang, H., Lin, Q.-H. y Chen, J. D. Z. (2005b). Application of the empirical mode decomposition to the analysis of esophageal manometric data in gastroesophageal reflux disease. *IEEE Transactions on Biomedical Engineering, 52*(10), 1692-1701.

Matos, S., Birring, S. S., Pavord, I. D. y Evans, D. H. (2006). Detection of cough signals in continuous audio recordings using hidden Markov models. *IEEE Transactions on Biomedical Engineering, 53*(6), 1078-1083. https://doi.org/10.1109/TBME.2006.873548

Poornachandra, S. (2008). Wavelet-based denoising using subband dependent threshold for ECG signals. *Digital Signal Processing, 18*(1), 49-55.

Pramono, R. X. A., Imtiaz, S. A. y Rodriguez-Villegas, E. (2016). A cough-based algorithm for automatic diagnosis of pertussis. *PloS one*, *11*(9), e0162128.

Ramos, D., Franco-Pedroso, J., Lozano-Diez, A. y Gonzalez-Rodriguez, J. (2018). Deconstructing cross-entropy for probabilistic binary classifiers. *Entropy, 20*(3), 1-20. https://doi.org/10.3390/e20030208

Shamsollahi, M. B. (2008). ECG denoising and compression using a modified extended Kalman filter structure. *IEEE Transactions on Biomedical Engineering, 55*(9), 2240–2248.

Shankar, A., Bhateja, V., Srivastava, A. y Taquee, A. (2020). Continuous wavelets for pre-processing and analysis of cough signals. En *Smart Intelligent Computing and Applications* (pp. 711-718). Springer.

Yan, J., Lu, Y., Liu, J., Wu, X. y Xu, Y. (2010). Self-adaptive model-based ECG denoising using features extracted by mean shift algorithm. *Biomedical Signal Processing and Control, 5*(2), 103-113.

Capítulo 6

Descripción del sistema integrado

Introducción

Durante el desarrollo del proyecto, en todo momento se trabajó con dos algoritmos por separado, uno donde se realizaron las pruebas para la etapa Front-End y otro donde se realizaron las pruebas para el Back-End. El paso previo a la integración del sistema fue la adaptación de ambos scripts de tal manera que puedan funcionar como un solo sistema integrado. El sistema final debe funcionar íntegramente ser integrado en Google Colaboratory (Colab) y por ello, es necesario que ambos códigos estén en un formato de producción y ya no en formato de desarrollo, por lo que, únicamente debe contener el código fuente necesario para realizar las tareas particulares de cada etapa.

El script del Front-End fue desarrollado desde un inicio en Google Colaborative (Colab), sin embargo, como la finalidad de este fue realizar pruebas, desarrollos y entrenamientos, es necesario adaptar el código y adecuarlo de tal forma que el script solo contenga el código de interés. De esta forma, el script del Front-End lee una carpeta que contiene todos los audios, los procesa y al final entrega una lista con todos los audios que contienen tos. En este punto ya no se realizan

entrenamientos de modelos ni validaciones, pues eso es necesario solo en la fase de desarrollo y no en la fase de producción.

Por otra parte, el Back-End sí que fue desarrollado a nivel local y por ello fue necesario replicar todas las configuraciones que se realizaron a nivel local y llevarlas al Colab. Una vez que se verificó que los experimentos eran reproductibles, tanto a nivel local como en el Colab, se procedió a adecuar el código para la fase de producción. Cabe destacar que, durante el proceso de adaptación, se verificó que las métricas sean las mismas que se obtuvieron durante la fase de pruebas y desarrollo en ambas etapas del sistema.

Metodología

El proyecto busca desarrollar un sistema que permita caracterizar y clasificar la señal de tos provocado por el Covid-19, por tanto, el sistema propuesto se componte de dos etapas: un Front-End y un Back-End. A continuación se describen las dos etapas del sistema:

Front End

Como se ha mencionado en capítulos anteriores, en el Front End, el objetivo es filtrar o seleccionar a todos los audios que contengan tos y para ello, los audios son procesados para extraer las características más relevantes que permitan identificar únicamente los audios que contienen tos. De esta manera, los audios que contengan voz, ruido y otros sonidos ambientales no serán enviados al Back-End.

El Front-End está compuesto por seis pasos:

a) Eliminación de silencios: con este paso inicial se busca eliminar todos los lapsos de silencio que contiene el audio, esto es necesario ya que, estos lapsos no contienen información relevante y de no eliminarlos, se perdería tiempo procesando información que no tiene ninguna relevancia en la caracterización de la tos y disminuiría en cierta medida el rendimiento general del sistema. Por lo que,

en este primer paso, el o los audios quedan recortados y contienen únicamente lapsos de tiempo donde existen un sonido (puede ser cualquier sonido tos, voz, ambiente, etc.).

b) Uniformización de la duración: durante las pruebas de los modelos para clasificar las señales de audio de tos y no tos, el modelo que mejor rendimiento tuvo fue el de redes neuronales convolucionales (CNN), este modelo fue entrenado para que reciba una imagen del espectrograma de un audio y a su salida identifique si se trata o no de un audio que contiene tos. El modelo se entrena con un tamaño específico de imagen del espectrograma y debe siempre recibir el mismo tamaño de imagen del espectrograma. Los audios no tienen ninguna restricción en cuanto a su duración, por lo que, es necesario realizar un proceso para uniformizar la duración sin que se pierda información. En este sentido existen dos enfoques: se puede utilizar padding para completar las imágenes de los espectrogramas y que siempre tengan el mismo tamaño o se puede colocar una duración específica y que todos los audios se recorten a dicha duración. Si se utiliza padding en los audios de muy corta duración, el espectrograma puede contener en su mayor parte información de padding y no información de la tos, por lo que, se optó por el segundo enfoque utilizando una duración fija de 2 segundos. Durante la uniformización de la duración existen tres casos dependiendo de si la duración del audio es:

- Igual a 2 segundos: no se realiza ninguna acción.
- Menor a 2 segundos: se repite el audio hasta completar los 2 segundos.
- Mayor a 2 segundos: se divide el audio en audios de duración fija de 2 segundos, si la duración es un múltiplo de 2, el audio se divide en varios audios de 2 segundos que sumados completan la duración del audio inicial. Si la duración es impar, se divide en audios de 2 segundos y el segmento de

tiempo restante se completa repitiendo el mismo segmento hasta completar los 2 segundos.

En el tercer caso se tendrían varios audios de 2 segundos en función de la duración del audio inicial, dado que todos estos audios son parte de un único audio, además de uniformizar la duración también se realiza un etiquetado considerando el audio original. De esta forma, todas las subdivisiones de un mismo audio tendrán la misma etiqueta y se clasificarán en tos y no tos como si fueran audios independientes. Luego, utilizando un algoritmo de votación (que se explica en el último paso) y la etiqueta de los audios iniciales, se define que audios originales tienen tos y cuales no la tienen.

c) Espectrograma en Escala de Mel: una vez se tiene audios de duración uniforme, lo siguiente es extraer sus características más relevantes, para ello se utiliza el espectrograma en escala de Mel de cada audio, que es una representación visual del audio de la señal de tos y permite identificar las variaciones de la frecuencia y la intensidad del sonido en el tiempo.

d) Modelo de clasificación CNN: el modelo de clasificación inicial fue entrenado considerando que los audios tienen una frecuencia de muestreo de 48 kHz, sin embargo, en toda la base de datos se tiene audios de 16 kHz y de 48 kHz. Por lo que, se podría procesar los audios de 16 kHz como si fueran de 48 kHz, pero esto afectaría a la duración y causaría que en todos los audios de 16 kHz la duración se reduzca al menos tres veces. Por lo que, se realizó lo inverso todos los audios fueron procesados considerando que tienen una frecuencia de muestreo de 16 kHz. En este caso, en los audios de 48 kHz también se cambia su duración, se va a incrementar, pero esto no afecta de forma significativa el audio, sino que únicamente se dilata en parte la señal de audio, pero como tiene más muestras por segundo, no se pierde información. Cabe destacar que se tenía inicialmente entrenado un modelo que procesa todos los audios como si fueran de 48 kHz, por lo que se volvió a entrenar un nuevo modelo para 16 kHz.

e) Clasificación: en este paso, utilizando el modelo entrenado para 16kHz, los espectrogramas en escala de Mel de los audios se ingresan como entrada y a la salida del modelo se determina que audios contienen tos (1) y cuales no contienen tos (0).

f) Algoritmo de votación: una vez que los audios uniformizados se clasificación (0-1), el último paso es utilizar la etiqueta de cada audio para determinar que audio original contiene tos y cual no contiene tos:

- Si se tienen dos audios con la misma etiqueta, y uno de los audios fue clasificado con 1 o los dos tiene una clasificación de 1, el audio original tiene tos. Caso contrario, no tiene tos.
- Si se tienen más de dos audios con la misma etiqueta, se considera la clasificación de la mayoría: si al menos dos audios tienen una clasificación de 1, el audio original tiene tos, si dos audios tienen una clasifican de 0, el audio original no tiene tos. Es decir, se considera la clasificación que tenga la mayoría de los audios uniformizados que pertenezcan a un mismo audio original.

Finalmente, se genera una lista con el directorio y el nombre de todos los audios originales que contienen tos y esta lista se envía al Back-End para la caracterización de la señal de la tos.

Back End

En el Back End se busca caracterizar la señal de la tos para distinguir características particulares de la tos causada por COVID-19. En esta etapa se recibe la lista de audios que contienen tos que fueron seleccionados en el Front-End. El Back-End está compuesto por dos pasos:

a) VGGish: en el primer paso se utiliza el modelo pre entrenado VGGish como extractor de características de los audios. VGGish es una red neuronal convolucional que fue entrenada con la base

de datos de AUDIOSET, la cual posee alrededor de 640 clases de audios y una de ellas es la tos. Por lo que, VGGish es adecuado para obtener los vectores de características de los audios con tos. Así, en este primer paso, los audios que se obtienen del Front-End son convertidos en vectores de características.

b) Clasificador: los vectores de características de los audios que se obtienen en el primer paso son la entrada de una red neuronal multicapa (MLP) que se encarga de identificar los audios que tienen características de una tos provocada por el COVID-19.

Resultados finales

Para verificar el rendimiento del sistema para la caracterización de la tos —se utilizó la base de datos de Cambridge (audios sin filtro y filtrados)— y se hizo una prueba adicional con los audios (audios web) que fueron recolectados en Ecuador utilizando la página web (https://databasecovid19.ups.edu.ec/).

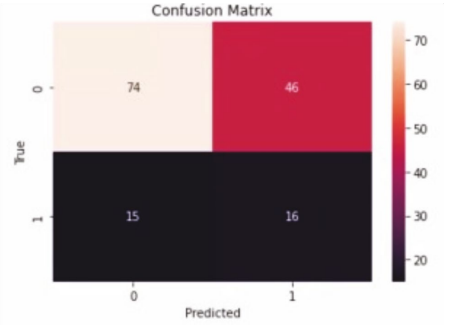
A continuación se muestran los resultados que se obtuvieron de todo el sistema de reconocimiento integrado, tanto para los audios como filtrados.

Tabla 1
Resultados del sistema integrado

	Métricas		
	Sensibilidad	**Especificidad**	**F1-Score**
Sin filtro	51 %	61 %	53 %
Filtrados	28 %	75 %	45 %
Audios Web	95 %	17 %	26 %

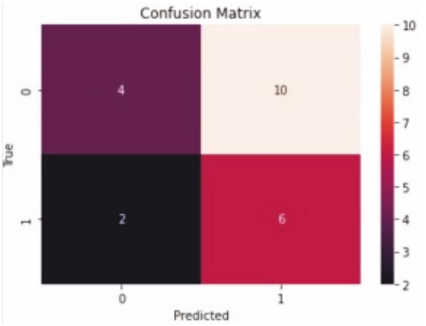
En la figura 1 se muestran la matriz de confusión para los audios sin filtro:

Figura 1
Matriz de confusión de audios sin filtro



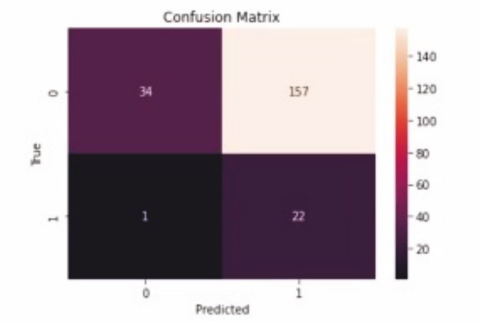
Mientras que en la figura 2 se muestra la matriz de confusión para los audios filtrados:

Figura 2
Matriz de confusión audios filtrados



Las figuras 1 y 2 están referidas al uso de la base de datos de Cambridge. En la figura 3, por otra parte, se muestra la matriz de confusión de la información obtenida con nuestra página web, es decir, nos muestra un panorama de lo que ha significado la pandemia en nuestro país. También nos permite cuantificar en qué medida el procedimiento seguido para la toma de información ha sido el adecuado.

Figura 3
Matriz de confusión audios web



Se puede observar que en el caso de los audios de Cambridge al utilizar audios filtrados únicamente se seleccionan 22 audios que según el Front-End contienen tos, mientras que en el caso de los audios sin filtrar se seleccionan 151 audios detectados como que contiene tos. Esto se da, debido a que el filtrado digital genera un tipo de señal de salida que le impide a los sistemas del Front-End diferenciar los audios correctamente. Al contrastar las métricas que se obtienen se puede notar que el que mejor funcionó es con los audios sin filtrar, incluso aunque se tienen más muestras por clasificar se obtiene una métrica de F1-Score mejor que en los audios filtrados. De igual manera se tiene un mejor equilibrio en la sensitividad y la especificidad.

Al comparar la matriz de confusión de los datos Web de nuestra base de datos con la de los datos filtrados de la base de Cambridge se puede ver que, si bien los valores son diferentes en cantidad, siguen la misma tendencia, se puede ver un acierto importante en relación con los casos positivos, en tanto que aparece una dificultad para detectar los negativos. Este resultado nos muestra que la base de datos de nuestra aplicación ha sido elaborada y recogida de una forma adecuada, además de asegurar que la ciudadanía aportó con información real y confiable.

Se realizó una prueba adicional con los audios web, ahí se puede notar que la clase positiva (tos COVID-19) se tiene una alta tasa de detección, pero para la clase negativa no se tiene una detección tan alta.

Conclusiones

El trabajo desarrollado "Tos por Covid-19: caracterización desde la Inteligencia Artificial", derivó en una producción de conocimiento muy importante en diferentes ámbitos de la ciencia de datos y de la inteligencia artificial aplicada a una necesidad puntual de un inmenso impacto social. Lo conseguido conlleva el convencimiento de que, para todos los grandes problemas que le atañen a la sociedad ecuatoriana, se deben obtener las soluciones siguiendo el camino de la investigación científica y del trabajo en equipo. Fue precisamente el trabajo en equipo lo que permitió obtener una cantidad muy importante y una calidad excepcional de resultados en un periodo de tiempo relativamente corto. Esto demuestra la importancia de las habilidades blandas en el proceso de construcción científica.

Al cierre del proyecto se pueden identificar cuatro ámbitos en los cuales se enfocó el trabajo desarrollado:

- El desarrollo del sitio web.
- La detección de la señal de tos y no tos.
- El filtrado de las señales de tos de otras fuentes de ruido.
- El modelado de la tos COVID-19, por medio de sistemas de aprendizaje automático.

En relación con el desarrollo del sitio web, se concluyó que la arquitectura cliente-servidor utilizada en este trabajo resultó la más conveniente, ya que era flexible y adaptable al servicio que se implementó, además de permitir el uso de diversas plataformas, bases de datos, redes y sistemas operativos, de diferentes fabricantes.

Dado que es el servidor el que controla el acceso a los datos y que se requiere de autorización del mismo servidor para el acceso, se tiene un mecanismo de seguridad en el sistema.

Por otra parte, el sistema de detección tos no-tos, es un sistema diseñado para realizar una clasificación binaria entre grabaciones de tos y de otros eventos sonoros (no-tos). En este sentido, se utilizaron las grabaciones de tos de la base de datos de Cambridge y 40 de las grabaciones de la base de datos de ruido ambiental para etiquetarlas como pertenecientes a la clase *tos*, mientras que se utilizaron las 1960 grabaciones de la base de datos de ruido ambiental para etiquetarlas como pertenecientes a la clase *no-tos*, de este modo se construyó un conjunto confiable de entrenamiento para el sistema de detección tos no-tos.

También se pudo observar que el sistema de reconocimiento de tos no-tos basado en redes neuronales convolucionales permite discriminar los sonidos de tos de otros eventos sonoros con una exactitud del 97,44 %, lo cual sentó una base sólida para desarrollar los próximos ensayos sobre el sistema modelado.

No fue del todo posible obtener un número significativo de grabaciones de tos, lo que dificultó la tarea de desarrollar el sistema automático de identificación de tos no-tos. Por ello, se trabajó en estudiar los efectos del aumento artificial de datos de entrada con el fin de balancear la base de datos y evitar el sobreajuste del sistema en la fase de evaluación.

Respecto a los métodos utilizados para conseguir el aumento de datos, se hizo uso combinado de los datos generados a partir de la variación del valor de frecuencia y la inyección de ruido en las grabaciones originales. Con ello se consiguió disminuir la diferencia en la medida de rendimiento del modelo entre los conjuntos de entrenamiento y validación, lo que indica que la técnica de aumento de datos funcionó eficazmente y se concluye que las transformaciones propuestas para incrementar la cantidad de datos ayudan a reducir el

sobreajuste del modelo, lo que permite utilizar estos procedimientos para entrenar redes neuronales convolucionales profundas orientadas a la diferenciación de la tos de otros eventos sonoros.

Si bien los resultados de este estudio son alentadores, es fundamental enfatizar que las muestras artificiales introducidas a la red neuronal aún están altamente correlacionadas, lo cual es una limitación. Esto se debe a que los datos aumentados provienen de una pequeña cantidad de datos originales, así no se está produciendo información completamente nueva, sino que solo se ha mezclado la información actual. En este sentido, esto podría no ser suficiente para deshacerse completamente del sobreajuste en algunos casos. Para estos casos, el trabajo futuro debería explorar algunas técnicas como el aprendizaje por transferencia o la extracción de características, e incluso preprocesar las grabaciones originales para eliminar segmentos no informativos.

En relación con el tercer ámbito de filtrado de la señal de tos para aislarla de sonidos externos, se creó un sistema basado en el uso de filtros adaptativos que es capaz de minimizar el ruido presente en un archivo de registro de tos. El ruido existente luego de aplicar el filtrado ha llegado a ser cercano a 0 dB con lo que se puede decir que la técnica propuesta ha funcionado adecuadamente.

Y, por último, en relación con el último ámbito que es el modelado de la tos COVID-19, se presentó, en primera instancia, el algoritmo de detección de sonidos realizado durante la actividad de reconocimiento de la tos en una señal de audio, consiguiendo obtener señales de audio donde esté presente únicamente el sonido de la tos, habiendo filtrado eventos externos.

Respecto al mismo ámbito y de una manera más concluyente, se presentó el algoritmo de detección de COVID-19 el cual tiene como objetivo caracterizar a una señal de tos integrando ya todo el sistema de reconocimiento.

Para la separación definitiva de señales tos no-tos, se utilizó la técnica de ensamblaje del modelo de clasificación por votación con una red neuronal convolucional, un bosque aleatorio y un clasificador basado en modelos ocultos de Márkov, para diferenciar los sonidos de señales de tos, de otros eventos de sonido ambiental, utilizando coeficientes para la representación del habla basados en la percepción auditiva humana (MFCC). Los resultados obtenidos experimentalmente mostraron que el enfoque de aprendizaje conjunto nos permite resolver las debilidades de algunos modelos con las fortalezas de otros. Se desarrolló un modelo robusto, con alta precisión de calificación, así como alta precisión y recuperación, en el sistema de reconocimiento de las señales de audio de tos.

Finalmente, se sabe que la clasificación de la señal de audio para la tos se ha utilizado con éxito para diagnosticar una variedad de afecciones respiratorias, y ha habido un interés significativo en aprovechar el aprendizaje automático para proporcionar una detección generalizada de COVID-19, por lo que la obtención del mejor rendimiento del sistema integrado de caracterización de las señales de audio analizadas, nos brindará una buena herramienta y aporte para una posterior identificación de esta enfermedad.