



**UNIVERSIDAD ESTATAL PENÍNSULA
DE SANTA ELENA
FACULTAD DE SISTEMAS Y TELECOMUNICACIONES**

TITULO DEL TRABAJO DE TITULACIÓN

Desarrollo de un Sistema Web para la Detección de Deepfakes de
Audio mediante Deep Learning y Análisis Espectral

AUTOR

SORIANO MEJILLONES KEVIN GERARDO

PROYECTO DE UNIDAD DE INTEGRACIÓN CURRICULAR

Previo a la obtención del grado académico en
INGENIERO EN TECNOLOGÍAS DE LA INFORMACIÓN

TUTOR

Ing. Rosero Vásquez Shendry Balmore, Msc.C.

Santa Elena, Ecuador

Año 2026



**UNIVERSIDAD ESTATAL PENÍNSULA
DE SANTA ELENA
FACULTAD DE SISTEMAS Y TELECOMUNICACIONES**

TRIBUNAL DE SUSTENTACIÓN

Ing. José Sánchez Aquino, Mgtr.
DIRECTOR DE LA CARRERA

Ing. Shendry Rosero Vázquez, Mgtr.
TUTOR

Lsi. Daniel Quirumbay Yagual, Mgtr.
DOCENTE ESPECIALISTA

Ing. Marjorie Coronel Suárez, Mgtr.
DOCENTE GUÍA UIC



**UNIVERSIDAD ESTATAL PENÍNSULA
DE SANTA ELENA
FACULTAD DE SISTEMAS Y TELECOMUNICACIONES**

CERTIFICACIÓN

Certifico que luego de haber dirigido científica y técnicamente el desarrollo y estructura final del trabajo, este cumple y se ajusta a los estándares académicos, razón por el cual apruebo en todas sus partes el presente trabajo de titulación que fue realizado en su totalidad por Soriano Mejillones Kevin Gerardo, como requerimiento para la obtención del título de Ingeniero en Tecnologías de la Información.

La Libertad, a los 23 días del mes de julio del año 2026

TUTOR

Shendry Rosero

Firmado digitalmente por Shendry Rosero
DN: cn=Shendry Rosero, gn=Shendry Rosero, c=ES,
Spain, l=ES, Spain, e=shendry.rosero.v@gmail.com
Motivo: Soy el autor de este documento
Ubicación:
Fecha: 2026-09-17 15:22:05.00

Ing. Shendry Rosero Vásquez, Mgtr.



**UNIVERSIDAD ESTATAL PENÍNSULA
DE SANTA ELENA
FACULTAD DE SISTEMAS Y TELECOMUNICACIONES**

DECLARACIÓN DE RESPONSABILIDAD

Yo, Soriano Mejillones Kevin Gerardo

DECLARO QUE:

El trabajo de Titulación, Desarrollo de un Sistema web para la detección de deepfakes de audio mediante Deep Learning y Análisis espectral previo a la obtención del título en Ingeniero en Tecnologías de la Información, ha sido desarrollado respetando derechos intelectuales de terceros conforme las citas que constan en el documento, cuyas fuentes se incorporan en las referencias o bibliografías. Consecuentemente este trabajo es de mi total autoría.

En virtud de esta declaración, me responsabilizo del contenido, veracidad y alcance del Trabajo de Titulación referido.

La Libertad, a los 23 días del mes de julio del año 2026

EL AUTOR



Validar únicamente en FirmaEC.
Firmado electrónicamente por:
**KEVIN GERARDO
SORIANO MEJILLONES**

Kevin Gerardo Soriano Mejillones



**UNIVERSIDAD ESTATAL PENÍNSULA
DE SANTA ELENA
FACULTAD DE SISTEMAS Y TELECOMUNICACIONES**

CERTIFICACIÓN DE ANTIPLAGIO

Certifico que después de revisar el documento final del trabajo de titulación denominado Desarrollo de un Sistema web para la detección de deepfakes de audio mediante Deep Learning y Análisis espectral, presentado por el estudiante, Soriano Mejillones Kevin Gerardo fue enviado al Sistema Antiplagio, presentando un porcentaje de similitud correspondiente al 6%, por lo que se aprueba el trabajo para que continúe con el proceso de titulación.



TUTOR

Shendry
Rosero

Firmado digitalmente por Shendry Rosero
DN: cn=Shendry Rosero, gn=Shendry
Rosero, c=ES, Spain, l=ES, Spain
e=shendry.rosero.v@gmail.com
Motivo: Soy el autor de este documento
Ubicación:
Fecha: 2026-09-17 15:23:05.00

Ing. Shendry Rosero Vásquez, Mgtr.



**UNIVERSIDAD ESTATAL PENÍNSULA
DE SANTA ELENA
FACULTAD DE SISTEMAS Y TELECOMUNICACIONES**

AUTORIZACIÓN

Yo, Soriano Mejillones Kevin Gerardo

Autorizo a la Universidad Estatal Península de Santa Elena, para que haga de este trabajo de titulación o parte de él, un documento disponible para su lectura consulta y procesos de investigación, según las normas de la Institución.

Cedo los derechos en línea patrimoniales del presente trabajo de titulación con fines de difusión pública, además apruebo la reproducción dentro de las regulaciones de la Universidad, siempre y cuando esta reproducción no suponga una ganancia económica y se realice respetando mis derechos de autor.

Santa Elena, a los 23 días del mes de julio del año 2026

EL AUTOR



Validar únicamente en FirmaBC.
Firmado electrónicamente por:
**KEVIN GERARDO
SORIANO MEJILLONES**

Kevin Gerardo Soriano Mejillones

AGRADECIMIENTO

Quiero agradecer a cada persona que fue parte de todo este arduo proceso, sobre todo a quienes confiaron en mí desde el primer día.

En primer lugar, quiero dar gracias a Dios por permitirme llegar hasta esta etapa.

También, agradezco profundamente a mi familia por su apoyo incondicional y su motivación constante. Ustedes fueron parte fundamental en este proceso, sin su ayuda nada de esto hubiera sido posible.

A mi tutor, Ing. Shendry Rosero, expreso mi profundo agradecimiento por cada charla, consejo, explicación y enseñanza brindada durante las tutorías. Su gran experiencia y conocimientos en el área fueron fundamentales para orientar el desarrollo de mi investigación.

A la Universidad Estatal Península de Santa Elena le extiendo mis más sinceros agradecimientos por haberme brindado la oportunidad de formarme en tan prestigiosa institución y por otorgarme los conocimientos y habilidades necesarios para mi futuro desempeño profesional.

Kevin Gerardo, Soriano Mejillones

DEDICATORIA

Quiero dedicar este trabajo a todas las personas que me han apoyado incondicionalmente a lo largo de mi vida.

A mis padres, Maritza Mejillones y Gerardo Soriano, por su amor, entrega y apoyo incondicional. Gracias a su esfuerzo diario, he podido superarme y alcanzar mis metas. Este trabajo se los dedico a ustedes, con mucho amor y gratitud por todo lo que han hecho por mí hasta hoy.

A mis hermanas, Kerly y Johanna Soriano, por ser una fuente constante de motivación y superación. Les dedico este trabajo realizado con todo mi esfuerzo, esperando que también sea una fuente de inspiración para que continúen alcanzando sus propios logros.

A mi pareja, Adriana González, por acompañarme durante toda mi formación, por cada desvelada, por su amor, paciencia, tiempo y apoyo incondicional. Gracias por estar presente en cada tropiezo y en cada acierto, por inspirarme a ser mejor persona y por permanecer a mi lado durante todo este proceso.

Este trabajo está dedicado a cada uno de ustedes como muestra de mi gratitud y afecto por todo el apoyo que me han brindado.

Kevin Gerardo, Soriano Mejillones

ÍNDICE GENERAL

TITULO DEL TRABAJO DE TITULACIÓN	II
TRIBUNAL DE SUSTENTACIÓN	III
CERTIFICACIÓN	IV
CERTIFICACIÓN DE ANTIPLAGIO	VI
AGRADECIMIENTO	II
DEDICATORIA	III
ÍNDICE GENERAL	IV
ÍNDICE DE TABLAS	VIII
ÍNDICE DE FIGURAS	X
RESUMEN	XI
ABSTRACT	XI
INTRODUCCIÓN	1
CAPITULO 1. FUNDAMENTACIÓN	2
1.1. Antecedentes	2
1.2. Descripción del Proyecto	5
1.3. Objetivos del Proyecto	8
1.3.1. Objetivo General	8
1.3.2. Objetivos Específicos	8
1.4. Justificación del Proyecto	8
1.5. Alcance del Proyecto	10
1.6. Metodología del Proyecto	11
1.6.1. Metodología de Investigación	11
1.6.2. Beneficiarios del Proyecto	14
1.6.3. Variables	14

1.6.4. Análisis de recolección de datos	14
1.7. Metodología de Desarrollo	19
CAPITULO 2. PROPUESTA	22
2.1. MARCO CONTEXTUAL	22
2.2. MARCO CONCEPTUAL	24
2.2.1. Inteligencia Artificial	24
2.2.2. Redes Neuronales	25
2.2.3. Procesamiento de Audio	26
2.2.4. Deepfakes	28
2.3. MARCO TEÓRICO	28
2.3.1. Enfoque de deep learning para la detección de audio sintético	28
2.3.2. Procesamiento de audio y extracción de características acústicas	29
2.3.3. Arquitecturas híbridas de aprendizaje profundo para clasificación de audio	30
2.4. Requerimientos	30
2.4.1. Requerimientos Funcionales	30
2.4.2. Requerimientos no Funcionales	34
2.5. Componentes de la Propuesta	35
2.5.1. Arquitectura del Sistema	35
2.5.2. Diagramas de Casos de Uso	38
2.5.3. Modelado de Datos	42
2.6. Diseño de Interfaces	44
2.6.1. Interfaz de registro e inicio de sesión	44
2.6.2. Interfaz de carga, análisis e historial de audios	45
2.6.3. Interfaz de visualización de resultados y generación de reportes	46

2.7. Pruebas	47
2.7.1. Caso N° 01 - Iniciar sesión y acceder a las funcionalidades protegidas	47
2.7.2. Caso N° 02 - Cargar y validar el formato del archivo del audio	48
2.7.3. Caso N° 03 - Validar la duración del archivo de audio	49
2.7.4. Caso N° 04 - Analizar y clasificar una muestra de audio	50
2.7.5. Caso N° 05 - Generar y descargar el reporte del análisis	51
2.8. Resultados	52
2.8.1. Resultados del desarrollo y optimización del modelo	52
CONCLUSIONES	64
RECOMENDACIONES	65
REFERENCIAS	67
ANEXOS	80
Anexo 1. Árbol de Problemas	80
Anexo 2. Formato de ficha de análisis	81
Anexo 3. Fragmento de Ficha de Análisis – Aishell	82
Anexo 4. Fragmento de Ficha de Análisis – Common Voice	83
Anexo 5. Fragmento de Ficha de Análisis – LibriSpeech	84
Anexo 6. Fragmento de Ficha de Análisis – VCTK Corpus	85
Anexo 7. Fragmento de Ficha de Análisis – ASVSpooof 2019	86
Anexo 8. Fragmento de Ficha de Análisis – CVoiceFake	87
Anexo 9. Fragmento de Ficha de Análisis – DFADD	88
Anexo 10. Fragmento de Ficha de Análisis – Fake-or-Real	89
Anexo 11. Fragmento de Ficha de Análisis – In-The-Wild	90
Anexo 12. Fragmento de Ficha de Análisis – LibriSeVoc	91
Anexo 13. Fragmento de Ficha de Análisis – MLAAD	92

Anexo 14. Fragmento de Ficha de Análisis – PartialSpoof	93
Anexo 15. Fragmento de Ficha de Análisis – WaveFake	94
Anexo 16. Fragmento de Ficha de Análisis – Coqui TTS	95
Anexo 17. Fragmento de Ficha de Análisis – Edge TTS	96
Anexo 18. Fragmento de Ficha de Análisis – ElevenLabs	97
Anexo 19. Fragmento de Ficha de Análisis – Higgs Audio	98
Anexo 20. Fragmento de Ficha de Análisis – Kokoro	99
Anexo 21. Fragmento de Ficha de Análisis – Piper TTS	100
Anexo 22. Caso N° 01 - Iniciar sesión y acceder a las funcionalidades protegidas	101
Anexo 23. Caso N° 02 - Cargar y validar el formato del archivo del audio	101
Anexo 24. Caso N° 03 - Validar la duración del archivo de audio	101
Anexo 25. Caso N° 04 - Analizar y clasificar una muestra de audio	102
Anexo 26. Caso N° 05 - Generar y descargar el reporte del análisis	102

ÍNDICE DE TABLAS

Tabla 1. Distribución del conjunto de datos seleccionados	18
Tabla 2. Requerimiento Funcional - Módulo de Autenticación y Gestión de Usuarios	31
Tabla 3. Requerimiento Funcional - Módulo de Carga y Procesamiento de Audios	32
Tabla 4. Requerimiento Funcional - Módulo de Análisis y Clasificación de Audios	32
Tabla 5. Requerimiento Funcional - Módulo de Visualización de Resultados	33
Tabla 6. Requerimiento Funcional - Módulo de Generación y Descarga de Reportes	33
Tabla 7. Requerimientos No Funcionales del Sistema	35
Tabla 8. Especificación del caso de uso de inicio de sesión	39
Tabla 9. Especificación del caso de uso de análisis de audio para la detección de deepfakes	42
Tabla 10. Caso de prueba N° 01 - Iniciar sesión y acceder a las funcionalidades protegidas	48
Tabla 11. Caso de prueba N° 02 - Cargar y validar el formato del archivo del audio	49
Tabla 12. Caso de prueba N° 03 - Validar la duración del archivo de audio	50
Tabla 13. Caso de prueba N° 04 - Analizar y clasificar una muestra de audio	51
Tabla 14. Caso de prueba N° 05 - Generar y descargar el reporte del análisis	52
Tabla 15. Métricas de evaluación del modelo base	55
Tabla 16. Métricas de evaluación del modelo optimizado	60
Tabla 17. Comparación del desempeño de los modelos de clasificación	61
Tabla 18. Formato de Ficha de Análisis	81

Tabla 19. Fragmento de Ficha de Análisis - Aishell	82
Tabla 20. Fragmento de Ficha de Análisis - Common Voice	83
Tabla 21. Fragmento de Ficha de Análisis – LibriSpeech	84
Tabla 22. Fragmento de Ficha de Análisis – VCTK Corpus	85
Tabla 23. Fragmento de Ficha de Análisis – ASVSpooF 2019	86
Tabla 24. Fragmento de Ficha de Análisis – CVoiceFake	87
Tabla 25. Fragmento de Ficha de Análisis – DFADD	88
Tabla 26. Fragmento de Ficha de Análisis – Fake-or-Real	89
Tabla 27. Fragmento de Ficha de Análisis – In-The-Wild	90
Tabla 28. Fragmento de Ficha de Análisis – LibriSeVoc	91
Tabla 29. Fragmento de Ficha de Análisis – MLAAD	92
Tabla 30. Fragmento de Ficha de Análisis – PartialSpooF	93
Tabla 31. Fragmento de Ficha de Análisis – WaveFake	94
Tabla 32. Fragmento de Ficha de Análisis – Coqui TTS	95
Tabla 33. Fragmento de Ficha de Análisis – Edge TTS	96
Tabla 34. Fragmento de Ficha de Análisis – ElevenLabs	97
Tabla 35. Fragmento de Ficha de Análisis – Higgs Audio	98
Tabla 36. Fragmento de Ficha de Análisis – Kokoro	99
Tabla 37. Fragmento de Ficha de Análisis – Piper TTS	100

ÍNDICE DE FIGURAS

Figura 1. Arquitectura General del Sistema Propuesto	36
Figura 2. Diagrama del caso de uso de inicio de sesión	38
Figura 3. Diagrama del caso de uso de análisis de audio para la detección de deepfakes	40
Figura 4. Modelo Relacional de la Base de Datos del Sistema	43
Figura 5. Interfaz de registro e inicio de sesión	44
Figura 6. Interfaz de carga, análisis e historial de audios	45
Figura 7. Interfaz de visualización de resultados y generación de reportes	46
Figura 8. Curva de pérdida de entrenamiento y validación del modelo base	54
Figura 9. Matriz de confusión del modelo base	56
Figura 10. Curva de pérdida del entrenamiento y validación del modelo optimizado	57
Figura 11. Curva ROC del modelo optimizado	58
Figura 12. Matriz de confusión del modelo optimizado	59
Figura 13. Representaciones espectrales antes y después del preprocesamiento	62
Figura 14. Resultado de una clasificación con nivel de confianza insuficiente	63
Figura 15. Árbol de Problemas	80
Figura 16. Caso de uso N° 01 - Iniciar sesión y acceder a las funcionalidades protegidas	101
Figura 17. Caso de uso N° 02 - Cargar y validar el formato del archivo del audio	101
Figura 18. Caso de uso N° 03 - Validar la duración del archivo de audio	101
Figura 19. Caso de uso N° 04 - Analizar y clasificar una muestra de audio	102
Figura 20. Caso de uso N° 05 - Generar y descargar el reporte del análisis	102

RESUMEN

Esta investigación aborda la detección de deepfakes de audio mediante técnicas de aprendizaje profundo y análisis espectral. Su objetivo fue desarrollar e integrar un modelo CNN-LSTM en una plataforma web para la clasificación de audios reales y deepfakes. Para ello, se recopiló y depuró un conjunto de datos conformado por muestras auténticas y sintéticas, del cual se obtuvieron representaciones espectrales basadas en coeficientes Mel, delta y delta-delta para el entrenamiento y la evaluación del modelo. Posteriormente, se comparó el desempeño del modelo base con una versión optimizada, seleccionándose esta última por alcanzar una precisión de 0.9961, un recall de 0.9933, un F1-Score de 0.9947, un AUC-ROC de 0.9990 y un EER de 0.0055. Finalmente, se implementó una plataforma web que permite analizar audios, visualizar los resultados de la clasificación y generar reportes, constituyendo una alternativa de apoyo en la detección de deepfakes de audio.

Palabras claves: Detección de deepfakes de audio, aprendizaje profundo, análisis espectral

ABSTRACT

This research addresses the detection of audio deepfakes using deep learning and spectral analysis techniques. Its objective was to develop and integrate a CNN-LSTM model into a web platform for classifying real audio and deepfakes. To achieve this, a dataset of authentic and synthetic samples was collected and refined. Spectral representations based on Mel, delta, and delta-delta coefficients were obtained from this dataset for training and evaluating the model. Subsequently, the performance of the base model was compared with an optimized version. The optimized version was selected due to its accuracy of 0.9961, recall of 0.9933, F1-Score of 0.9947, AUC-ROC of 0.9990, and EER of 0.0055. Finally, a web platform was implemented that allows for audio analysis, visualization of classification results, and report generation, providing an alternative tool for detecting audio deepfakes.

Keywords: Audio deepfake detection, deep learning, spectral analysis

INTRODUCCIÓN

Los avances recientes en la inteligencia artificial han propiciado el desarrollo de modelos de síntesis voz con un alto nivel de realismo, capaces de generar audios que replican fielmente las características de las voces humanas, lo que dificulta distinguirlos de grabaciones reales; dichas tecnologías, conocidas como deepfakes de audio, representan un importante avance en ámbitos como la asistencia virtual, producción multimedia y otras temáticas vinculadas a la accesibilidad. No obstante, su uso también plantea nuevos desafíos relacionados con la autenticidad del contenido digital, lo que ha impulsado el desarrollo de herramientas orientadas al análisis objetivo las señales de audio y a la detección de posibles alteraciones.

En este contexto, la detección automática de audios sintéticos se ha convertido en una línea de investigación de creciente interés para la inteligencia artificial, procesamiento digital de señales y la ciberseguridad; sin embargo, la distinción entre voces naturales y sintetizadas requiere del análisis de características acústicas que, en la mayoría de los casos, no pueden detectarse mediante una inspección convencional; por lo tanto, el uso de los modelos de aprendizaje profundo y las técnicas de procesamiento espectral representan una alternativa prometedora para mejorar el desempeño de los sistemas de clasificación frente a los continuos avances de las tecnologías de síntesis de voz. Como respuesta a esta problemática, la presente investigación propone el desarrollo de un sistema web para la detección de deepfakes de audio mediante una arquitectura híbrida CNN-LSTM con mecanismo de atención, cuya implementación permitirá analizar audios y visualizar las predicciones generadas por el modelo.

Finalmente, el documento se estructura en dos capítulos que presentan de manera ordenada el desarrollo de la propuesta investigativa. El primer capítulo está conformado por los antecedentes, la descripción del proyecto, los objetivos, la justificación, el alcance y la metodología de la investigación; mientras que el segundo capítulo desarrolla los marcos contextual, conceptual y teórico, la metodología de desarrollo, la arquitectura del sistema, las fases de implementación de la propuesta y los resultados obtenidos, junto con las conclusiones y recomendaciones derivadas de la investigación.

CAPITULO 1. FUNDAMENTACIÓN

1.1. Antecedentes

El desarrollo acelerado de la inteligencia artificial y las tecnologías de aprendizaje profundo ha generado inquietud en gran parte de la sociedad, debido al aumento de desinformación y fraude que han provocado los avances significativos en cuanto a generación de imágenes, textos, videos y audios falsos, conocidos comúnmente como deepfakes [1]. Los deepfakes tienen la capacidad de modificar o generar contenido de todo tipo de manera tan realista, lo que dificulta cada vez más distinguir entre medios auténticos y manipulados; según un estudio [2], [3], YouTube es la principal plataforma de distribución de deepfakes, representando el 40% de los casos registrados. Sin embargo, también se evidenciaron volúmenes considerables de deepfakes en Twitter, Facebook y otras redes sociales.

Los deepfakes se han desplegado a nivel mundial, debido a la accesibilidad de herramientas de generación que facilitan su uso indebido generalizado [4]. Según [5], [6], de 2018 a 2019, el número de artículos relacionados con deepfakes aumentó significativamente, de 60 a 309; mientras que para finales de 2020 se estimaba que incrementaría a más de 730 artículos, pero el número fue mucho mayor, alcanzando los 1323 artículos publicados hasta esa fecha. Esta tecnología ha evolucionado rápidamente con el paso del tiempo, provocando fraudes, pérdidas financieras, suplantación de identidad y otros incidentes críticos; según [7], [8], los países más afectados por estas tecnologías en 2024 fueron: China 32%, Corea del Sur e Indonesia 18%, Singapur 12%, Hong Kong 11%, Vietnam 6% e India 3%.

Los deepfakes de audio implican desafíos directos hacia la integridad y autenticidad del contenido digital; a pesar de que su importancia no es comparable con la de otros tipos de deepfakes, son considerados una amenaza significativa, debido a su potencial para crear situaciones en las que los individuos tienen que tomar decisiones rápidas [9]. Un claro ejemplo de su impacto sucedió durante la contienda entre Ucrania y Rusia, cuando se propagó un audio falso del presidente ucraniano Volodymyr Zelenskyy pidiendo la rendición militar; este evento demostró que los deepfakes tienen la capacidad de socavar la confianza, propagar desinformación e influir en los relatos geopolíticos [4].

El incremento en la difusión de contenidos manipulados a través del mundo digital se ve impulsado por la desinhibición en línea generada por el aparente anonimato en la red y la percepción de escasa aplicación de medidas correctivas [10]. Uno de los principales retos que se enfrentan en la actualidad, consiste en reducir la propagación indiscriminada de ese tipo de contenidos por parte de los usuarios, fomentando la sensibilización acerca de las consecuencias de su divulgación, con el objetivo de evitar su circulación en la red y así reducir el daño que puedan ocasionar [11]. Debido a la facilidad de acceso a herramientas de generación, numerosas personas la emplean con fines malintencionados, lo que provoca un deterioro en la reputación de personas o instituciones [12]. Tal como ocurre en ciertos tipos de deepfakes, donde se usa el rostro o la voz de figuras tanto públicas como privadas, con la finalidad de afectar su reputación personal y profesional [13].

En la actualidad, se dispone de muy pocas herramientas de detección de contenido manipulado, y las que existen suelen ser de pago o no son ampliamente utilizadas, lo que deja a los usuarios sin un mecanismo eficaz para detectar y defenderse de contenidos maliciosos; por ello, se resalta la necesidad de desarrollar nuevas herramientas que faciliten la detección rápida de información falsa para evitar que los usuarios sean engañados [14]. En este contexto, sin las herramientas sofisticadas para la detección de deepfakes, la generación de este tipo de contenido malicioso supone múltiples riesgos e implicaciones legales que influyen en el bienestar social, ya que en algunas ocasiones se usan deepfakes de audio o video en salas de audiencia y en investigaciones policiales, lo que puede ser perjudicial para las víctimas, si se exponen evidencias falsas en su contra [15].

La falta de normativas y mecanismos para comprobar la veracidad de la información permiten que esta tecnología pueda alterar contenido hasta el punto de que resulta prácticamente imposible distinguirlo del material auténtico, influyendo en la opinión pública, grupos sociales y discursos políticos [16]. La escasa supervisión y control genera un entorno propicio para la desinformación, lo cual puede afectar negativamente la salud mental de las víctimas, provocándoles sentimientos de inseguridad, temor y desconfianza generalizada con respecto a la información difundida por los medios de comunicación [17]. Esto subraya la

necesidad de desarrollar sistemas que permitan detectar deepfakes de todo tipo, con el objetivo de verificar la autenticidad de la información y disminuir los riesgos vinculados a la difusión de contenidos manipulados.

En el ámbito internacional, existen varios trabajos de investigación que han alcanzado porcentajes considerables de precisión y exactitud en la detección de deepfakes, tal como el trabajo de investigación denominado “Detección de deepfakes mediante redes neuronales convolucionales y modelado LSTM”, realizado por la Universidad Sigma de la India [1], que obtuvo un 97% de precisión, exactitud y puntuación F1, evidenciando una mejora significativa con respecto a otros modelos que solo alcanzaron una precisión del 53%, lo que destaca la eficiencia del modelo y la disminución de la tasa de error. Este trabajo de investigación servirá de referencia para la construcción de este proyecto en base a las técnicas y métricas que se aplican en su desarrollo.

A nivel de América Latina, el trabajo "Modelo de Detección de Audios en Español y Videos Deepfake Contra Ataques de Ingeniería Social" realizado por la Universidad de los Andes, comprobó la factibilidad de la detección de audios deepfake mediante un sistema basado en la representación de las señales de voz usando coeficientes MFCC y la evaluación de modelos de aprendizaje profundo, como MLP o CNN; los resultados revelan que la arquitectura CNN obtuvo un mejor desempeño, alcanzando una precisión de 97,49 %, un recall de 96,89 % y un F1-score de 97,45 % [18]. Este trabajo de investigación contribuye al desarrollo de este proyecto, al respaldar el uso de MFCC para procesar audio, las redes CNN para la clasificación de audios falsificados y las métricas como precisión, recall y F1-Score para evaluar el rendimiento del modelo.

Dentro del contexto local, en el trabajo de titulación denominado “Aplicación de técnicas de IA y acústica forense para identificar audios sintéticos”, de la Universidad Estatal Península de Santa Elena, se implementaron técnicas de IA y acústica forense, usando la extracción de características a través de MFCC, así como herramientas especializadas para el análisis espectral y el cálculo de coeficientes cepstrales [19]. Sin embargo, en este proyecto no se abordará el uso de aplicaciones sofisticadas para analizar, comparar y detectar contenido manipulado,

ya que se pretende desarrollar una arquitectura propia que integre todos estos componentes en un único sistema, capaz de procesar el audio y determinar su autenticidad.

Las investigaciones revisadas muestran una clara preocupación por el aumento de deepfakes no sólo en el audio sino también en otras formas de contenido digital. De igual forma, subrayan las restricciones en aspectos esenciales, como la detección de audios fraudulentos y evidencian la evolución constante de métodos y recursos para crear contenido sintético. Como respuesta a esta situación, este proyecto sugiere la creación de un sistema web para detectar deepfakes de audio, que contemple la generación de reportes fundamentados en los análisis realizados, con el objetivo de minimizar los riesgos asociados al uso indebido de estas tecnologías y herramientas.

1.2. Descripción del Proyecto

A lo largo del tiempo, los deepfakes se han vuelto cada vez más reales, generando escenarios complejos en los que detectarlos resulta casi una misión imposible; los desarrollos tecnológicos están introduciendo herramientas y técnicas avanzadas que permiten alterar diversos tipos de contenido, lo que representa un riesgo para la seguridad e integridad de las personas y los países [20]. Debido a esto, se propone el desarrollo de un sistema en línea que integre un algoritmo de aprendizaje profundo, capaz de procesar, examinar y detectar deepfakes de audios a partir de sus características espectrales; además, se elaborará un informe técnico fundamentado en este análisis, con el fin de garantizar y corroborar la autenticidad y la validez de los archivos de audio.

El sistema sugerido adopta un proceso que comienza con la recopilación de audios reales y sintéticos, que son sometidos a un preprocesamiento que abarca su conversión a espectrogramas log-Mel, debido a su capacidad para modelar la percepción auditiva humana, y la incorporación de coeficientes delta y delta que permiten capturar variaciones temporales; dichas representaciones sirven como entrada para un modelo fundamentado en una arquitectura CNN-LSTM, responsable de extraer características y modelar dependencias en las señales de audio; de este proceso se obtienen embeddings que se emplean en la fase de

clasificación; por último, se aplican métodos de regularización y se ajustan los parámetros importantes para mejorar la capacidad de generalización del modelo. Los resultados se muestran en una página web que facilita la comprobación de la veracidad de los audios, lo que se acompaña de un informe técnico.

Finalmente, este estudio enfatiza la importancia de la inteligencia artificial en la detección de deepfakes de audio; no solo desde el punto de vista técnico, sino también como una manera de reducir los peligros relacionados con la desinformación; desde esta perspectiva, el uso de modelos de aprendizaje profundo permite abordar este problema mediante el análisis de patrones complejos de forma automática; esto propicia la creación de soluciones más precisas y confiables [21], [22], [23], [24]. Desde esta perspectiva, el proyecto se organiza en etapas que abarcan desde el procesamiento de datos hasta la implementación del sistema, garantizando un desarrollo ordenado y coherente. A continuación se detallan las fases previstas:

Fase 1: En esta fase, se recopilarán audios reales y sintéticos de repositorios públicos y herramientas de generación de voz con inteligencia artificial; posteriormente, se efectuará el proceso de estandarización para adaptar el formato de los archivos, la frecuencia de muestreo y el número de canales, el cual abarcará el recorte de silencios y la eliminación de muestras que no cumplan con los mínimos en su duración. Durante el proceso, se llevará a cabo la verificación de la integridad del conjunto de datos disponible, se registrará la información correspondiente de cada archivo y se procederá a la partición en los subconjuntos de entrenamiento y validación y prueba.

Fase 2: Esta fase estará orientada a la obtención de representaciones acústicas que permitan describir las características más relevantes de las señales de audio para el proceso de clasificación. Para ello, los audios serán segmentados en ventanas temporales, normalizados y transformados a espectrogramas log-Mel, que serán posteriormente complementados con coeficientes delta y delta-delta; además, se aplicarán técnicas de data augmentation y las características generadas serán almacenadas en un formato adecuado para su utilización durante el entrenamiento y evaluación del modelo.

Fase 3: En esta fase, se desarrollará una arquitectura híbrida basada en redes neuronales convolucionales (CNN), redes LSTM bidireccionales y un mecanismo de atención para el análisis de las representaciones acústicas obtenidas; como parte de esta etapa, se definirá la configuración inicial del entrenamiento, incluyendo estrategias de regularización para potenciar la estabilidad del aprendizaje y reducir el sobreajuste. Como resultado de este proceso, se obtendrá una arquitectura base que servirá como punto de partida para el proceso de optimización.

Fase 4: Esta fase estará dedicada a mejorar la arquitectura base mediante el ajuste de hiperparámetros y la incorporación de mejoras orientadas a potenciar la capacidad de clasificación y generalización del modelo; para ello, se realizarán modificaciones en la configuración del entrenamiento, el procesamiento de las muestras, las estrategias de atención y muestreo balanceado, con el fin de mejorar su desempeño. Finalmente, se evaluará el modelo optimizado mediante métricas de rendimiento, con el propósito de seleccionar la configuración óptima que será integrada en el sistema.

Fase 5: En esta fase, se desarrollará el sistema web que integrará los módulos de autenticación y gestión de usuarios, carga y procesamiento de audios, análisis y clasificación de la muestra, visualización de los resultados y generación de reportes. El sistema incluirá el modelo optimizado para analizar muestras de audio y presentará al usuario la probabilidad de autenticidad resultante de la clasificación. Asimismo, se realizarán pruebas con audios externos para comprobar su correcto funcionamiento y se integrará la generación de reportes en formato PDF que incluirán los resultados obtenidos y las representaciones gráficas del análisis.

Este proyecto se enmarca en la línea de investigación Tecnología y Sistemas de la Información (TSI), específicamente en la sub-línea de Inteligencia Computacional, ya que plantea el desarrollo de un sistema web que emplea técnicas de Deep Learning y análisis espectral para detectar deepfakes de audio; la elección de esta línea se fundamenta en el uso de modelos de redes neuronales avanzadas y métodos de aprendizaje automático orientados al reconocimiento y clasificación de patrones acústicos, con la finalidad de desarrollar una solución tecnológica que contribuya a verificar la autenticidad de los audios de forma automática y confiable [25].

1.3. Objetivos del Proyecto

1.3.1. Objetivo General

Desarrollar un sistema web de detección de deepfakes de audio mediante una arquitectura CNN-LSTM y técnicas de procesamiento espectral, orientado a la clasificación de audios auténticos y sintéticos.

1.3.2. Objetivos Específicos

- Construir un conjunto de datos de audios reales y sintéticos provenientes de repositorios públicos y herramientas de generación artificial, adecuados para el entrenamiento y evaluación del sistema.
- Depurar el conjunto de datos mediante procesos de validación, normalización y eliminación de registros duplicados o inconsistentes, asegurando la calidad y confiabilidad de los datos utilizados en el modelo.
- Generar representaciones espectrales de tipo log-Mel y coeficientes delta y delta-delta a través de técnicas de procesamiento digital de señales, facilitando la identificación de patrones relevantes para el modelo.
- Implementar una arquitectura CNN-LSTM basada en representaciones espectrales log-Mel y técnicas de aumento de datos para la clasificación de audios auténticos y manipulados.
- Evaluar el desempeño del modelo CNN-LSTM mediante métricas como accuracy, recall, F1-score, AUC-ROC y EER, determinando su capacidad de discriminación en la detección de deepfakes de audio.
- Integrar el modelo de detección en una plataforma web a través de módulos de análisis, visualización y generación de reportes, facilitando la interpretación de los resultados obtenidos.

1.4. Justificación del Proyecto

En los últimos años, la inteligencia artificial se ha expandido de forma considerable y con ello, las técnicas y métodos para generar y propagar desinformación se vuelven cada vez más sofisticadas, lo que plantea desafíos significativos para la seguridad y confiabilidad de los contenidos digitales; por ello, resulta fundamental desarrollar sistemas de detección confiables y accesibles que contribuyan a reducir

los riesgos derivados del uso indebido de estas técnicas [26].

En el ámbito social, la relevancia del sistema propuesto radicará en su capacidad para detectar de forma automática deepfakes de audio, lo que podría contribuir a evitar que los usuarios sean engañados y a mitigar las implicaciones legales que conlleva el uso de esta tecnología, como fraudes, actividades ciberdelictivas y la usurpación de identidad en contenido pornográfico; así como la difusión indebida y malintencionada por parte de los usuarios, y la alteración de pruebas que podrían incriminar erróneamente a un individuo por un presunto delito [27], [28].

El sistema de detección de deepfakes de audio, surge en base a la necesidad urgente de contar con mecanismos eficaces para la detección y verificación de este tipo de contenido, con el objetivo de responder a la creciente evolución de las técnicas y herramientas de generación y manipulación de voz mediante inteligencia artificial, favoreciendo análisis más eficientes y reduciendo la necesidad de una revisión técnica humana [10], [17].

Al integrar redes neuronales, algoritmos de clasificación, extracción de características auditivas y análisis espectral, el modelo de Deep Learning será capaz de identificar patrones asociados a alteraciones en los archivos de audio, proporcionando una herramienta tecnológica que permitirá fortalecer la autenticidad de la información digital [9], [29]. Este aporte es de suma importancia en un contexto donde la credibilidad en los medios digitales se ha debilitado a causa de la proliferación de contenidos manipulados.

Este estudio contribuye a la creación de métodos y métricas para examinar audios mediante el uso de técnicas de Deep Learning, fomentando la seguridad en el ámbito tecnológico. De igual manera, proporciona un marco de referencia para investigaciones futuras, en las cuales se podría examinar la introducción de otros tipos de algoritmos de clasificación o técnicas de Deep Learning, así como la evaluación de métricas alternativas que podrían ayudar a mejorar la precisión de los resultados obtenidos en este estudio.

El presente proyecto está alineado con el objetivo del plan de desarrollo “Ecuador No Se Detiene”, en el siguiente eje:

Eje: Institucional

Objetivo 8.- Fortalecer la institucionalidad pública de forma eficiente, transparente y participativa [30].

Política 8.3.- Impulsar la transformación digital del Estado, la adopción del modelo de Estado Abierto, la protección de la información, con un entorno digital seguro y confiable en todos los niveles de gobierno, así como la integridad pública y la lucha contra la corrupción, que promueva la gestión pública eficiente, inclusiva, transparente y participativa [30].

1.5. Alcance del Proyecto

El sistema se centrará en la clasificación binaria de los audios, determinando si estos corresponden a contenido auténtico o a deepfakes; además, se incorporará un mecanismo de atención que facilitará la identificación de las áreas más relevantes de la señal durante el proceso de decisión, fortaleciendo así la capacidad de discriminación del modelo. El sistema también incluirá un proceso de inferencia que permitirá examinar nuevos audios, produciendo como resultado probabilidades asociadas a las clases “real” y “falso”; con base en estos resultados, se generarán informes que contendrán representaciones gráficas del análisis acústico realizado, acompañadas de información relevante sobre las características identificadas en las señales de audio, esto no solo brindará una evaluación del desempeño del sistema, sino que también facilitará la comprensión de los resultados obtenidos.

Con respecto a las limitaciones, el enfoque del proyecto se centra exclusivamente en el análisis de señales de audio, por lo que no se contempla la inclusión de otros tipos de contenido como imágenes o vídeos, ni el procesamiento o la detección de audios en tiempo real; por otra parte, la detección se restringe únicamente a los tipos de deepfakes de audio que forman parte del conjunto de entrenamiento, el cual reúne diversas tecnologías de síntesis de voz (TTS) desarrolladas en los últimos años. Por esta razón, el sistema no abarca la totalidad de las técnicas existentes, ya que su incorporación requeriría conjuntos de datos específicos y un nuevo proceso de entrenamiento del modelo.

Por otra parte, el manejo de datos considera únicamente archivos de audio en

formato estándar, como WAV, FLAC, MP3, OGG y OPUS, por su compatibilidad con las funciones de procesamiento implementadas en el sistema; además, únicamente se admiten audios con una duración comprendida entre 4 segundos y 5 minutos, con el objetivo de asegurar condiciones homogéneas en el procesamiento. Finalmente, no se integran las funcionalidades de identificación de hablantes, reconocimiento automático del contenido hablado, ni tampoco el análisis semántico del audio, ya que el sistema está orientado principalmente al análisis de las características acústicas de la señal.

1.6. Metodología del Proyecto

1.6.1. Metodología de Investigación

1.6.1.2. Contexto de la investigación

Este proyecto se fundamenta en la combinación de metodologías propuestas en investigaciones previas, con la finalidad de seleccionar y adaptar componentes para conformar un marco metodológico propio que aborde adecuadamente los requerimientos de la presente investigación; para ello, se consideró la metodología propuesta en el trabajo de investigación “Deepfake Audio Detection via MFCC Features Using Machine Learning”, que evidenció la efectividad del enfoque combinado de CNN-LSTM junto con un clasificador adicional en la detección de deepfakes de audio, donde la secuencia ordenada de las fases propuestas permitió abordar de manera adecuada el preprocesamiento de los datos, la extracción de características acústicas, el desarrollo de los modelos de clasificación y la comparativa de estos; proporcionando así una base metodológica que orientará el desarrollo de este proyecto [7], [8], [31].

La implementación de este proyecto se llevará a cabo en un entorno simulado, mediante pruebas experimentales que utilizarán conjuntos de datos previamente seleccionados y clasificados como auténticos o manipulados. Esto hace posible realizar análisis comparativos y evaluar el rendimiento del sistema en la identificación de alteraciones en los archivos de audio. Este proceso es fundamental para validar la efectividad del modelo propuesto bajo condiciones controladas y sin comprometer datos reales. Además, permitirá identificar posibles limitaciones y oportunidades de mejora durante su aplicación.

1.6.1.3. Diseño y alcance de la investigación

1.6.1.3.1. Diseño

El presente trabajo de investigación adopta un diseño experimental de carácter aplicado y centrado en la implementación de una arquitectura híbrida CNN-LSTM en la que se incorpora un mecanismo de atención, orientada a la detección de audios deepfake; durante la fase experimental, se realizarán varias configuraciones de entrenamiento, así como el ajuste de los hiperparámetros, con el objetivo de optimizar el rendimiento computacional del modelo. Posteriormente, el sistema será analizado mediante el uso de métricas objetivas tales como precisión, recall, F1-Score, AUC-ROC y EER; de este modo, se podrá validar el funcionamiento del modelo bajo condiciones controladas y determinar su capacidad para clasificar audios reales y sintéticos de forma automatizada [32].

1.6.1.3.2. Alcance

El alcance de la presente investigación tiene un enfoque descriptivo, puesto que estará orientado a describir el comportamiento del modelo de Deep Learning; para ello, se analizarán las métricas de rendimiento registradas durante las etapas de entrenamiento, validación y pruebas, permitiendo describir de forma objetiva la capacidad del modelo para el análisis de audios [33]. De forma complementaria, se identificarán las fortalezas y limitaciones de la propuesta con base en los resultados obtenidos.

1.6.1.4. Tipo y método de investigación

La presente investigación es de índole cuantitativa, ya que se basará en la recopilación, procesamiento y análisis de datos numéricos derivados del conjunto de datos utilizado durante las fases de entrenamiento y evaluación del modelo; los resultados obtenidos serán interpretados mediante indicadores estadísticos que permitirán medir objetivamente su rendimiento y comparar los resultados obtenidos en cada etapa del proceso [34]. Este enfoque permitirá disponer de información cuantificable que facilitará la valoración del desempeño de la propuesta y la interpretación de los resultados alcanzados.

Con respecto al método de investigación, este estudio emplea el método hipotético-

deductivo, al plantear la hipótesis de que una arquitectura híbrida CNN-LSTM con mecanismo de atención permitirá detectar audios deepfake con un rendimiento adecuado; dicha hipótesis se contrastará a través del desarrollo, entrenamiento y evaluación experimental del modelo bajo un procedimiento experimental [35]. Además, se utilizará el método analítico para analizar las características acústicas extraídas de las señales de audio durante la fase de procesamiento, y el método sintético que permitirá integrar los resultados obtenidos en una solución tecnológica funcional. [36].

1.6.1.5. Población

La población del presente estudio estará conformada de grabaciones de audio auténticas y sintéticas obtenidas de repositorios públicos y herramientas de generación de voz mediante IA; inicialmente, se recopilaron 119.808 ficheros de audio y, tras una revisión inicial, se descartaron tres registros, dado que dos de ellos presentaban una duración excesiva y uno tenía una duración de cero; posteriormente, se aplicó un proceso de estandarización que incluyó la normalización del formato, la frecuencias de muestreo y el recorte de silencios, durante el cual se descartaron 1.355 audios cuya duración final fue inferior a un segundo; como resultado, la población quedó conformada por 118.450 ficheros de audio, los cuales fueron utilizados en su totalidad, por lo que no fue necesario aplicar un procedimiento de muestreo [32], [37], [38].

1.6.1.6. Técnicas e instrumentos de recolección de datos

Para la recolección de datos se empleará la técnica de análisis documental, ya que esta investigación se basa en la revisión, selección y organización de archivos de audio provenientes de repositorios públicos especializados y de muestras sintéticas generadas mediante herramientas de síntesis de voz. La incorporación de muestras con diferentes características de generación permitirá conformar un conjunto de datos más diverso y representativo para el entrenamiento y evaluación del modelo propuesto; además, esta técnica permitirá comprobar de manera sistemática factores como el formato, la frecuencia de muestreo, la duración, el etiquetado y los metadatos existentes, garantizando que los archivos seleccionados cumplan con los criterios definidos para el desarrollo de la investigación [39].

Como apoyo a este proceso, se utilizarán scripts desarrollados en Python para facilitar la organización, validación y etiquetado de los archivos de audio, complementando la revisión manual realizada durante la conformación del conjunto de datos; por otra parte, el instrumento de recolección de datos será la ficha de análisis, la cual permitirá registrar de forma estructurada la información técnica de cada archivo, incluyendo los metadatos, características acústicas y la categoría correspondiente, contribuyendo a garantizar una gestión ordenada, coherente y verificable de la información recopilada [39].

1.6.2. Beneficiarios del Proyecto

Entre los beneficiarios se incluyen:

- **Público en general:** Personas que reciben o detectan grabaciones de audios en plataformas digitales y desean asegurarse de que no han sido alteradas o manipuladas.

1.6.3. Variables

En el presente estudio, se identifican las siguientes variables:

- **Precisión de detección de deepfakes de audio (%):** Mide el porcentaje de audios clasificados correctamente por el modelo CNN-LSTM con mecanismo de atención, permitiendo evaluar su desempeño en la detección de contenido auténtico y sintético [31], [40].

1.6.4. Análisis de recolección de datos

1.6.4.1. Recopilación de audios genuinos y manipulados

La recopilación de datos se fundamentó en la técnica de análisis documental, entendiendo como el procedimiento de búsqueda, selección y organización e interpretación de la información proveniente de documentos digitales. En esta investigación, los documentos corresponden a archivos de audio que se obtuvieron de repositorios públicos especializados como OpenSLR, Kaggle, Zenodo y Hugging Face, los cuales ofrecen datasets ampliamente utilizados en investigaciones asociados con el procesamiento de voz y la detección de deepfakes; con base en estas fuentes, se recopilaron audios reales y sintéticos que presentan

diversas características acústicas y condiciones de grabación, permitiendo conformar un conjunto de datos más amplio, diverso y representativo para el desarrollo de la investigación.

Durante la fase de selección, se revisaron los metadatos correspondientes de cada uno de los datasets, los cuales facilitaron la identificación de información relevante como el hablante, la etiqueta asignada y otras características técnicamente relevantes de las muestras; posteriormente, se realizó un proceso de balanceo por speakers, seleccionando un número equivalente de audios por cada hablante hasta llegar a la dimensión del dataset definido en el plan de investigación. Esta estrategia permitió reducir posibles sesgos asociados a una representación desmedida de determinados hablantes, favoreciendo una distribución más equilibrada de las muestras empleadas para el entrenamiento y evaluación del modelo.

En la fase de organización, se verificó la existencia de posibles casos de data leakage en los subconjuntos de entrenamiento, validación y prueba. Como resultado de esta revisión, no fueron identificados archivos duplicados, ni muestras compartidas entre las particiones, lo que confirma la independencia de cada uno de los subconjuntos y, por tanto, contribuye a una evaluación confiable del modelo. Únicamente se aislaron tres audios en cuarentena, dos de ellos por presentar una duración superior al promedio del dataset y uno adicional por registrar una duración de cero segundos, condiciones que podrían afectar el procesamiento y el análisis de las muestras.

Finalmente, el conjunto de datos fue distribuido de manera que el 70% de las muestras se destinó al entrenamiento, el 15% se asignó a la validación y el 15% restante a las pruebas. Esta distribución permitió separar las etapas de aprendizaje, ajuste y evaluación del modelo, garantizando que las métricas de desempeño se obtuvieran a partir de muestras no empleadas durante la fase de entrenamiento. De este modo, el proceso de análisis documental dio lugar a un conjunto de datos organizado, balanceado y consistente, adecuado para el desarrollo de las etapas posteriores de la investigación. Además, esta organización de los datos contribuyó a mejorar la confiabilidad del proceso experimental y a obtener resultados más sólidos durante la evaluación del modelo.

1.6.4.2. Generación de audios sintéticos

La generación de audios sintéticos se integró como estrategia complementaria para ampliar la variabilidad del dataset y poder reflejar diferentes patrones de síntesis de voz de las tecnologías actuales; en este caso se utilizaron diferentes herramientas de conversión de texto a voz como ElevenLabs, Coqui TTS, Piper TTS, Edge TTS, Kokoro, y Higgs Audio, mediante las que se generaron 3000 audios a partir de oraciones comunes y diferentes configuraciones de voz.

Cada motor de síntesis de voz implementa sus propios modelos y procesos para generar el habla, lo que se refleja en la aparición de diferencias perceptibles en los audios producidos. Estas diferencias abarcan aspectos como la pronunciación, la entonación, el ritmo y el timbre de voz, aportando una mayor riqueza acústica a las muestras obtenidas; de esta forma, el conjunto de audios también incluyó registros con características heterogéneas, acordes con las tecnologías de síntesis de voz disponibles en la actualidad. Al concluir con esta fase, se dispuso de un conjunto de datos compuesto por audios reales y sintéticos, organizado y balanceado en función de los criterios establecidos para en la investigación; este conjunto de muestras constituyó la base para dar continuidad a las etapas posteriores del estudio.

1.6.4.3. Aplicación de la ficha de análisis

La ficha de análisis se empleó como instrumento para organizar y sistematizar la información obtenida durante la recopilación de los conjuntos de datos, permitiendo registrar las principales características técnicas y descriptivas de cada dataset; como resultado, se elaboraron 19 fichas de análisis en las que se incluyó información relacionada con el origen de los datos, formato de los archivos, frecuencia de muestreo, duración de las muestras, categorías disponibles y los metadatos proporcionados por cada repositorio. Este registro facilitó la comparación de los diferentes conjuntos de datos y sirvió como una base objetiva para respaldar las decisiones adoptadas en las fases posteriores del procesamiento de los audios.

El análisis de las fichas permitió identificar diferencias importantes entre los conjuntos de datos, especialmente en aspectos como la frecuencia de muestreo, la duración de los audios, el formato de almacenamiento y la disponibilidad de

metadatos. Estos hallazgos constituyeron una de las razones por las que se consideró necesario aplicar procesos de estandarización y validación antes del entrenamiento del modelo, garantizando que todas las muestras cumplieran con condiciones técnicas homogéneas para su procesamiento.

Además de documentar las características de cada conjunto de datos, la ficha de análisis constituyó en un mecanismo de control que permitió verificar la consistencia de la información recopilada antes de su integración en el conjunto de datos definitivo. Este procedimiento favoreció la trazabilidad de las muestras usadas y proporcionó un respaldo documental de las decisiones adoptadas durante la etapa de la recolección. Las 19 fichas de análisis se incluyen como anexos (**Véase Anexo 3 al 21**), ya que, por su extensión no pueden ser incluidos como parte del cuerpo del documento; no obstante, permanecen disponibles como evidencia del proceso de organización y caracterización de los datos.

1.6.4.4. Conformación del conjunto de datos

Una vez culminado el procedimiento de recolección, generación y selección de las muestras, se conformó el conjunto de datos que se utilizará en esta investigación. Dicho procedimiento permitió agrupar la información procedente de las diferentes fuentes consideradas, obteniendo una estructura ordenada para su uso en los procedimientos posteriores. La distribución de las muestras seleccionadas se presenta en la **Tabla 1**.

Dataset	Fuente	Tipos de audios	Cantidad
Aishell	OpenSLR	Reales	14999
LibriSpeech	OpenSLR	Reales	14807
CommonVoice	OpenSLR	Reales	15000
VCTK Corpus	OpenSLR	Reales	15000
ASVSpooof 2019	Kaggle	Falsos	6000

Dataset	Fuente	Tipos de audios	Cantidad
Fake-or-Real	Kaggle	Falsos	3999
WaveFake	Kaggle	Falsos	5000
MLAAD	Hugging Face	Falsos	5998
CVoiceFake	Kaggle	Falsos	4000
LibriSevoc	OpenSLR	Falsos	3996
In-The-Wild	OpenSLR	Falsos	4000
DFADD	Hugging Face	Falsos	5000
PartialSpoof	Zenodo	Falsos	4000
Higgs Audio	Higgs Audio	Falsos	3000
Edge TTS	Edge TTS	Falsos	3000
Kokoro	Kokoro	Falsos	3000
Piper TTS	Piper TTS	Falsos	3000
Coqui TTS	Coqui TTS	Falsos	3000
ElevenLabs	ElevenLabs	Falsos	3000

Tabla 1. Distribución del conjunto de datos seleccionados

Como se observa en la **Tabla 1**, el conjunto de datos quedó conformado por muestras obtenidas de diferentes repositorios públicos y herramientas de síntesis de voz, las cuales fueron seleccionadas en función de las cantidades definidas en la planificación de la investigación; esta distribución permitió reunir fuentes con determinadas características acústicas y con diferentes tecnologías de síntesis de voz, lo que dio lugar a un conjunto de datos más representativo para el desarrollo y

evaluación del modelo propuesto.

La diversidad de las muestras integradas favorece la representación de diferentes tipos de escenarios analíticos, al incluir voces naturales y audios que han sido generados a partir de múltiples tecnologías de síntesis; esto permitió ampliar los patrones acústicos presentes en el conjunto de datos, contribuyendo a que el modelo adquirirá una mayor capacidad para distinguir entre audios reales y sintéticos durante el proceso de clasificación.

1.7. Metodología de Desarrollo

La metodología que orientará el desarrollo de este proyecto se construirá a partir de varias metodologías propuestas en investigaciones previas, de esta forma, se seleccionarán y ajustarán componentes metodológicos de varios proyectos de investigación relevantes [7], [8], [31], [41], con el objetivo de aprovechar sus fortalezas y conformar un marco metodológico propio que aborde de manera adecuada los requerimientos de la presente investigación.

Fases del Proyecto

Fase 1: Procesamiento de datos

- Recopilar audios auténticos y sintéticos procedentes de repositorios públicos y de herramientas de generación de voz a partir de IA.
- Estandarizar el dataset mediante la conversión de los audios a un formato homogéneo, frecuencia de muestreo de 16 kHz y canal mono.
- Eliminar los silencios de inicio y final, descartando los audios cuya duración final sea inferior a 1 segundo.
- Verificar que no existan filtraciones de información entre los conjuntos de datos.
- Separar el dataset en conjuntos de entrenamiento, validación y prueba, empleando una distribución del 70 %, 15 % y 15 % respectivamente.
- Registrar la información de los audios procesados a un archivo de metadatos para garantizar la trazabilidad del proceso.

Fase 2: Extracción de características acústicas

- Segmentar los archivos de audio en ventanas de tiempo para facilitar su

procesamiento en el entrenamiento del modelo.

- Normalizar la amplitud de las señales de audio para mitigar la variabilidad en las señales y favorecer la estabilidad del aprendizaje.
- Extraer espectrogramas log-Mel como representación espectral de las señales de audio.
- Incluir coeficientes delta y delta-delta para complementar la información temporal que contienen los espectrogramas.
- Aumentar las representaciones espectrales con técnicas de aumento de datos para mejorar la generalización del modelo.
- Almacenar las características acústicas generadas en un formato estructurado que optimice su uso en las etapas posteriores.



Fase 3: Desarrollo del modelo de clasificación

- Diseñar una arquitectura híbrida basada en redes neuronales convolucionales, redes LSTM bidireccionales y un mecanismo de atención para la tarea de clasificación de audios.
- Implementar la arquitectura propuesta para el procesamiento de las características acústicas obtenidas durante la etapa extracción.
- Configurar el proceso de entrenamiento mediante la selección de la función de pérdida, el optimizador elegido y los parámetros base del modelo.
- Incorporar técnicas de regularización orientadas a mitigar el sobreajuste y potenciar la estabilidad del proceso de aprendizaje.
- Entrenar el modelo utilizando los conjuntos de entrenamiento, validación y pruebas previamente definidos.
- Evaluar el rendimiento del modelo mediante métricas de clasificación y representaciones gráficas, con el objetivo de establecer un modelo base para la etapa optimización.



Fase 4: Optimización y ajuste de hiperparámetros

- Optimizar la arquitectura base mediante el ajuste de hiperparámetros y la implementación de mejoras para aumentar la capacidad de clasificación y el grado de generalización.
- Refinar la configuración del proceso de entrenamiento mediante la

modificación de la función de pérdida, el optimizador, la tasa de aprendizaje y las estrategias de regularización.

- Mejorar el procesamiento de las muestras implementando ventanas deslizantes, técnicas de normalización y estrategias de aumento de datos.
- Fortalecer las estrategias de atención y muestreo balanceado para mejorar la capacidad de discriminación del modelo y reducir el sesgo entre clases y conjuntos de datos.
- Evaluar el modelo optimizado mediante métricas de rendimiento, matrices de confusión, curvas ROC y tasa de error igual, seleccionando la configuración final para su integración en el sistema.

Fase 5: Implementación del sistema

- Desarrollar una plataforma web que permita la carga, análisis y clasificación de los audios a través de los módulos de autenticación y gestión de usuarios, carga y procesamiento de audios, análisis y clasificación, visualización de resultados y generación de reportes.
- Incorporar al sistema el modelo de clasificación optimizado para el análisis de audios y la estimación de la probabilidad de autenticidad de las muestras evaluadas.
- Validar el funcionamiento del sistema mediante pruebas con audios externos que no fueron incluidos durante la etapa de entrenamiento.
- Presentar los resultados del análisis mediante la probabilidad de clasificación, indicando si el audio corresponde a una muestra auténtica, un deepfake o un resultado sospechoso.
- Generar reportes en formato PDF que contengan la información del análisis, los resultados de la clasificación y las gráficas obtenidas durante la evaluación.

CAPITULO 2. PROPUESTA

2.1. MARCO CONTEXTUAL

En los años recientes, el progreso de la inteligencia artificial ha facilitado la creación de herramientas que generan representaciones audiovisuales muy realistas, imitando tanto el aspecto como la voz de las personas; estas simulaciones, llamadas deepfakes, pueden generarse a partir de fotografías, grabaciones y sonidos, lo que complica su detección por su elevado grado de veracidad; el uso de estas tecnologías se ha incrementado considerablemente en todo el mundo, motivado por la existencia de aplicaciones accesibles y sencillas que facilitan la producción de contenido sintético sin necesidad de habilidades técnicas avanzadas; este escenario ha aumentado los peligros relacionados con la suplantación de identidad, la manipulación de datos y la violación de la privacidad, impactando en la reputación de personas y empresas [23], [24], [42], [43].

Además, los deepfakes han pasado de ser utilizados únicamente con fines de entretenimiento a ser empleados de forma maliciosa, como en la propagación de desinformación, estafas en línea y la suplantación de identidad, lo que ejemplifica su efecto adverso en el entorno digital; en los últimos años, también se ha registrado un aumento significativo en el uso de dichas tecnologías con fines engañosos; dicho incremento, se reflejó entre 2023 y 2024 en casos de fraude financiero y en la amplia circulación en línea de contenido falso, lo que confirma que los deepfakes constituyen una amenaza latente a nivel mundial [23], [24], [44], [45].

En el caso de Ecuador, la adjudicación de tecnologías de inteligencia artificial para la generación de deepfakes ha adquirido relevancia en los últimos años, manifestando su lado más perjudicial en la coyuntura de las elecciones de 2023 y 2025, en las que se evidenció una notable difusión de contenido fake, provocando confusión entre la población y acentuando la polarización política, debido a su rápida circulación en redes sociales; por otra parte, casos como la difusión de información falsa sobre el Fondo Monetario Internacional muestran el impacto de estas tecnologías en la percepción social [46], [47].

En este contexto, el uso indebido de los deepfakes constituye una amenaza

significativa para derechos fundamentales como la privacidad, el honor, la imagen y la identidad de las personas, lo cual refleja la falta de regulación en la normativa ecuatoriana actual, pues no existe una regulación específica que aborde este tipo de situaciones, lo que impide la posibilidad de sancionar la conducta infractora y ofrecer la debida protección a las víctimas, incrementando el riesgo de impunidad; en consecuencia, el desafío no se limita únicamente al ámbito tecnológico, sino que también plantea implicaciones jurídicas y sociales que deben ser abordadas de forma integral [48], [49], [50], [51].

Frente a este panorama, la implementación del presente proyecto se llevará a cabo en un entorno simulado, utilizando pruebas experimentales basadas en conjuntos de datos etiquetados previamente como audios reales y falsos. Este enfoque permitirá realizar análisis comparativos orientados a evaluar la eficacia del sistema en términos de precisión y capacidad para detectar posibles alteraciones en los archivos de audio. En este contexto, el uso de datos organizados será indispensable para el entrenamiento y validación de los algoritmos propuestos, garantizando que el modelo pueda distinguir de manera efectiva entre contenido auténtico y manipulado; además, el entorno de simulación permitirá reproducir condiciones representativas del escenario de aplicación dentro de un entorno controlado, facilitando la validación funcional y el rendimiento del sistema sin comprometer datos sensibles, lo que permitirá garantizar la confiabilidad y la consistencia de los resultados obtenidos.

Bajo esta perspectiva, la detección de deepfakes constituye un desafío relevante en el área tecnológica, debido al elevado nivel de realismo que alcanzan estas alteraciones, lo que dificulta distinguirlos de contenido auténtico mediante la inspección humana; a ello se suma la escasa disponibilidad de herramientas especializadas para detectar este tipo de alteraciones de forma precisa y accesible [52]. Frente a esta problemática, resulta necesario desarrollar modelos de deep learning que sean capaces de adaptarse al avance de las técnicas de síntesis de voz [23], [24]. Bajo este enfoque, las arquitecturas híbridas que integran CNN y LSTM representan una alternativa adecuada para fortalecer la detección de audios sintéticos frente a las limitaciones de los métodos convencionales y favorecen una

mejor generalización frente a diferentes técnicas de síntesis de voz [7], [8].

2.2. MARCO CONCEPTUAL

2.2.1. Inteligencia Artificial

Se entiende como la capacidad de los sistemas para realizar tareas que comúnmente requieren la inteligencia humana; esto se logra mediante la aplicación de algoritmos, la recolección de datos y la aplicación de dicha información en el proceso de toma de decisiones; también, incluye modelos conceptuales, conjuntos de datos, razonamiento automatizado, enfoques heurísticos y aprendizaje profundo [53], [54], [55]. En el presente estudio, la inteligencia artificial constituye el pilar tecnológico que sustenta el desarrollo de la solución propuesta, ya que brinda las capacidades necesarias para el procesamiento automático y el análisis eficiente de la información [56], [57], [58], [59].

2.2.1.1. Aprendizaje Automático

Es una rama de la inteligencia artificial que permite a las máquinas aprender de la información sin programación explícita; consiste en la detección automática de patrones y la mejora de la eficacia de un sistema a través de la experiencia; se basa en el uso de algoritmos que procesan grandes volúmenes de datos para extraer información útil; y se ha consolidado como un recurso crucial en tareas que requieren el análisis y la interpretación de datos [60], [61], [62], [63]. En este proyecto, el aprendizaje automático proporciona el enfoque metodológico para desarrollar modelos capaces de identificar patrones en los datos y mejorar progresivamente su desempeño conforme avanza el proceso de entrenamiento [58], [59], [64], [65].

2.2.1.1.1. Deep Learning

Es una rama del aprendizaje automático que permite a las máquinas aprender de la experiencia y entender el entorno a través de una jerarquía de conceptos, donde los conceptos más complejos se construyen a partir de los más simples, sin necesidad de que un operador humano defina toda la base de conocimiento necesaria [66]. Deep Learning permite poder modelar las relaciones complejas presentes en grandes volúmenes de datos mediante redes neuronales profundas, favoreciendo el

desarrollo de modelos con un mayor nivel de abstracción y de discriminación, lo que lo convierte en un enfoque adecuado para poder hacer frente a problemas de clasificación que requieren identificar patrones difíciles de representar mediante métodos tradicionales [4].

2.2.2. Redes Neuronales

Se definen como representaciones matemáticas que imitan los modos de funcionamiento del cerebro humano, se componen de partes interconectadas que gestionan información de forma activa y ofrecen la posibilidad de un aprendizaje propio basado en datos y experiencias; de esta forma, son capaces de realizar cálculos basados en conceptos biológicos para resolver problemas de una manera efectiva [67], [68], [69]. Las redes neuronales representan un elemento clave en la solución propuesta, ya que permiten transformar los datos de entrada en representaciones cada vez más abstractas a partir de un proceso conformado por múltiples capas, lo que facilita la identificación de patrones complejos y el desarrollo de soluciones con un alto grado de automatización [9], [70], [71].

2.2.2.1. Redes Neuronales Convolucionales (CNN)

Son un tipo de red neuronal orientada al procesamiento de datos con forma de rejilla, como imágenes o espectrogramas de audio; para ello, incorporan capas convolucionales que emplean filtros para extraer características locales y espaciales, facilitando la identificación de patrones relevantes presentes en las señales, como frecuencias y componentes espectrales [9]. En este trabajo, se emplean redes neuronales convolucionales debido a su capacidad para aprender automáticamente las características más relevantes de los datos de entrada mediante la aplicación de filtros convolucionales, lo que permite extraer patrones representativos de la información en diversos niveles de abstracción y favorece el desarrollo de un modelo con una mayor capacidad discriminativa para la tarea de clasificación [31].

2.2.2.2. Redes de Memoria a Corto y Largo Plazo (LSTM)

Son un tipo de red neuronal recurrente diseñadas para gestionar datos secuenciales y dependencias a largo plazo, en tareas de aprendizaje automático, como en la

traducción automática, generación de texto y reconocimiento de voz, y a diferencia de las redes recurrentes tradicionales, integran celdas de memoria que permiten retener información relevante por largos períodos [72], [73]. En la presente investigación, las redes de memoria a corto y largo se emplean con el propósito de modelar la información temporal presente en las señales acústicas, facilitando el aprendizaje de patrones secuenciales que contribuyen a diferenciar entre audios auténticos y manipulados [1].

2.2.2.3. Modelo Híbrido CNN-LSTM

Es un enfoque de aprendizaje profundo que emplea redes neuronales convolucionales y redes de memoria a corto y largo plazo; las CNN permiten extraer las características espaciales de los datos de entrada, mientras que las redes LSTM se enfocan en las dependencias temporales de la información, logrando una representación más completa de los datos [7], [8], [74], [75]. En la presente investigación, se implementó un modelo híbrido CNN-LSTM, con el objetivo de integrar las potencialidades de las dos arquitecturas; esta integración permite aprovechar las características espaciales y temporales de las señales de audio, lo que a su vez mejora a la clasificación y la discriminación de audio.

2.2.3. Procesamiento de Audio

Es una rama del procesamiento de señales digitales que tiene como objetivo estudiar las características matemáticas de la señal de audio digital, permitiendo así la extracción, el análisis y la modificación de sus atributos; esta rama comprende diferentes tipos de fuentes fonéticas, incluyendo la música, el ruido ambiental y la voz humana, y ha cobrado mucha importancia gracias al desarrollo de las tecnologías de reconocimiento de voz y los asistentes virtuales [76]. Dentro del flujo de procesamiento establecido, los archivos de audio se transforman y adaptan en varias etapas que permiten obtener una representación homogénea de las señales; estas operaciones preparan los datos de entrada que el modelo utiliza para el análisis de la información acústica existente en cada muestra de audio [70], [71].

2.2.3.1. Análisis Espectral

Es un método que se utiliza para evaluar la intensidad de una señal con relación a

su frecuencia, a través del análisis de su espectro o la densidad espectral de potencia; este proceso, que a menudo se realiza utilizando la transformada de Fourier y su versión optimizada, la FFT, permite identificar cómo se distribuye la energía dentro de una señal; sin embargo, la precisión de este tipo de análisis depende del número y la calidad de las muestras disponibles [77], [78]. El análisis espectral resulta valioso para identificar deepfakes de audio, ya que ayuda a revelar patrones y características en el espectro de frecuencias que diferencian la voz generada artificialmente de la voz auténtica; además, esto contribuye a mejorar la precisión de modelos combinados como el CNN-LSTM, que fusionan componentes espectrales y temporales para lograr una identificación más sólida y confiable [79], [80].

2.2.3.2. Espectrograma en escala Mel

Es una representación en el dominio tiempo–frecuencia de la señal de audio donde las frecuencias se organizan según la escala Mel, basada en cómo percibe el oído humano, permitiendo así una disposición no lineal en el eje de frecuencias; además, puede ser presentada de manera logarítmica (log-Mel), mostrando la amplitud en decibelios para enfatizar cambios significativos en la energía de la señal y simplificar el estudio de los patrones acústicos [81], [82]. La generación de espectrogramas log-Mel permite convertir cada uno de los fragmentos de audio en una representación tiempo-frecuencia, adaptada para el procesamiento computacional; esta transformación, junto con las operaciones de preparación aplicadas sobre la señal, permiten obtener una representación consistente de las muestras de audio y preservar información acústica relevante para el análisis posterior [9].

2.2.3.3. Coeficientes delta y delta-delta

Los coeficientes delta son una forma de medir cómo cambian las propiedades acústicas de una señal de audio a lo largo del tiempo, miden la tasa de cambio entre momentos consecutivos y brindan información dinámica o de velocidad; por otro lado, los coeficientes delta-delta están relacionados con la segunda derivada de estas propiedades, es decir, miden cómo varían los coeficientes delta; esto nos permite captar cambios de aceleración en la señal, mejorando así la representación

temporal del audio [83], [84], [85]. Los coeficientes delta y delta-delta se utilizan para registrar la dinámica temporal de la señal de audio a través de sus cambios de primer y segundo orden, complementando la representación espectral, lo que contribuye a enriquecer los datos utilizados en las etapas posteriores del sistema [83], [86], [87].

2.2.4. Deepfakes

Son materiales en formato de audio o vídeo que han sido alterados con elementos de IA, con el objetivo de generar la percepción de que una persona dijo o hizo algo que nunca ocurrió; este tipo de material puede generar consecuencias graves e incluso comprometer la reputación de personas o provocar conflictos de tipo político o social [88], [89]. En este contexto, los deepfakes constituyen el fenómeno de estudio sobre el cual se desarrolla la investigación, al representar una forma de manipulación multimedia cuya expansión ha impulsado la necesidad de desarrollar mecanismos automáticos para su detección [7], [8].

2.2.4.1. Deepfakes de audio

Son sonidos producidos a través de métodos de síntesis o conversión de voz, pensados para replicar el habla humana con un alto nivel de realismo, ya sea copiando una voz específica o alterando el contenido original de una grabación; estas tecnologías facilitan la producción de audios que pueden parecer reales, aunque no tengan su origen en una fuente humana real [5], [6]. En este trabajo, los deepfakes de audio representan el objeto de análisis del sistema desarrollado, al constituir las muestras manipuladas que permiten evaluar la capacidad del modelo para diferenciar entre audios reales y manipulados [58], [59].

2.3. MARCO TEÓRICO

2.3.1. Enfoque de deep learning para la detección de audio sintético

La teoría del modelado profundo sostiene que las arquitecturas de aprendizaje profundo permiten la detección de audios manipulados porque aprenden patrones acústicos complejos capaces de distinguir entre el habla real y la sintética; entre estas arquitecturas destacan las redes neuronales convolucionales (CNN) y las redes de memoria a corto y largo plazo (LSTM), las cuales examinan y procesan las

representaciones espectrales y temporales del sonido, lo que permite obtener una representación más robusta de la señal y, por tanto, una mayor capacidad para la detección de deepfakes de audio [90], [91].

Estudios recientes respaldan esta estrategia al comprobar que los modelos basados en Deep Learning son capaces de incrementar el desempeño de los algoritmos convencionales, alcanzando precisiones superiores al 97 % en la detección de audios sintéticos, lo que valida la eficacia del aprendizaje profundo para el desarrollo de sistemas automáticos de autenticación de la voz; en consecuencia, estos resultados evidencian el potencial que pueden tener las arquitecturas profundas para la detección y clasificación de audio generado artificialmente. [92], [93].

En este contexto, existen varios enfoques de deep learning que se han desarrollado para detectar audios sintéticos [4], [65]. Donde los MLP, son rápidos y sencillos de entrenar, las CNNs son efectivas para detectar patrones claves en espectrogramas y finalmente las LSTM permiten modelar las dependencias temporales presentes en la señal de audio; como resultado, estas estructuras permiten realizar un análisis exhaustivo de representaciones complejas del audio y son consideradas como un método eficaz para detectar contenido generado de forma artificial [18].

2.3.2. Procesamiento de audio y extracción de características acústicas

La teoría del preprocesamiento y la extracción de características acústicas establece las bases necesarias para convertir las señales de voz en representaciones numéricas que los modelos de aprendizaje automático y profundo pueden entender; en la identificación de deepfakes de audio, este proceso es crucial para reconocer patrones sutiles que diferencian el habla genuina de la artificial; mediante técnicas como la normalización, los espectrogramas Mel y los coeficientes MFCC, se generan representaciones tanto temporales como espectrales del audio que mejoran la precisión en la detección de audios alterados [94], [95].

Bajo esta perspectiva, el estudio de las propiedades del sonido facilita entender con mayor profundidad el comportamiento de una señal de audio; para ello, se consideran aspectos como los MFCC, la energía, el centroide y el ancho de banda

espectral, entre otros; los cuales reflejan la estructura espectral y temporal de la señal de audio, la selección correcta de estas características es crucial para potenciar la detección de deepfakes de audio, puesto que, el uso de los MFCC, junto con sus variaciones a lo largo del tiempo y otros descriptores espectrales, brindan información significativa sobre el comportamiento de la señal, contribuyendo a una mayor precisión en la clasificación y a un uso más eficaz de los recursos computacionales [70], [71], [96], [97].

2.3.3. Arquitecturas híbridas de aprendizaje profundo para clasificación de audio

La teoría del aprendizaje colaborativo en modelos profundos propone que la combinación de diferentes arquitecturas de redes neuronales mejora la precisión y la robustez en la clasificación de señales de audio; este método reúne modelos con habilidades complementarias, como las redes convolucionales (CNN y ResNet) que extraen rasgos espectrales y las redes recurrentes (LSTM) que analizan dependencias temporales; la combinación de los resultados de varios modelos facilita una interpretación más integral de la señal, disminuyendo los errores de clasificación y optimizando la detección de patrones complejos. [98].

De manera similar, la teoría propone que la aplicación de métodos de fusión tardía o en niveles intermedios facilita la combinación adaptable de las salidas de diferentes modelos, utilizando su diversidad para integrar información complementaria y redundante, la cual produce representaciones más precisas del audio, mejorando la habilidad de los sistemas para diferenciar entre señales genuinas y alteradas, lo que fortalece su desempeño en aplicaciones de detección forense y seguridad [99], [100].

2.4. Requerimientos

2.4.1. Requerimientos Funcionales

2.4.1.1. Módulo de Autenticación y Gestión de Usuarios

Código	Especificación de requisitos
RF-01	El sistema debe permitir al usuario iniciar sesión mediante el ingreso

Código	Especificación de requisitos
	de sus credenciales
RF-02	El sistema debe validar las credenciales ingresadas por el usuario mediante la API REST
RF-03	El sistema debe generar un token JWT cuando las credenciales proporcionadas sean válidas
RF-04	El sistema debe restringir el acceso a las funcionalidades protegidas a los usuarios autenticados
RF-05	El sistema debe gestionar la información de los usuarios registrados en la base de datos
RF-06	El sistema debe asociar cada análisis realizado con el usuario autenticado que lo ejecutó

Tabla 2. Requerimiento Funcional - Módulo de Autenticación y Gestión de Usuarios

2.4.1.2. Módulo de Carga y Procesamiento de Audios

Código	Especificación de requisitos
RF-07	El sistema debe permitir al usuario cargar archivos de audio para su análisis
RF-08	El sistema debe validar que el archivo de audio cargado tenga un formato permitido
RF-09	El sistema debe validar que la duración del archivo de audio se encuentre entre 4 segundos y 5 minutos
RF-10	El sistema debe rechazar los archivos que no cumplan con los requisitos de formato o duración establecidos

Código	Especificación de requisitos
RF-11	El sistema debe enviar los archivos de audio válidos al modelo CNN-LSTM desplegado en Hugging Face
RF-12	El sistema debe gestionar la comunicación entre la API REST y el modelo de inteligencia artificial para el procesamiento de los audios

Tabla 3. Requerimiento Funcional - Módulo de Carga y Procesamiento de Audios

2.4.1.3. Módulo de Análisis y Clasificación de Audios

Código	Especificación de requisitos
RF-13	El sistema debe procesar la muestra de audio mediante el modelo CNN-LSTM para obtener una predicción
RF-14	El sistema debe clasificar la muestra de audio en una de las categorías: Real, Deepfake o Sospechoso
RF-15	El sistema debe asignar la categoría Sospechoso cuando el nivel de confianza del modelo no permita determinar de manera concluyente si la muestra corresponde a un audio Real o Deepfake
RF-16	El sistema debe recibir y gestionar el resultado de la clasificación generado por el modelo de inteligencia artificial
RF-17	El sistema debe registrar los resultados de la clasificación en la base de datos MySQL alojada en AlwaysData.
RF-18	El sistema debe conservar la relación entre el usuario autenticado, la muestra analizada y el resultado obtenido

Tabla 4. Requerimiento Funcional - Módulo de Análisis y Clasificación de Audios

2.4.1.4. Módulo de Visualización de Resultados

Código	Especificación de requisitos
RF-19	El sistema debe mostrar al usuario el resultado de la clasificación obtenida para la muestra de audio analizada
RF-20	El sistema debe mostrar la categoría asignada a la muestra de audio: Real, Deepfake o Sospechoso
RF-21	El sistema debe permitir al usuario consultar la información asociada al análisis realizado

Tabla 5. Requerimiento Funcional - Módulo de Visualización de Resultados

2.4.1.5. Módulo de Generación y Descarga de Reportes

Código	Especificación de requisitos
RF-22	El sistema debe generar un reporte en formato PDF con los resultados del análisis realizado
RF-23	El sistema debe almacenar los reportes PDF generados en Cloudinary
RF-24	El sistema debe registrar en la base de datos la URL segura (secure_url) y el public_id asociados al reporte almacenado
RF-25	El sistema debe poner generado a disposición del usuario mediante la interfaz web
RF-26	El sistema debe permitir al usuario descargar el reporte PDF correspondiente al análisis realizado

Tabla 6. Requerimiento Funcional - Módulo de Generación y Descarga de Reportes

2.4.2. Requerimientos no Funcionales

Código	Especificación de requisitos
RNF-01	El sistema debe utilizar autenticación basada en tokens JWT para controlar el acceso a las funcionalidades protegidas
RNF-02	El sistema debe validar la autenticidad del token JWT en las solicitudes que requieren autenticación
RNF-03	La comunicación entre la aplicación web y la API REST debe realizarse mediante solicitudes HTTP/HTTPS
RNF-04	El sistema debe mantener una arquitectura modular que permita separar las responsabilidades de la aplicación web, la API REST, el modelo de inteligencia artificial, la base de datos y el almacenamiento de archivos
RNF-05	El sistema debe permitir la comunicación entre la API REST y el modelo CNN-LSTM desplegado en hugging Face para el procesamiento de los audios
RNF-06	El sistema debe conservar la integridad de la información asociada a los análisis, relacionando los resultados obtenidos con el usuario que realizó el proceso
RNF-07	El sistema debe garantizar la persistencia de los resultados de clasificación almacenados en la base de datos MySQL alojada en AlwaysData
RNF-08	El sistema debe garantizar la disponibilidad de los reportes PDF almacenados en Cloudinary mediante la URL segura registrada en la base de datos
RNF-09	El sistema debe gestionar los errores que puedan producirse durante

Código	Especificación de requisitos
	la comunicación entre la aplicación web, la API REST y el modelo inteligencia artificial
RNF-10	El sistema debe informar al usuario cuando no sea posible completar una operación debido a errores durante el procesamiento, generación, almacenamiento o descarga de los reportes
RNF-11	El sistema debe proporcionar una interfaz web que permita al usuario interactuar con las funciones de autenticación, carga de audios, visualización de resultados y descarga de reportes
RNF-12	El sistema debe mantener la separación entre el almacenamiento de los datos de los análisis en la base de datos y el almacenamiento de los archivos PDF en Cloudinary

Tabla 7. Requerimientos No Funcionales del Sistema

2.5. Componentes de la Propuesta

2.5.1. Arquitectura del Sistema

La **Figura 1** presenta la arquitectura general del sistema propuesto, conformada por cinco capas funcionales: capa de presentación, capa de servicios, capa de inteligencia artificial, capa de almacenamiento de archivos y capa de persistencia de datos. Esta estructura distribuye las responsabilidades que tiene cada uno de los componentes y define el flujo de comunicación entre la aplicación web, la API REST, el modelo de inteligencia artificial, el servicio de almacenamiento de archivos y la base de datos; de esta manera, se establece una arquitectura modular que favorece la integración de los procesos que intervienen en la detección de los audios deepfake, favoreciendo un mantenimiento más sencillo, una mejor escalabilidad y la incorporación de futuras mejoras sin impactar en el funcionamiento del resto de los componentes. Asimismo, esta organización facilita la administración del sistema y la interacción eficiente entre sus diferentes capas funcionales.

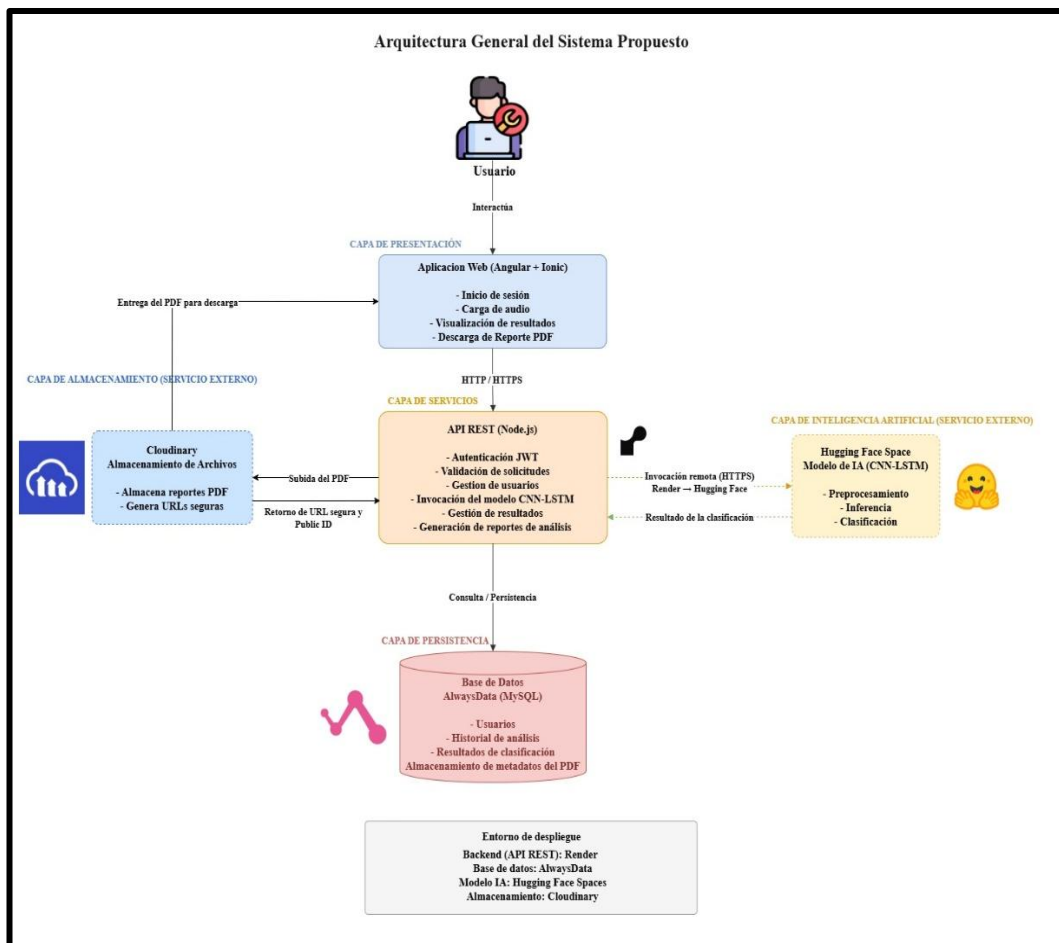


Figura 1. Arquitectura General del Sistema Propuesto

La capa de presentación corresponde a la aplicación web GlowVox, desarrollada con Angular e Ionic, la cual constituye el punto de contacto entre el usuario y el sistema; a través de esta interfaz, se gestionan acciones como el registro, el inicio de sesión, la carga de archivos de audio, la consulta del historial de análisis, la visualización de los resultados de clasificación o la descarga de los reportes generados en formato PDF. Todas las peticiones del usuario son enviadas a la API REST mediante los protocolos HTTP/HTTPS, lo que garantiza una comunicación continua y segura con los servicios que se encargan del procesamiento de la información.

La capa de servicios está implementada mediante una API REST, desarrollada en Node.js y desplegada en Render, encargada de centralizar la lógica de negocio y coordinar el flujo de ejecución del sistema. Entre las principales funcionalidades

que realiza la API REST se encuentran la autenticación de usuarios mediante tokens JWT, la validación de las solicitudes que recibe, la gestión de usuarios y la administración de los procesos asociados al análisis de los audios.

Una vez recibido un archivo, la API REST lo envía al modelo de inteligencia artificial desplegado en Hugging Face Spaces para realizar la inferencia correspondiente. Posteriormente, recibe el resultado de la clasificación, genera el reporte de análisis en formato PDF, gestiona su almacenamiento en Cloudinary y realiza el registro de la información generada durante el proceso en la base de datos, actuando como el componente encargado de integrar y orquestar el funcionamiento de todo el sistema.

La capa de inteligencia artificial integra el modelo CNN-LSTM desarrollado en Python utilizando la biblioteca PyTorch, el cual se desplegó como un servicio de inferencia en Hugging Face Spaces. Este elemento recibe los audios que la API REST le envía, aplica toda la lógica de preprocesamiento, extracción de características, inferencia y clasificación, y devuelve la predicción correspondiente, indicando si el audio analizado es auténtico o sintético. Al encontrarse desacoplado de la lógica de negocio, el modelo puede ser reemplazado en cualquier momento, sin afectar el funcionamiento de los demás componentes, lo que permite mejorar la mantenibilidad y la escalabilidad del sistema.

La capa de almacenamiento utiliza Cloudinary como servicio destinado para el almacenamiento de los informes de análisis generados por la API REST. Una vez que el fichero ha sido correctamente almacenado, Cloudinary devuelve la URL segura (`secure_url`) y el ID único (`public_id`), los cuales permiten acceder al reporte de forma segura y facilitan su administración dentro de la plataforma. Esta capa desacopla el almacenamiento de los ficheros de la lógica de negocio, garantizando una gestión eficiente de los ficheros generados durante el proceso de análisis.

Finalmente, la capa de persistencia de datos está implementada mediante una base de datos MySQL hospedada en AlwaysData, responsable de almacenar la información estructurada del sistema. En la base de datos se registran los usuarios, las credenciales de acceso, el historial de análisis, los resultados obtenidos en cada clasificación y los metadatos asociados a los reportes generados, incluyendo la URL

segura y el ID único que proporciona Cloudinary. De esta manera, el sistema mantiene un registro persistente de las operaciones realizadas, facilitando la consulta posterior de los análisis efectuados por cada usuario y garantiza la integridad de la información almacenada

2.5.2. Diagramas de Casos de Uso

Para representar las principales interacciones entre el usuario y el sistema de detección de deepfakes de audio, se utilizan diagramas de casos de uso y tablas de especificación; los diagramas permiten observar las funcionalidades principales y las relaciones entre el usuario y los procesos que intervienen en la autenticación y en el análisis de los archivos de audio, mientras que las tablas especifican con detalles el objetivo, las precondiciones, el flujo principal, los flujos alternativos y las postcondiciones de cada caso de uso; en este contexto, se pudieron establecer dos casos de uso principales: Iniciar sesión y Analizar audio, donde el caso de uso de analizar audio incluye las actividades vinculadas a la clasificación, el almacenamiento de resultados, la generación y descarga del reporte PDF.

2.5.2.1. Caso de Uso 1: Inicio de Sesión

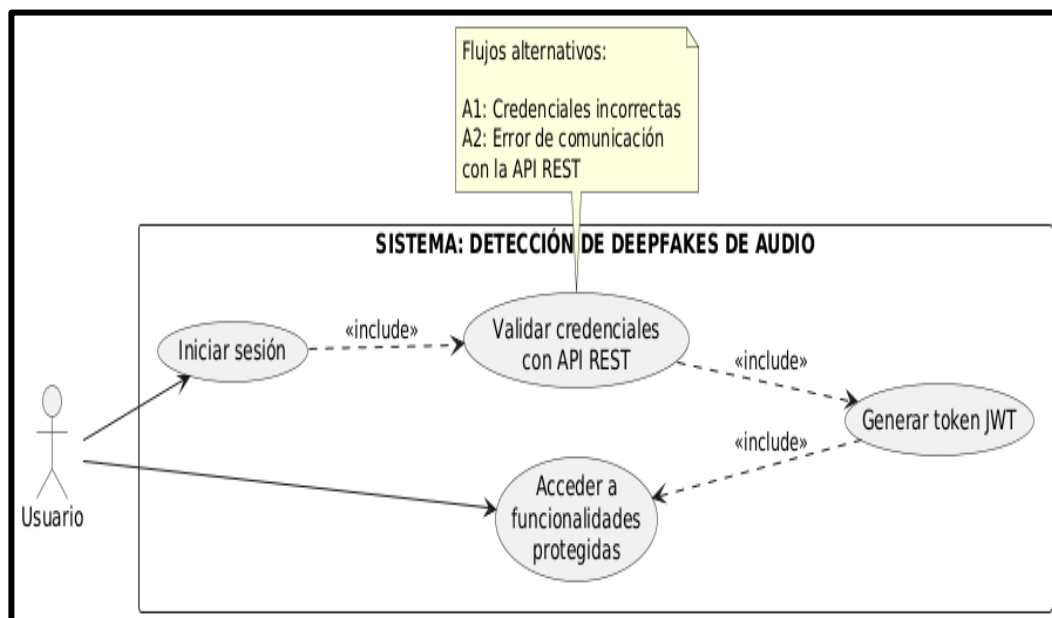


Figura 2. Diagrama del caso de uso de inicio de sesión

Elemento	Descripción
Caso de uso	Iniciar sesión
Descripción	Permite al usuario autenticarse en el sistema mediante la validación de sus credenciales y la generación de un token JWT para acceder a las funcionalidades protegidas
Actor	Usuario
Evento desencadenador	El usuario ingresa sus credenciales y selecciona la opción de inicio de sesión
Precondición	El usuario debe estar registrado en el sistema y disponer de credenciales válidas
Secuencia principal	1. El usuario ingresa sus credenciales. 2. El sistema envía las credenciales a la API REST. 3. La API REST valida las credenciales proporcionadas. 4. Si las credenciales son válidas, la API REST genera un token JWT. 5. El sistema concede al usuario el acceso a las funcionalidades protegidas. 6. El usuario puede acceder al módulo de análisis de audio.
Errores/ Alternativas	A1. Credenciales incorrectas: la API REST rechaza las credenciales y el sistema informa al usuario que los datos ingresados no son válidos. A2. Error de comunicación con la API REST: el sistema informa que no fue posible completar el proceso completar el proceso de autenticación.
Postcondición	El usuario queda autenticado y puede acceder a las funcionalidades protegidas mediante un token JWT válido.

Tabla 8. Especificación del caso de uso de inicio de sesión

2.5.2.2. Caso de uso 2: Análisis de audio para la detección de deepfakes

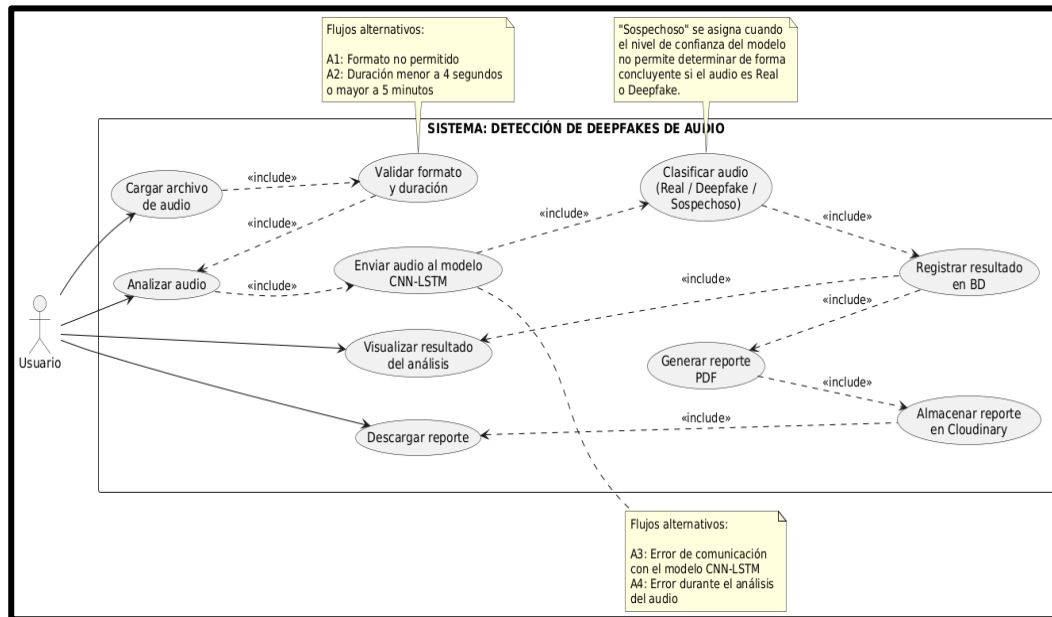


Figura 3. Diagrama del caso de uso de análisis de audio para la detección de deepfakes

Elemento	Descripción
Caso de uso	Análisis de audio para la detección de deepfakes.
Descripción	Permite al usuario cargar un archivo de audio para analizarlo mediante un modelo CNN-LSTM, obtener una clasificación como Real, Deepfake o Sospechoso, consultar los resultados y generar un reporte PDF para su descarga
Actor	Usuario
Evento desencadenador	El usuario selecciona un archivo de audio e inicia el proceso de análisis
Precondición	El usuario debe haber iniciado sesión correctamente y contar con un token JWT válido
Secuencia principal	1. El usuario carga un archivo de audio. 2. El sistema valida el formato y la duración del archivo.

Elemento	Descripción
	3. El usuario inicia el análisis del audio. 4. La API REST envía el audio al modelo CNN-LSTM desplegado en Hugging Face. 5. El modelo procesa el audio y ejecuta la clasificación. 6. El audio es clasificado como Real, Deepfake o Sospechoso. 7. La API REST recibe el resultado de la clasificación. 8. El sistema registra el resultado del análisis en la base de datos. 9. El sistema muestra al usuario el resultado del análisis. 10. El sistema genera un reporte en formato PDF. 11. El reporte se almacena en Cloudinary. 12. Cloudinary devuelve la URL segura y el public_id del reporte. 13. El sistema registra la URL segura y el public_id en la base de datos. 14. El reporte queda disponible para que el usuario pueda descargarlo. 15. El usuario descarga el reporte PDF.
Errores/Alternativas	A1. Formato no permitido: el sistema rechaza el archivo e informa que el formato no es compatible A2. Duración no válida: el sistema rechaza al archivo cuando su duración es menor a 4 segundos o mayor a 5 minutos A3. Error de comunicación con el modelo CNN-LSTM: la API REST informa que no fue posible comunicarse con el modelo y el análisis no se completa A4. Error durante el análisis del audio: el sistema informa que ocurrió un error durante el procesamiento o la clasificación A5. Clasificación con confianza insuficiente: el sistema

Elemento	Descripción
	asigna la categoría Sospechoso cuando el modelo no tiene suficiente confianza para determinar de manera concluyente si el audio es Real o Deepfake A6. Error al generar el reporte PDF: el sistema informa que no fue posible generar el reporte A7. Error al almacenar el reporte: el sistema informa que el reporte no pudo ser almacenado en Cloudinary A8. Error al registrar los metadatos: el sistema informa que no fue posible registrar la URL segura o el public_id en la base de datos A9. Error durante la descarga: el sistema informa que el reporte no está disponible o que la descarga no pudo completarse
Postcondición	El resultado de la clasificación queda almacenado en la base de datos. El reporte PDF se almacena en Cloudinary y su URL segura y public_id quedan registrados en la base de datos. Finalmente, el reporte queda disponible para su descarga por parte del usuario

Tabla 9. Especificación del caso de uso de análisis de audio para la detección de deepfakes

2.5.3. Modelado de Datos

El modelado de datos permite representar la estructura de la información gestionada por el sistema de detección de deepfakes de audio, así como también las relaciones establecidas entre las entidades más relevantes del mismo; para ello, se ha diseñado un modelo relacional orientado a organizar la información de los usuarios, los análisis realizados, la generación de reportes y las actividades registradas durante el uso de la plataforma, garantizando la integridad y consistencia de los datos, así como la correcta implementación de las relaciones entre las entidades del sistema.

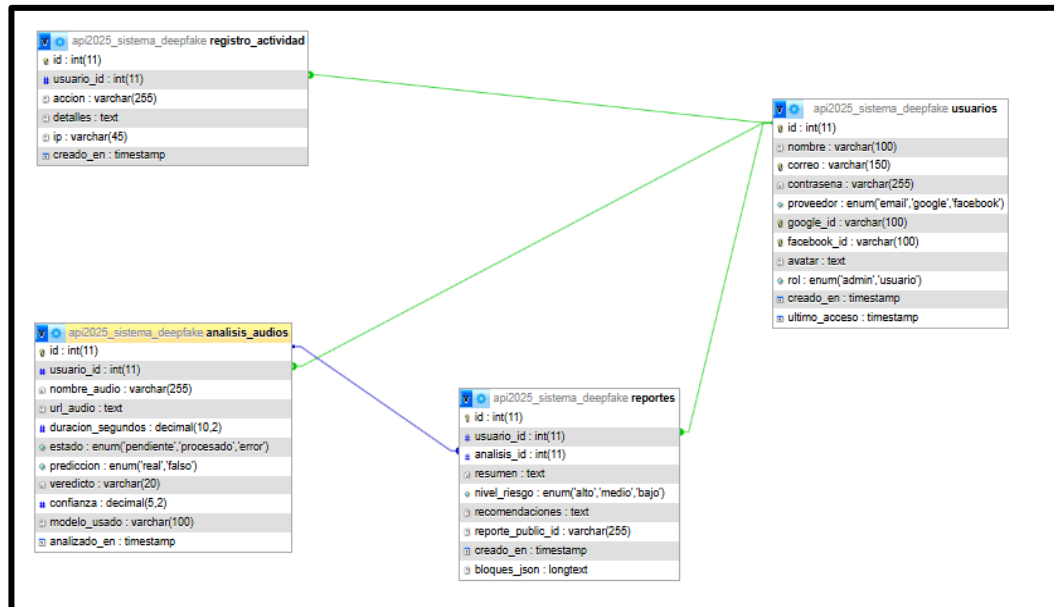


Figura 4. Modelo Relacional de la Base de Datos del Sistema

Como se muestra en la **Figura 4**, el modelo relacional está conformado por las entidades usuarios, analisis_audios, reportes y registro_actividad. La entidad usuarios constituye el eje principal del modelo, ya que se relaciona con los análisis de audio realizados, los reportes generados y las actividades ejecutadas en el sistema. Estas relaciones nos permiten asociar cada operación con el usuario correspondiente y mantener la trazabilidad de la información almacenada.

Por su parte, la entidad analisis_audios almacena la información relacionada con las muestras procesadas y los resultados obtenidos tras el proceso de clasificación. Esta entidad considera datos referentes al archivo analizado, su duración, el estado del proceso de análisis, la predicción generada, el veredicto y el nivel de confianza del modelo. Por otro lado, la entidad reportes permite la relación entre los resultados de los análisis y los informes generados a partir de ellos, conservando la información requerida para su gestión dentro del sistema.

Finalmente, la entidad registro_actividad almacena las acciones ejecutadas por los usuarios y los detalles asociados a cada operación. La organización de estas entidades permite establecer una relación lógica entre el usuario que realiza una acción, el análisis de audio que se ejecuta, el resultado de clasificación que se obtiene y el reporte que, posteriormente se genera. De esta forma, el modelo

relacional proporciona una estructura organizada para gestionar la información generada por la plataforma durante el proceso de detección.

2.6. Diseño de Interfaces

El diseño de interfaces se orientó a facilitar la interacción del usuario con las principales capacidades del sistema de detección de deepfakes de audio. La aplicación web fue estructurada para acompañar al usuario de manera progresiva durante las principales etapas del proceso; de este modo, el flujo inicia con la autenticación y continúa con la carga y análisis de la muestra, la visualización de los resultados, para finalmente permitir la descarga del reporte generado.

2.6.1. Interfaz de registro e inicio de sesión

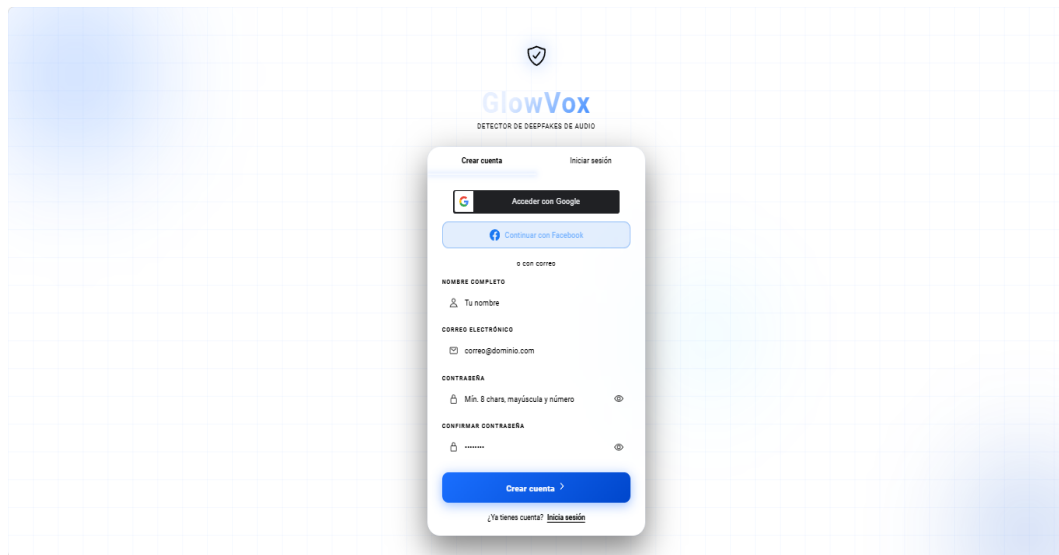


Figura 5. Interfaz de registro e inicio de sesión

La **Figura 5** presenta la interfaz de registro e inicio de sesión, la cual constituye el punto de acceso a la plataforma. En esta interfaz, el usuario puede autenticarse con sus credenciales o podrá registrar una nueva cuenta para utilizar las funcionalidades del sistema. La vista incluye formularios de entrada y controles de interfaz para validar los datos proporcionados por el usuario, permitiendo gestionar el acceso de forma controlada; una vez completado este proceso, el usuario puede acceder a la plataforma y continuar con el flujo de trabajo establecido. A partir de esta interfaz, se establece el inicio de la interacción del usuario con los demás componentes de la plataforma.

Elementos principales:

- **Formulario de inicio de sesión:** Permite introducir las credenciales de acceso.
- **Formulario de registro:** Permite crear una nueva cuenta de usuario.
- **Campos de entrada:** Permiten introducir la información necesaria para cada proceso.
- **Botones de acción:** Permiten iniciar la sesión o finalizar el registro.
- **Mensajes de validación:** Informa en caso de error relacionado con los datos introducidos.

2.6.2. Interfaz de carga, análisis e historial de audios



Figura 6. Interfaz de carga, análisis e historial de audios

La interfaz mostrada en la **Figura 6** concentra las funcionalidades relacionadas con la gestión de las muestras de audio. El usuario puede elegir el archivo que desea procesar, y antes de iniciar el análisis, el sistema verifica que el fichero cumpla las condiciones establecidas en cuanto a formato y duración; además, la vista incorpora el historial de análisis realizados previamente, permitiendo consultar los registros asociados a las muestras procesadas; una vez que el archivo ha sido validado, el usuario puede iniciar el proceso de clasificación a través de la opción correspondiente.

Elementos principales:

- **Área de carga:** Permite seleccionar el archivo de audio que se desea analizar.

- **Información de la muestra:** Presenta los detalles más relevantes del archivo seleccionado.
- **Validación del audio:** Comprueba el formato y la duración de la muestra, que debe estar entre 4 segundos y 5 minutos.
- **Botón de análisis:** Permite iniciar el proceso de análisis de la muestra validada.
- **Historial de análisis:** Permite ver los registros de los análisis que ya se han realizado.

2.6.3. Interfaz de visualización de resultados y generación de reportes

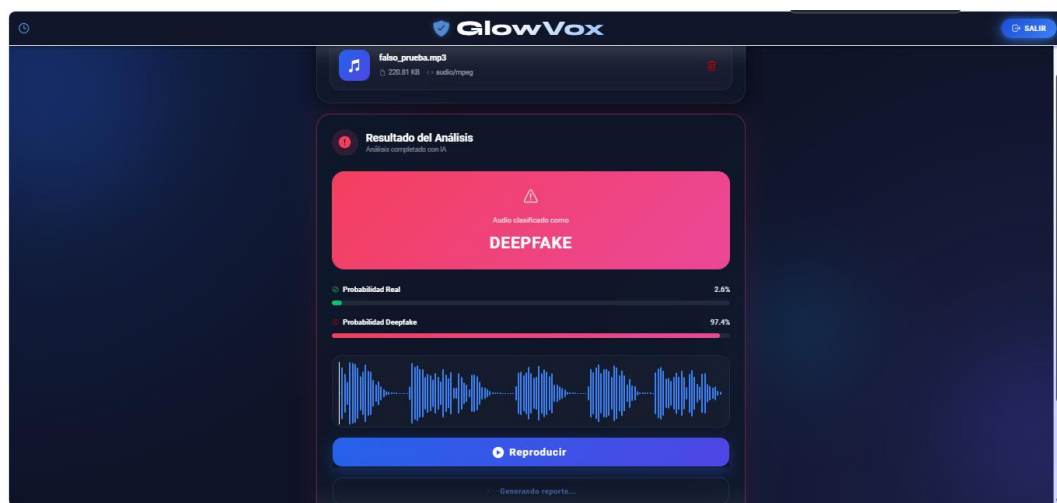


Figura 7. Interfaz de visualización de resultados y generación de reportes

Como se aprecia en la **Figura 7**, la plataforma presenta la información obtenida a partir de la clasificación de la muestra. En este apartado, el usuario puede consultar el veredicto asignado por el modelo, que puede corresponder a las categorías de Real, Deepfake o Sospechoso, junto con los datos relevantes del procesamiento y el nivel de confianza asociado; asimismo, la interfaz proporciona el acceso al reporte generado en formato PDF, permitiendo descargarlo para su posterior consulta.

Elementos principales:

- **Resultado de clasificación:** Presenta el veredicto obtenido de la muestra analizada.
- **Categoría del resultado:** Identifica el audio analizado como Real, Deepfake o Sospechoso.

- **Nivel de fiabilidad:** Muestra la confianza asociada a la predicción del modelo.
- **Acceso al reporte:** Permite acceder al documento que ha sido generado a partir de los resultados.
- **Botón de descarga:** Permite almacenar el reporte en formato PDF.

2.7. Pruebas

2.7.1. Caso N° 01 - Iniciar sesión y acceder a las funcionalidades protegidas

Información del caso de prueba		
Caso de prueba N°	01	
Caso de uso	Iniciar sesión y acceder a las funcionalidades protegidas	
Objetivos de la prueba	Comprobar que el sistema valide correctamente las credenciales del usuario y permita el acceso a las funcionalidades protegidas cuando la autenticación sea exitosa	
Rol	Usuario	
Pasos de la prueba		
1. El usuario accede a la interfaz de inicio de sesión. 2. Ingresa sus credenciales de acceso. 3. El sistema envía la información a la API REST. 4. La API REST valida las credenciales proporcionadas. 5. El sistema genera el token JWT y permite el acceso a las funcionalidades protegidas.		
Resultados de prueba		
Resultados esperados	Evaluación	
El sistema debe validar correctamente las credenciales del	Exitoso: ✓	Fallido:

usuario y permitir el acceso a las funcionalidades protegidas cuando la autenticación sea exitosa.		
--	--	--

Tabla 10. Caso de prueba N° 01 - Iniciar sesión y acceder a las funcionalidades protegidas

2.7.2. Caso N° 02 - Cargar y validar el formato del archivo del audio

Información del caso de prueba		
Caso de prueba N°	02	
Caso de uso	Cargar y validar el formato del archivo del audio	
Objetivos de la prueba	Comprobar que el sistema valide el formato del archivo de audio y permita continuar únicamente cuando corresponda a un formato compatible con el proceso de análisis.	
Rol	Usuario	
Pasos de la prueba		
1. El usuario accede al módulo de carga de archivos de audio. 2. Selecciona un archivo de audio. 3. El sistema verifica el formato del archivo. 4. Si el formato es válido, el sistema permite continuar con el proceso de análisis. 5. Si el formato no es compatible, el sistema rechaza el archivo y muestra un mensaje de error.		
Resultados de prueba		
Resultados esperados	Evaluación	
El sistema debe aceptar los	Exitoso: ✓	Fallido:

archivos de audio que cumplan con los formatos permitidos y rechazar aquellos que no sean compatibles.		
--	--	--

Tabla 11. Caso de prueba N° 02 - Cargar y validar el formato del archivo del audio

2.7.3. Caso N° 03 - Validar la duración del archivo de audio

Información del caso de prueba	
Caso de prueba N°	03
Caso de uso	Validar la duración del archivo de audio.
Objetivos de la prueba	Verificar que el sistema acepte únicamente archivos de audio con una duración comprendida entre 4 segundos y 5 minutos antes de iniciar el proceso de análisis.
Rol	Usuario
Pasos de la prueba	
1. El usuario selecciona un archivo de audio. 2. El sistema verifica la duración de la muestra. 3. El sistema compara la duración del archivo con los límites establecidos. 4. Si la duración se encuentra entre 4 segundos y 5 minutos, el archivo es aceptado. 5. Si la duración no cumple con los límites establecidos, el sistema rechaza el archivo y muestra un mensaje de error.	
Resultados de prueba	
Resultados esperados	Evaluación

El sistema debe aceptar archivos con una duración mínima de 4 segundos y máxima de 5 minutos, rechazando aquellos que se encuentren fuera de este rango.	Exitoso: ✓	Fallido:
--	-------------------	-----------------

Tabla 12. Caso de prueba N° 03 - Validar la duración del archivo de audio

2.7.4. Caso N° 04 - Analizar y clasificar una muestra de audio

Información del caso de prueba	
Caso de prueba N°	04
Caso de uso	Analizar y clasificar una muestra de audio
Objetivos de la prueba	Comprobar que el sistema procese correctamente el audio validado, lo envíe al modelo CNN-LSTM y muestre al usuario el resultado de la clasificación junto con la información obtenida durante el análisis.
Rol	Usuario
Pasos de la prueba	
1. El usuario selecciona un archivo de audio válido. 2. Inicia el proceso de análisis. 3. La API REST envía el audio al modelo CNN-LSTM. 4. El modelo procesa la muestra y determina la clasificación correspondiente. 5. La API REST recibe el resultado del modelo. 6. El sistema registra el resultado del análisis. 7. El sistema muestra al usuario la clasificación obtenida y la información asociada al análisis.	

Resultados de prueba		
Resultados esperados	Evaluación	
El sistema debe procesar correctamente el audio y mostrar el resultado de la clasificación en una de las categorías definidas: Real, Deepfake o Sospechoso, junto con la información correspondiente al análisis.	Exitoso: ✓	Fallido:

Tabla 13. Caso de prueba N° 04 - Analizar y clasificar una muestra de audio

2.7.5. Caso N° 05 - Generar y descargar el reporte del análisis

Información del caso de prueba	
Caso de prueba N°	05
Caso de uso	Generar y descargar el reporte del análisis
Objetivos de la prueba	Comprobar que el sistema genere correctamente el reporte en formato PDF, lo almacene y permita al usuario descargarlo posteriormente.
Rol	Usuario
Pasos de la prueba	
1. El usuario finaliza el análisis del audio. 2. El sistema genera el reporte con la información obtenida durante el análisis. 3. El reporte es almacenado en Cloudinary.	

4. El sistema registra la referencia del reporte en la base de datos. 5. El usuario selecciona la opción para descargar el reporte. 6. El sistema proporciona el archivo PDF generado. 7. El usuario abre el documento descargado para verificar su contenido.		
Resultados de prueba		
Resultados esperados	Evaluación	
El sistema debe generar correctamente el reporte en formato PDF, almacenarlo y permitir su descarga por parte del usuario. El documento descargado debe contener la información correspondiente al análisis realizado.	Exitoso: ✓	Fallido:

Tabla 14. Caso de prueba N° 05 - Generar y descargar el reporte del análisis

Se realizaron cinco pruebas funcionales para comprobar el buen funcionamiento de las principales funcionalidades del sistema en un entorno de uso concreto. Estas pruebas incluyeron la autenticación del usuario, la validación del formato y la duración de los archivos de audio, el análisis y la clasificación de las muestras en las categorías Real, Deepfake y Sospechoso, así como la generación y descarga de informes en formato PDF; cada prueba fue diseñada para comprobar el cumplimiento de las funcionalidades esperadas y también comprobar el comportamiento del sistema en diferentes condiciones de operación. Las evidencias de los casos de prueba se pueden encontrar en los anexos de este trabajo (**Véase Anexo 22 al 26**).

2.8. Resultados

2.8.1. Resultados del desarrollo y optimización del modelo

2.8.1.1. Resultados del modelo base

Como punto de partida para el desarrollo del sistema, se implementó un modelo base orientado a la clasificación binaria de audios en las categorías real y deepfake. Para esta fase, se contemplaron los conjuntos de datos seleccionados previamente,

los cuales pasaron a ser divididos en subconjuntos de entrenamiento, validación y prueba; de esta forma, se logró establecer una primera referencia acerca del comportamiento del modelo antes de realizar modificaciones posteriores en la arquitectura, hiperparámetros y la composición de los datos empleados.

La arquitectura implementada estuvo compuesta por una red convolucional con bloques residuales y convoluciones separables por profundidad, complementada con mecanismos de atención y una red recurrente bidireccional. En la etapa convolucional, se incorporaron técnicas como SEBlock, MixStyle y stochastic depth mediante drop path; posteriormente, las características extraídas fueron procesadas a través de un mecanismo de atención escalada basado en consultas, claves y valores, seguido de una capa LSTM bidireccional con una dimensión oculta de 64 y una única capa.

Para la fase de entrenamiento, se empleó una función de pérdida Focal Loss, $\gamma=2$, el optimizador AdamW, una tasa de aprendizaje de 2×10^{-4} y weight decay de 1×10^{-3} . El ajuste de la tasa de aprendizaje consistió de dos fases, la primera consistía en una fase de calentamiento lineal durante cinco epochs y la segunda consistió en un scheduling basado en una función coseno. El entrenamiento se realizó en batches de tamaño 8 y acumulación de gradientes en cuatro pasos, lo que equivale a un tamaño de batch efectivo de 32.

Las muestras fueron procesadas en segmentos fijos de cuatro segundos por archivo de audio. Las señales fueron recortadas o completadas mediante el padding hasta una longitud máxima de 250 frames. Para la representación de las muestras se utilizaron espectrogramas Mel convertidos a una escala logarítmica mediante Amplitude-to-dB, utilizando un rango de frecuencias de entre 20-7900 Hz. Además, se incorporaron las características derivadas mediante deltas y técnicas de SpecAugment, junto con una pequeña cantidad de ruido gaussiano para aumentar la variabilidad de las muestras durante el entrenamiento.

El modelo utilizó un muestreo equilibrado entre las clases real y deepfake. Asimismo, el umbral de decisión fue optimizado mediante la métrica F- $\beta(2)$, mientras que la calibración de la confianza se realizó mediante temperature scaling. Como criterio de monitoreo y detención temprana, se empleó la métrica F- $\beta(2)$, con

una paciencia de 12 épocas. Finalmente, se fijó una semilla de reproducibilidad con el valor 42 para mantener condiciones experimentales consistentes.

2.8.1.1.1. Curva de pérdida del entrenamiento del modelo base

La **Figura 8** presenta la evolución de la función de pérdida a lo largo del entrenamiento del modelo base, respecto a los datos contenidos en los conjuntos de entrenamiento y de validación. En las primeras épocas, se puede observar una pérdida de entrenamiento decreciente, lo que representa un aprendizaje sensible y rápido de los patrones que se pueden encontrar en los datos; posteriormente, la curva se va desarrollando de forma más moderada hasta alcanzar un comportamiento estable.

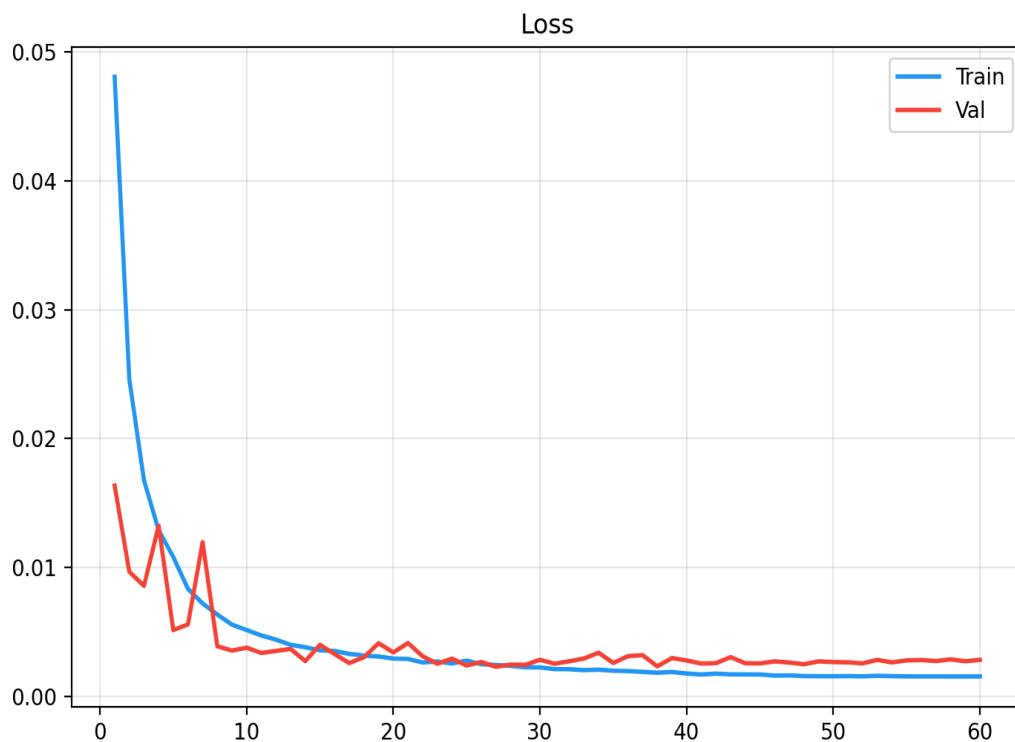


Figura 8. Curva de pérdida de entrenamiento y validación del modelo base

Por su parte, la curva correspondiente con la pérdida de validación aportó cierto tipo de variaciones en las primeras etapas del entrenamiento, estableciéndose posteriormente en un comportamiento relativamente estable alrededor de valores bajos; la proximidad entre ambas curvas durante gran parte del proceso indica que el modelo ha logrado aprender las características de las muestras sin observarse una separación excesiva respecto al comportamiento de la curva de pérdida de

entrenamiento. Esta evolución facilitó una forma de referencia sobre la estabilidad del proceso de aprendizaje del modelo base.

2.8.1.1.2. Métricas de evaluación del modelo base

Métrica	Resultado
Precisión	0.9982
Recall	0.9901
F1-Score	0.9941
F-Beta	0.9917
AUC	0.9989
EER	0.0064

Tabla 15. Métricas de evaluación del modelo base

La **Tabla 15** muestra los resultados que se obtuvieron del modelo base tras realizar la evaluación sobre los datos del conjunto de prueba. Se obtuvo para dicha evaluación una precisión de 0.9982 y un recall de 0.9901, lo que refleja un alto porcentaje de aciertos en la clasificación de las muestras evaluadas; además, el valor F1-Score alcanzó 0.9941, mientras que la métrica $F-\beta(2)$ obtuvo un valor de 0.9917. Por otra parte, el AUC-ROC fue de 0.9989 y el EER alcanzó un valor de 0.0064, lo que indica una baja tasa de error de clasificación en el punto de equilibrio de clasificación entre falsos positivos y falsos negativos.

2.8.1.1.3. Matriz de confusión del modelo base

La **Figura 9** muestra la matriz de confusión obtenida en la evaluación del modelo base, aplicado sobre el conjunto de prueba. Este modelo clasificó correctamente 16.792 muestras reales y 15.766 muestras deepfake; por otro lado, se registraron 29 falsos positivos, correspondientes a muestras reales clasificadas como deepfake, y 158 falsos negativos, correspondientes a muestras deepfake clasificadas como reales.

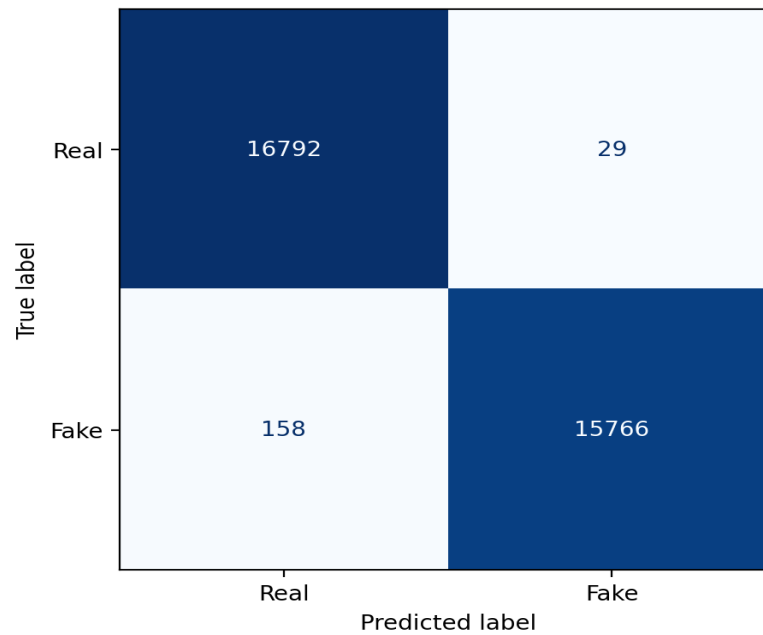


Figura 9. Matriz de confusión del modelo base

Estos resultados evidencian un rendimiento elevado del modelo base en la clasificación binaria de las muestras evaluadas. La cantidad de falsos negativos fue mayor que la de falsos positivos, lo que indica que el tipo de error predominante se relaciona más a muestras deepfake clasificados como muestras reales. En términos relativos, el FNR fue el siguiente 0.0099 y el FPR fue el 0.0017. Estos resultados permitieron identificar el comportamiento inicial del modelo y sirvieron como referencia para la siguiente etapa de optimización, en la que se modificaron elementos de la arquitectura, los hiperparámetros y la composición de los conjuntos de datos.

2.8.1.2. Resultados del modelo optimizado

Como resultado de la fase de optimización y ajuste del modelo, se obtuvo una nueva configuración orientada a mejorar la capacidad de generalización del sistema y a realizar una evaluación más representativa de los audios completos; para ello, se modificaron la arquitectura de la red, así como las estrategias de entrenamiento y procesamiento de las muestras. Entre los principales cambios destacan la elección de una arquitectura convolucional más sencilla, el uso de atención aditiva, y la aplicación de una red LSTM bidireccional de mayor capacidad, constituida por dos capas con 128 unidades ocultas.

Además, se incorporó un esquema de ventaneo deslizante para el tratamiento de los audios, utilizando ventanas de cinco segundos con un 50% de solapamiento. Las predicciones obtenidas a partir de las diferentes ventanas son posteriormente añadidas a nivel de archivo, permitiendo obtener una clasificación global para cada muestra de audio; por último, se aplicó normalización RMS sobre la señal original y se modificó la representación espectral utilizada durante el entrenamiento.

Dicha configuración también incluyó el uso de BCEWithLogitsLoss con suavizado de etiquetas durante el proceso de entrenamiento, el optimizador AdamW con una tasa de aprendizaje de 1×10^{-4} y el planificador ReduceLROnPlateau. El proceso de entrenamiento fue monitoreado mediante métricas de evaluación a nivel de archivo, aplicando un mecanismo de detención temprana que permitió finalizar el entrenamiento cuando no se evidenciaban mejoras relevantes en el rendimiento del modelo.

2.8.1.2.1. Curva de pérdida del entrenamiento del modelo optimizado

En la **Figura 10**, se describe la evolución de la pérdida de entrenamiento y validación durante el proceso de aprendizaje del modelo optimizado. Como se observa, la pérdida de entrenamiento decrece de manera progresiva desde las primeras épocas, pasando de un valor inicial cercano a 0.24, para finalizar estabilizándose en torno a un valor de 0.12; por su parte, la pérdida de validación disminuye de forma considerable durante las primeras épocas, y posteriormente, presenta ciertas fluctuaciones, manteniéndose en niveles inferiores a los registrados al inicio de la etapa de entrenamiento.

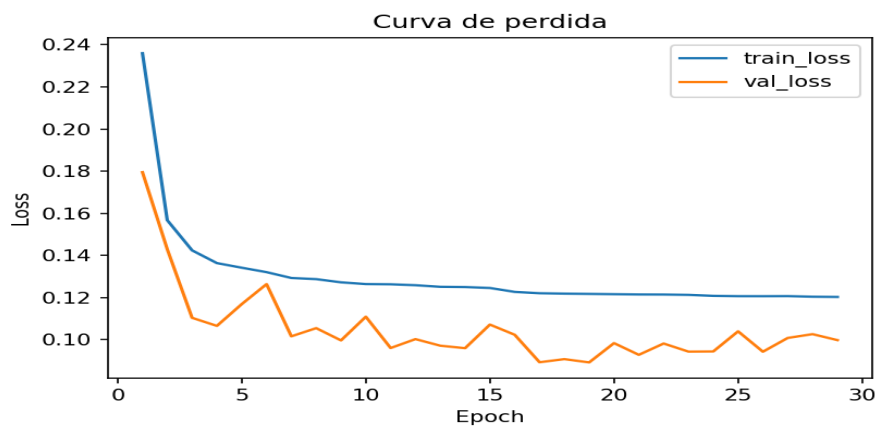


Figura 10. Curva de pérdida del entrenamiento y validación del modelo optimizado

La disminución de ambas curvas evidencia que el modelo ha aprendido patrones relevantes para diferenciar entre muestras reales y modificadas; adicionalmente, el comportamiento de la pérdida de validación no muestra un incremento progresivo, mientras que la pérdida de entrenamiento decrece de forma continua, lo que permite considerar que la fase de entrenamiento alcanzó una estabilización adecuada bajo la configuración empleada.

2.8.1.2.2. Curva ROC del modelo optimizado

La **Figura 11** muestra la curva característica operativa del receptor (ROC), la cual se construyó a partir de los resultados obtenidos durante la evaluación del modelo optimizado. Esta curva permite aproximar la relación que existe entre las tasas de verdaderos y falsos positivos para diferentes valores de los umbrales de decisión. El área bajo la curva obtuvo un valor de 0.9990, lo que evidencia una elevada capacidad de discriminación entre las clases real y deepfake en el conjunto de pruebas.

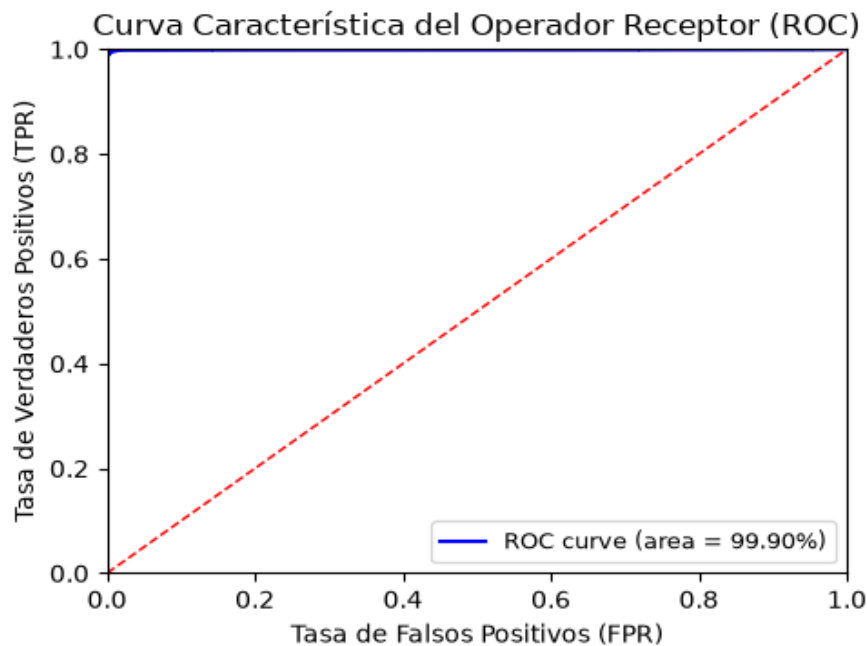


Figura 11. Curva ROC del modelo optimizado

La curva ROC queda muy próxima al extremo superior izquierdo de la gráfica, lo que indica que el modelo detecta de manera correcta la mayor parte de las muestras positivas manteniendo una baja tasa de falsos positivos; sin embargo, el valor obtenido debe interpretarse dentro del conjunto de datos y las condiciones de

evaluación empleadas, ya que no garantiza un rendimiento óptimo frente a todas las condiciones de audio que podrían presentarse.

2.8.1.2.3. Matriz de confusión del modelo optimizado

La **Figura 12** muestra la matriz de confusión obtenida durante la evaluación del modelo optimizado sobre el conjunto de prueba. La matriz permite observar la cantidad de muestras reales y deepfake correctamente clasificadas y los errores cometidos por el modelo en ambas clases. Los resultados muestran que de las 17.649 muestras reales, el modelo clasificó 17.615 como reales y 34 incorrectamente como deepfake; en el caso de las muestras deepfake, de 8.813 muestras, 8.753 fueron etiquetadas correctamente, mientras que 60 fueron identificadas como reales. En consecuencia, se observa una cantidad muy reducida de errores en ambas clases, aunque los falsos negativos son ligeramente superiores a los falsos positivos.

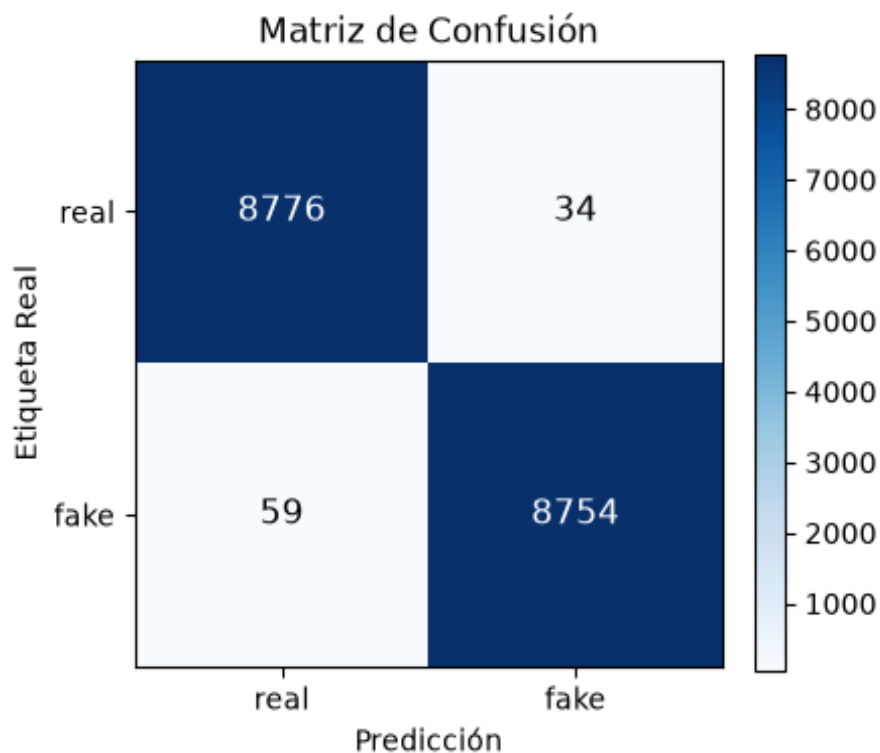


Figura 12. Matriz de confusión del modelo optimizado

Estos resultados muestran que el modelo optimizado mantiene un comportamiento consistente en la clasificación de las muestras evaluadas; no obstante, los errores

observados demuestran que el modelo no se comporta como un clasificador perfecto, aspecto que debe considerarse tanto en la interpretación de los resultados, así como en la posterior incorporación de mecanismos de gestión de confianza dentro del sistema en producción.

2.8.1.2.4. Métricas de evaluación del modelo optimizado

Una vez seleccionada la configuración final del modelo, se evaluó su desempeño a partir de un conjunto de métricas que permiten medir la clasificación entre audios reales y deepfake; los resultados obtenidos permiten valorar aspectos diferentes del comportamiento del modelo como, por ejemplo, la precisión de sus predicciones, la capacidad para identificar correctamente muestras de cada una de las respectivas clases, el equilibrio entre precisión y recall y la tasa de error equivalente. Los valores obtenidos se pueden consultar en la **Tabla 16**.

Métrica	Resultado
Precisión	0.9961
Recall	0.9933
F1-Score	0.9947
AUC	0.9990
EER	0.0055

Tabla 16. Métricas de evaluación del modelo optimizado

2.8.1.3. Comparación de resultados

Métrica	Modelo base	Modelo optimizado
Precisión	0.9982	0.9961
Recall	0.9901	0.9933
F1-Score	0.9941	0.9947

Métrica	Modelo base	Modelo optimizado
AUC	0.9989	0.9990
EER	0.0064	0.0055

Tabla 17. Comparación del desempeño de los modelos de clasificación

La **Tabla 17** muestra los resultados en una comparación entre el modelo base y el modelo optimizado; de forma general, el modelo optimizado presenta un comportamiento similar. En este caso, las métricas de evaluación como el recall, el F1-Score, el AUC y el EER han mejorado, alcanzando un AUC de 0.9990, mientras que la tasa de error equivalente reduce de 0.0064 a 0.0055; a pesar de que la precisión ha disminuido para este modelo de 0.9982 a 0.9961, el comportamiento conjunto de las métricas evidencia un mejor equilibrio en la capacidad del modelo para reconocer correctamente ambas clases.

2.8.1.4. Resultados de la implementación del sistema

Durante la implementación del sistema de detección de deepfake de audio, se integró el modelo de clasificación optimizado en la plataforma web desarrollada. La incorporación del modelo permitió procesar las muestras de audio cargadas por el usuario y obtener una predicción relacionada con las categorías real y deepfake, junto con las probabilidades correspondientes a cada resultado; por lo tanto, la aplicación permite presentar al usuario no solo la clasificación obtenida, sino también información adicional relacionada con el grado de confianza asociado a la predicción y la clasificación realizada.

Como parte del procesamiento que realiza el sistema, las muestras de audio son convertidas en representaciones espectrales antes de ser analizadas por el modelo de clasificación. La **Figura 13** muestra ejemplos de dichas representaciones antes y después del procesamiento aplicado a una muestra real y a una muestra deepfake, permitiendo visualizar los cambios obtenidos a partir de las transformaciones realizadas sobre las señales de audio. Este procedimiento permite mantener una estructura de entrada uniforme para las diferentes muestras analizadas por el sistema.

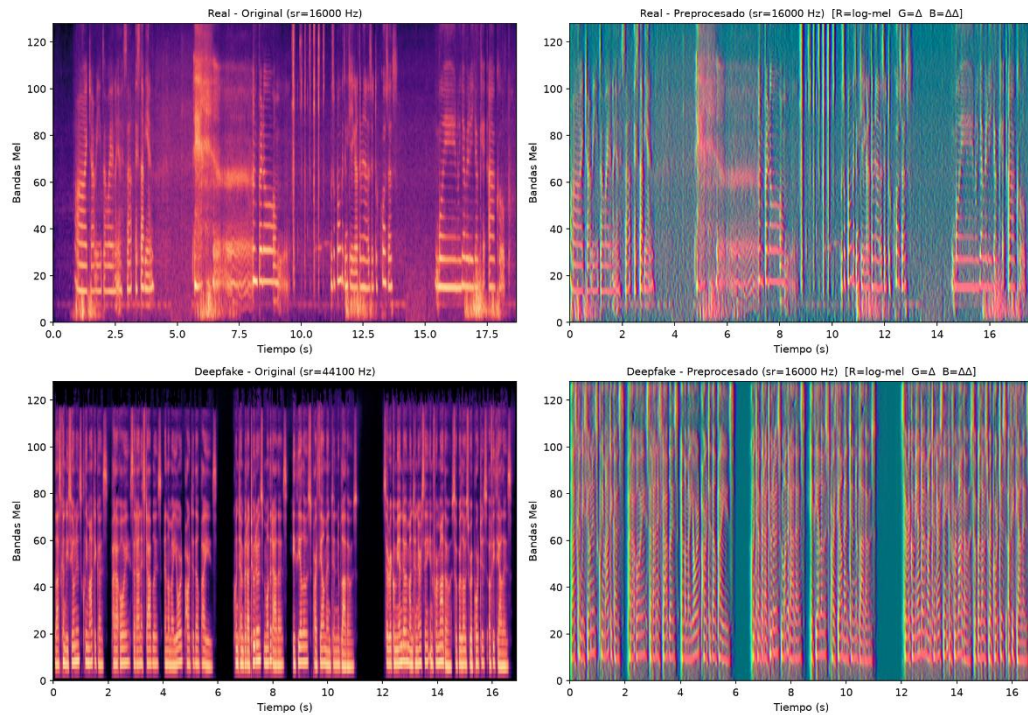


Figura 13. Representaciones espectrales antes y después del preprocesamiento

Como se observa en la **Figura 13**, las muestras originales presentan variaciones en su frecuencia de muestreo y en su duración; al ser procesadas, estas características se ajustan mediante el pipeline que está implementado en el sistema, a partir del cual se generan las representaciones basadas en coeficientes Mel y en sus componentes delta y delta-delta, que constituyen la información que posteriormente utiliza el modelo para realizar la clasificación de las muestras.

Por otro lado, debido a que el rendimiento del modelo fue evaluado a partir de un conjunto de datos específico y bajo condiciones controladas, se incorporó un mecanismo adicional para gestionar aquellos casos donde la predicción no resultaba suficientemente concluyente. Este mecanismo permite que el sistema no emita una clasificación cuando las probabilidades asociadas a la clase real y deepfake presentan una diferencia reducida.

Para este tipo de casos, se introdujo la categoría "Sospechoso" como un mecanismo de gestión de confianza de la aplicación; cabe señalar que la categoría "Sospechoso" no representa una tercera clase que el modelo haya aprendido, ya que el modelo de

clasificación fue desarrollado para distinguir únicamente entre audios reales y deepfakes. En consecuencia, el resultado "Sospechoso" se determina posteriormente, a nivel de la lógica de la aplicación y a partir de las probabilidades generadas por el modelo.

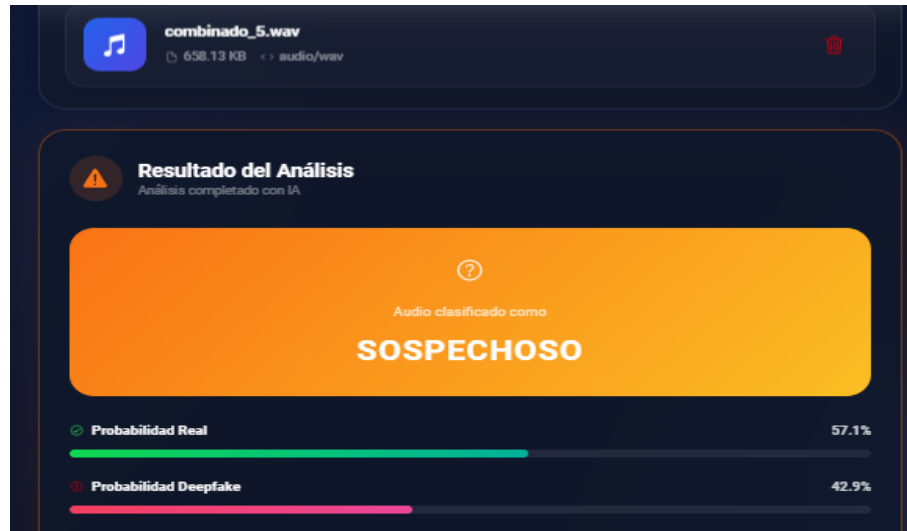


Figura 14. Resultado de una clasificación con nivel de confianza insuficiente

Como se observa en la **Figura 14**, el sistema asigna un resultado "Sospechoso" cuando la predicción obtenida no tiene una distancia suficientemente amplia entre las clases. En este caso, el audio tiene una probabilidad de 57,1 % de corresponder a un audio real y un 42,9 % de ser un deepfake; debido a que los valores están tan cerca, el sistema no se compromete a reportar el resultado como tal y lo hace utilizando la categoría "Sospechoso".

Esta implementación permite que el nivel de incertidumbre de la predicción se traslade a la interfaz del sistema, proporcionando una respuesta más prudente ante muestras con características ambiguas. De esta manera, la aplicación no considera automáticamente cada predicción como concluyente, sino que emplea un mecanismo alternativo para gestionar aquellos momentos en los que el modelo no proporciona un nivel de certeza suficiente para dictaminar con claridad la autenticidad del audio.

Por dicho motivo, la incorporación de este mecanismo complementa a la predicción binaria generada por el modelo y permite que el sistema gestione de una forma más adecuada las predicciones potencialmente ambiguas durante su funcionamiento en producción. Esto resulta especialmente relevante, ya que las métricas obtenidas

durante la fase de evaluación reflejan el comportamiento del modelo sobre los conjuntos de datos empleados en dicha etapa, mientras que las muestras procesadas posteriormente pueden presentar características distintas a las observadas durante el entrenamiento y la evaluación del modelo.

CONCLUSIONES

- El conjunto de datos recopilado presentó la diversidad necesaria para abordar el problema relacionado con la detección de deepfakes de audio, al incluir muestras reales y sintéticas generadas a partir de diferentes fuentes y métodos. La diversidad de las muestras permite incluir diversas características acústicas relacionadas con las diferentes clases de audios y proporcionar una amplia base para el desarrollo del modelo de clasificación.
- La depuración de las muestras permitió mejorar la consistencia del conjunto de datos que se utiliza en la realización de este trabajo; la validación de los ficheros, la normalización de las señales de audio y la eliminación de las muestras repetidas o inconsistentes permitieron reducir la cantidad de información que podría afectar de forma negativa, tanto al aprendizaje del modelo, como a la fiabilidad de los resultados obtenidos.
- Las representaciones basadas en coeficientes Mel, junto con sus componentes delta y delta-delta, proporcionaron una representación adecuada de las características espectrales y temporales de las señales de audio. Estas representaciones permitieron obtener la información necesaria para que la arquitectura analizara los patrones acústicos que permiten diferenciar entre audios reales y sintéticos.
- La arquitectura CNN-LSTM desarrollada alcanzó un desempeño adecuado para la clasificación de audios reales y deepfake. La comparación entre el modelo base y su configuración optimizada mostró un comportamiento global más favorable de esta última en métricas como recall, F1-Score, AUC-ROC y EER, razón por la cual fue seleccionada como la configuración final empleada en el sistema.
- La evaluación del modelo optimizado reflejó un alto desempeño, alcanzando una precisión de 0.9961, recall de 0.9933, F1-Score de 0.9947, AUC-ROC de

0.9990 y un EER de 0.0055. Sin embargo, el análisis de su funcionamiento en producción evidenció la necesidad de contemplar situaciones en las que el modelo no presenta suficiente certeza para establecer una clasificación definitiva; por ello, se incorporó la clase "sospechoso" como un mecanismo de gestión de la incertidumbre. Esta categoría no representa una tercera clase del modelo, sino un resultado definido por la lógica del sistema cuando las probabilidades de las clases real y fake son demasiado cercanas.

- La integración del modelo en la plataforma web permitió consolidar un sistema funcional para el análisis de audios, articulando de manera conjunta los procesos de autenticación, carga y procesamiento de audios, análisis y clasificación, visualización de resultados y generación de reportes en un solo sistema; de este modo, se obtuvo una plataforma que centraliza el proceso de detección de deepfakes de audio y presenta al usuario los resultados del análisis de forma organizada.

RECOMENDACIONES

- Se aconseja ampliar progresivamente el conjunto de datos, introduciendo nuevas muestras reales y sintéticas procedentes de distintas fuentes, lenguas, factores de grabación y métodos de creación; de esta forma, se podría aumentar la diversidad de los audios analizados y evaluar de forma más exhaustiva la capacidad del modelo para generalizar sobre características diferentes a las presentes en los datos utilizados durante su desarrollo.
- También se recomienda mantener un proceso de revisión y control de calidad de las muestras, especialmente a medida que se expanda el conjunto de datos. La integración de nuevas validaciones y métodos automatizados de detección de registros duplicados, incompletos o inconsistentes permitiría conservar la calidad de la información utilizada por el sistema.
- Se sugiere investigar otras técnicas de representación de las señales de audio, abarcando métodos de extracción alternativos, así como la combinación de características espectrales y temporales. Esto permitiría comprobar si otras representaciones de las señales de audio añaden nueva información que ayude a distinguir entre audios reales y generados de forma artificial.

- Se recomienda continuar haciendo pruebas con las diversas configuraciones de la arquitectura CNN-LSTM y sus hiperparámetros, teniendo en cuenta conjuntos de datos de documentos más amplios y diversos. Estas pruebas podrían determinar si otras configuraciones mejoran el comportamiento del modelo ante muestras con características distintas a las empleadas durante su desarrollo.
- Se considera conveniente profundizar en el mecanismo utilizado para identificar predicciones con baja confianza, considerando nuevos conjuntos de audios y diferentes condiciones de análisis. Esto permitirá ajustar con más precisión los criterios empleados para determinar cuándo una muestra debe considerarse "Sospechosa" y mejorar la gestión de aquellos casos en los que el modelo no presenta suficiente certeza.
- Se aconseja realizar actualizaciones regulares del modelo y de la plataforma web, teniendo en cuenta el surgimiento de nuevas técnicas de generación y manipulación de voz. De este modo, el propio podría adaptarse a nuevos métodos de síntesis de voz y mantener un desempeño adecuado en la detección de deepfakes.

REFERENCIAS

- [1] Vishal Manishbhai Patel and Dr. Sheshang Degadwala, “Deepfake Detection Using Convolutional Neural Networks and LSTM Modelling,” *Int. J. Sci. Res. Sci. Technol.*, vol. 12, no. 3, pp. 448–456, May 2025, doi: 10.32628/ijrst2512361.
- [2] S. M. Qureshi, A. Saeed, S. H. Almotiri, F. Ahmad, and M. A. A. Ghamdi, “Deepfake forensics: a survey of digital forensic methods for multimodal deepfake identification on social media,” *PeerJ Comput. Sci.*, vol. 10, p. e2037, 2024, doi: 10.7717/PEERJ-CS.2037.
- [3] “Análisis forense de deepfakes: un estudio de métodos forenses digitales para la identificación multimodal de deepfakes en redes sociales - PMC.” Accessed: Aug. 19, 2025. [Online]. Available: <https://pmc.ncbi.nlm.nih.gov/articles/PMC11157519/>
- [4] O. A. Shaaban and R. Yildirim, “Audio Deepfake Detection Using Deep Learning,” *Engineering Reports*, vol. 7, no. 3, Mar. 2025, doi: 10.1002/eng2.70087.
- [5] Z. Khanjani, G. Watson, and V. P. Janeja, “Audio deepfakes: A survey,” *Front. Big Data*, vol. 5, p. 1001063, Jan. 2023, doi: 10.3389/FDATA.2022.1001063/FULL.
- [6] “Deepfakes de audio: una encuesta - PMC.” Accessed: Aug. 19, 2025. [Online]. Available: <https://pmc.ncbi.nlm.nih.gov/articles/PMC9869423/>
- [7] L. H. Singh, P. Charanarur, and N. K. Chaudhary, “Advancements in detecting Deepfakes: AI algorithms and future prospects – a review,” *Discover Internet of Things*, vol. 5, no. 1, pp. 1–30, Dec. 2025, doi: 10.1007/S43926-025-00154-0/TABLES/2.
- [8] “Avances en la detección de deepfakes: algoritmos de IA y perspectivas futuras: un análisis | Descubra el Internet de las Cosas.” Accessed: Aug. 19, 2025. [Online]. Available: <https://link.springer.com/article/10.1007/s43926-025-00154-0>

- [9] “Ciberseguridad: los datos tienen la respuesta - Editorial Universidad Icesi.” Accessed: Apr. 14, 2026. [Online]. Available: <https://www.icesi.edu.co/editorial/ciberseguridad-datos/>
- [10] D. Cristina *et al.*, “‘Deepfakes’ un nuevo desafío para los derechos de las mujeres,” 2004, doi: 10.14679/3206.
- [11] I. Jiménez-Castellanos Ballesteros, “CAPÍTULO 5 LA DIFÍCIL CONVIVENCIA ENTRE LA INTELIGENCIA ARTIFICIAL Y LOS DERECHOS FUNDAMENTALES: LA TECNOLOGÍA DEEPFAKE”, Accessed: Sep. 08, 2025. [Online]. Available: <https://www.gov.uk/government/case-studies/innovating-to-detect-deep->
- [12] A. Olivós-Farfán, “Challenges for communication managers: artificial intelligence and reputation management”, Accessed: Sep. 08, 2025. [Online]. Available: <http://hdl.handle.net/10757/676096>
- [13] C. DE Derecho, F. Geovanny Sinaluisa Sagñay, and M. Wendy Pilar Romero Noboa, “UNIVERSIDAD NACIONAL DE CHIMBORAZO FACULTAD DE CIENCIAS POLÍTICAS Y ADMINISTRATIVAS,” 2024, Accessed: Aug. 26, 2025. [Online]. Available: <http://dspace.unach.edu.ec/bitstream/51000/13841/1/Sinaluisa%20Sag%20c3%20b1ay%20F%282024%29%20Deepfakes%20Dificultades%20probatorias%20%20y%20su%20incidencia%20en%20la%20vulneraci%20b3n%20al%20derecho%20de%20intimidad.%28Tesis%20de%20Pregrado%29%20Universidad%20Nacional%20de%20Chimborazo%20Riobamba%20Ecuador.pdf>
- [14] “TellezRodríguez_Alvaro_TFM_ANTONIO ROBLES GOMEZ”, Accessed: Sep. 08, 2025. [Online]. Available: <https://oai.e-spacio.uned.es/server/api/core/bitstreams/31384430-3dba-41bb-b74f-0be01aaf8a01/content>
- [15] “Vista de Repercusiones éticas sobre el uso indebido del deepfake en el ámbito de las TIC mediante un análisis cualitativo documental | Reto.” Accessed: Aug. 26, 2025. [Online]. Available:

<https://revistas.sena.edu.co/index.php/RETO/article/view/3040/5520>

- [16] P. L. Kharvi, “Understanding the Impact of AI-Generated Deepfakes on Public Opinion, Political Discourse, and Personal Security in Social Media,” *IEEE Secur. Priv.*, vol. 22, no. 4, pp. 115–122, 2024, doi: 10.1109/MSEC.2024.3405963.
- [17] “Tesis Liz_Kely_Rodriguez_Saenz_compressed”, Accessed: Aug. 26, 2025. [Online]. Available: https://repositorio.upn.edu.pe/bitstream/handle/11537/40538/Tesis%20Liz_Kely_Rodriguez_Saenz_compressed.pdf?sequence=1&isAllowed=y
- [18] “Modelo de detección de audios en español y videos deepfake contra ataques de ingeniería social.” Accessed: Apr. 19, 2026. [Online]. Available: <https://repositorio.uniandes.edu.co/entities/publication/a0c39515-488b-4c10-9aa0-eb253919dc59>
- [19] O. Acosta, S. Yulexy, M. De Titulación, and S. Elena, “UNIVERSIDAD ESTATAL PENÍNSULA DE SANTA ELENA FACULTAD DE SISTEMAS Y TELECOMUNICACIONES TITULO DEL TRABAJO DE TITULACIÓN APLICACIÓN DE TÉCNICAS DE IA Y ACÚSTICA FORENSE PARA IDENTIFICAR AUDIOS SINTÉTICOS AUTOR.”
- [20] F. J. García-Ull, “Deepfakes: el próximo reto en la detección de noticias falsas*”, doi: 10.5565/rev/analisi.3378.
- [21] “Deepfake: Análisis de sus implicancias tecnológicas y jurídicas en la era de la Inteligencia Artificial.” Accessed: Apr. 09, 2026. [Online]. Available: https://www.scielo.org.mx/scielo.php?pid=S2448-51362024000200359&script=sci_arttext
- [22] F. Ramos-Zaga and F. Ramos-Zaga, “Deepfake: Análisis de sus implicancias tecnológicas y jurídicas en la era de la Inteligencia Artificial,” *Derecho global. Estudios sobre derecho y justicia*, vol. 9, no. 27, pp. 359–387, Jul. 2024, doi: 10.32870/DGEDJ.V9I27.754.
- [23] “Deepfakes: Revisión sistemática de tecnologías, impacto y estrategias de

- detección | Revista de Investigación en Tecnologías de la Información.” Accessed: Apr. 09, 2026. [Online]. Available: <https://riti.es/index.php/riti/article/view/335>
- [24] J. Y. M. Serrano, C. E. D. la T. Noriega, S. C. Zatarain, L. Y. A. Contreras, and G. Á. D. Lucas, “Deepfakes: Revisión sistemática de tecnologías, impacto y estrategias de detección,” *Revista de Investigación en Tecnologías de la Información*, vol. 13, no. 29, pp. 19–37, Feb. 2025, doi: 10.36825/RITI.13.29.003.
- [25] “Líneas.” Accessed: Oct. 14, 2025. [Online]. Available: <https://incyt.upse.edu.ec/index.php/investigacion/lineas>
- [26] A. Olivos-Farfán, “Challenges for communication managers: artificial intelligence and reputation management”, Accessed: Sep. 09, 2025. [Online]. Available: <http://hdl.handle.net/10757/676096>
- [27] F. Ramos-Zaga and F. Ramos-Zaga, “Deepfake: Análisis de sus implicancias tecnológicas y jurídicas en la era de la Inteligencia Artificial,” *Derecho global. Estudios sobre derecho y justicia*, vol. 9, no. 27, pp. 359–387, Jul. 2024, doi: 10.32870/DGEDJ.V9I27.754.
- [28] “Deepfake: Análisis de sus implicancias tecnológicas y jurídicas en la era de la Inteligencia Artificial.” Accessed: Sep. 08, 2025. [Online]. Available: https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S2448-51362024000200359
- [29] S. Elena, “UNIVERSIDAD ESTATAL PENÍNSULA DE SANTA ELENA FACULTAD DE SISTEMAS Y TELECOMUNICACIONES,” 2024, Accessed: Sep. 09, 2025. [Online]. Available: <https://repositorio.upse.edu.ec/server/api/core/bitstreams/1fdda2a6-79b3-44a1-b844-fa36e24e5e08/content>
- [30] “PLAN NACIONAL DE DESARROLLO,” 2025. Accessed: Sep. 10, 2025. [Online]. Available: https://www.planificacion.gob.ec/wp-content/uploads/2025/08/PlanNacionalDeDesarrollo25-29_EcuadorNoSeDetiene.pdf

- [31] A. Hamza *et al.*, “Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000. Deepfake Audio Detection via MFCC features using Machine Learning”, doi: 10.1109/ACCESS.2017.DOI.
- [32] J. Arias-Gonzales, “DISEÑO Y METODOLOGÍA DE LA INVESTIGACIÓN,” 2021, Accessed: Oct. 07, 2025. [Online]. Available: <https://www.researchgate.net/publication/352157132>
- [33] C. A. Ramos-Galarza, “Alcances de una investigación,” *CienciAmérica*, vol. 9, no. 3, pp. 1–6, Oct. 2020, doi: 10.33210/ca.v9i3.336.
- [34] A. F. Haro Sarango, E. R. Chisag Pallmay, J. P. Ruiz Sarzosa, and J. E. Caicedo Pozo, “Tipos y clasificación de las investigaciones,” *LATAM Revista Latinoamericana de Ciencias Sociales y Humanidades*, vol. 5, no. 2, Apr. 2024, doi: 10.56712/latam.v5i2.1927.
- [35] I. Reyes Blácido, E. Damián Guerra, N. Ciriaco Reyes, B. Oscar Corimayhua Luque, and L. Marcelino Urbina Olortegui, “TITLE: Scientific methods and their application in pedagogical research.”
- [36] E. Maya, “Métodos y técnicas de investigación Una propuesta ágil para la presentación de trabajos científicos en las áreas de arquitectura, urbanismo y disciplinas afines,” 2014.
- [37] “openslr.org.” Accessed: Apr. 08, 2026. [Online]. Available: <https://www.openslr.org/>
- [38] “Kaggle: Your Home for Data Science.” Accessed: Apr. 08, 2026. [Online]. Available: <https://www.kaggle.com/>
- [39] M. Medina Romero Rómulo Rojas León Wilder Bustamante Hoces Raquel Loaiza Carrasco Christian Martel Carranza Roxana Castillo Acobo, “METODOLOGÍA DE LA INVESTIGACIÓN TÉCNICAS E INSTRUMENTOS DE INVESTIGACIÓN”, doi: 10.35622/inudi.b.080.
- [40] X. Li, P.-Y. Chen, and W. Wei, “Measuring the Robustness of Audio Deepfake Detectors,” 2025, Accessed: Oct. 14, 2025. [Online]. Available: <https://www-ft-com.ezphost.dur.ac.uk/content/bd1bc5b4-f540-48f8-9cda->

- [41] G. Thawani, S. Shukla, R. Wable, U. Chugh, A. Gondal, and F. Y. Student, "A NOVEL APPROACH FOR DEEPFAKE AUDIO DETECTION," *www.irjmets.com @International Research Journal of Modernization in Engineering*, vol. 9447, doi: 10.56726/IRJMETs74467.
- [42] "(13) Deepfakes: el próximo reto en la detección de noticias falsas." Accessed: Apr. 13, 2026. [Online]. Available: https://www.academia.edu/49488127/Deepfakes_el_pr%C3%B3ximo_reto_en_la_detecci%C3%B3n_de_noticias_falsas
- [43] F. José and G. U. García-Ull, "DeepFakes: The Next Challenge in Fake News Detection," *Analisi*, no. 64, pp. 103–120, Jun. 2021, doi: 10.5565/REV/ANALISI.3378.
- [44] "La metamorfosis de la verdad: Deepfakes y el desafío de la autenticidad en la sociedad digital - Dialnet." Accessed: Apr. 14, 2026. [Online]. Available: <https://dialnet.unirioja.es/servlet/articulo?codigo=10345849>
- [45] A. A. Gallo, U. Rey, and J. Carlos, "La metamorfosis de la verdad: Deepfakes y el desafío de la autenticidad en la sociedad digital," *Derecom: Revista Internacional de Derecho de la Comunicación y las Nuevas Tecnologías*, ISSN-e 1988-2629, Vol. 38, Nº. 1, 2025, págs. 45-57, vol. 38, no. 1, pp. 45–57, 2025, doi: 10.5209/dere.102344.
- [46] "Deepfake como estrategia para desinformar en las redes sociales durante las campañas electorales en Ecuador - Dialnet." Accessed: Apr. 14, 2026. [Online]. Available: <https://dialnet.unirioja.es/servlet/articulo?codigo=10563038>
- [47] A. E. M. Romero *et al.*, "Deepfake como estrategia para desinformar en las redes sociales durante las campañas electorales en Ecuador," *Revista de ciencias sociales*, ISSN-e 1315-9518, Vol. 31, Nº. 4, 2025, págs. 223-238, vol. 31, no. 4, pp. 223–238, 2025, Accessed: Apr. 14, 2026. [Online]. Available: <https://dialnet.unirioja.es/servlet/articulo?codigo=10563038&info=resume>

n&idioma=SPA

- [48] C. DE Derecho, F. Geovanny Sinaluisa Sagñay, and M. Wendy Pilar Romero Noboa, “Deepfakes, dificultades probatorias y su incidencia en la vulneración al derecho de intimidad,” Oct. 2024, Accessed: Apr. 14, 2026. [Online]. Available: <http://dspace.unach.edu.ec/handle/51000/13841>
- [49] “Repositorio Digital UNACH: Deepfakes, dificultades probatorias y su incidencia en la vulneración al derecho de intimidad.” Accessed: Apr. 14, 2026. [Online]. Available: <http://dspace.unach.edu.ec/handle/51000/13841>
- [50] “Deepfakes Pornográficos: Impacto jurídico-probatorio y social en el Ecuador - Dialnet.” Accessed: Apr. 14, 2026. [Online]. Available: <https://dialnet.unirioja.es/servlet/articulo?codigo=9959552>
- [51] F. G. S. Sagñay, W. P. R. Noboa, N. F. Freire, <https://dialnet.unirioja.es/servlet/autor?codigo=6724809>, <https://dialnet.unirioja.es/servlet/autor?codigo=6724810>, and <https://dialnet.unirioja.es/servlet/autor?codigo=6724811>, “Deepfakes Pornográficos: Impacto jurídico-probatorio y social en el Ecuador,” *REINCISOL: Revista de Investigación Científica y Social, ISSN-e 2953-6421, Vol. 3, Nº. 6, 2024 (Ejemplar dedicado a: Science and Social Development)*, págs. 2912-2934, vol. 3, no. 6, pp. 2912–2934, 2024, doi: 10.59282/reincisol.V3(6)2912-2934.
- [52] “TellezRodríguez_Alvaro_TFM_ANTONIO ROBLES GOMEZ (1)”.
- [53] R. L. de Mántaras, P. B. Crosa, <https://dialnet.unirioja.es/servlet/autor?codigo=40920>, and <https://dialnet.unirioja.es/servlet/autor?codigo=287165>, “¿Qué es la inteligencia artificial?,” *Papeles de relaciones ecosociales y cambio global, ISSN 1888-0576, Nº. 164 (Riesgos, ventajas y repercusiones. de la Inteligencia Artificial), 2023 (Ejemplar dedicado a: Inteligencia Artificial)*, págs. 13-21, no. 164, pp. 13–21, 2023, Accessed: Apr. 14, 2026. [Online]. Available: <https://dialnet.unirioja.es/servlet/articulo?codigo=9287111&info=resumen>

&idioma=ENG

- [54] “¿Qué es la inteligencia artificial? - Dialnet.” Accessed: Apr. 14, 2026. [Online]. Available: <https://dialnet.unirioja.es/servlet/articulo?codigo=9287111>
- [55] “INTELIGENCIA ARTIFICIAL 101 COSAS QUE DEBES SABER HOY SOBRE NUESTRO FUTURO INTELIGENCIA ARTIFICIAL,” 2018, Accessed: Apr. 14, 2026. [Online]. Available: www.planetadelibros.com
- [56] “El uso de la inteligencia artificial generativa en la investigación de la ciberdelincuencia de género: ante el auge de los deepfakes | IUS ET SCIENTIA.” Accessed: Apr. 14, 2026. [Online]. Available: <https://revistascientificas.us.es/index.php/ies/article/view/24691>
- [57] I. G. Pulido, “El uso de la inteligencia artificial generativa en la investigación de la ciberdelincuencia de género: ante el auge de los deepfakes,” *IUS ET SCIENTIA*, vol. 9, no. 2, pp. 157–180, Dec. 2023, doi: 10.12795/IESTSCIENTIA.2023.I02.08.
- [58] B. Zhang, H. Cui, V. Nguyen, and M. Whitty, “Audio Deepfake Detection: What Has Been Achieved and What Lies Ahead,” *Sensors* 2025, Vol. 25, Page 1989, vol. 25, no. 7, p. 1989, Mar. 2025, doi: 10.3390/S25071989.
- [59] “Detección de audio deepfake: Lo que se ha logrado y lo que nos espera.” Accessed: Aug. 19, 2025. [Online]. Available: <https://www.mdpi.com/1424-8220/25/7/1989>
- [60] “Aprendizaje Automático y Aprendizaje Profundo.” Accessed: Apr. 14, 2026. [Online]. Available: https://www.scielo.cl/scielo.php?pid=S0718-33052021000200180&script=sci_arttext
- [61] J. Díaz-Ramírez and J. Díaz-Ramírez, “Aprendizaje Automático y Aprendizaje Profundo,” *Ingeniare. Revista chilena de ingeniería*, vol. 29, no. 2, pp. 180–181, Jun. 2021, doi: 10.4067/S0718-33052021000200180.
- [62] “Técnicas de aprendizaje automático aplicadas en los sistemas de predicción | Tecnología Investigación y Academia.” Accessed: Apr. 14,

2026. [Online]. Available:
<https://revistas.udistrital.edu.co/index.php/tia/article/view/17325>
- [63] I. N. S. Rativa, “Técnicas de aprendizaje automático aplicadas en los sistemas de predicción,” *Tecnología Investigación y Academia*, vol. 8, no. 1, pp. 37–53, Jul. 2021, Accessed: Apr. 14, 2026. [Online]. Available: <https://revistas.udistrital.edu.co/index.php/tia/article/view/17325>
- [64] O. A. Shaaban and R. Yildirim, “Audio Deepfake Detection Using Deep Learning,” *Engineering Reports*, vol. 7, no. 3, p. e70087, Mar. 2025, doi: 10.1002/ENG2.70087;ISSUE:ISSUE:DOI.
- [65] “Detección de deepfakes de audio mediante aprendizaje profundo - Shaaban - 2025 - Informes de ingeniería - Biblioteca en línea de Wiley.” Accessed: Apr. 14, 2026. [Online]. Available: <https://onlinelibrary.wiley.com/doi/10.1002/eng2.70087>
- [66] K. G. Kim, “Book Review: Deep Learning,” *Healthc. Inform. Res.*, vol. 22, no. 4, p. 351, 2016, doi: 10.4258/hir.2016.22.4.351.
- [67] C. J. Tablada, G. A. Torres, <https://dialnet.unirioja.es/servlet/autor?codigo=5938189>, and <https://dialnet.unirioja.es/servlet/autor?codigo=5938190>, “Redes Neuronales Artificiales,” *Revista de Educación Matemática (RevEM)*, ISSN-e 1852-2890, ISSN 0326-8780, Vol. 24, N°. 3, 2009, vol. 24, no. 3, p. 2, 2009, Accessed: Apr. 14, 2026. [Online]. Available: <https://dialnet.unirioja.es/servlet/articulo?codigo=8832469&info=resumen&idioma=SPA>
- [68] “Redes Neuronales Artificiales - Dialnet.” Accessed: Apr. 14, 2026. [Online]. Available: <https://dialnet.unirioja.es/servlet/articulo?codigo=8832469>
- [69] C. Alberto Ruiz Marta Susana Basualdo Autor and D. Jorge Matich, “Cátedra: Informática Aplicada a la Ingeniería de Procesos-Orientación I Redes Neuronales: Conceptos Básicos y Aplicaciones.”
- [70] “Análisis de la influencia de características vocales en modelos de

- detección de voces clonadas mediante IA.” Accessed: Apr. 14, 2026. [Online]. Available: <https://uvadoc.uva.es/handle/10324/79502>
- [71] D. Fernando, M. Martínez, D. Blas, and T. Gracia, “Análisis de la influencia de características vocales en modelos de detección de voces clonadas mediante IA,” 2025, Accessed: Apr. 14, 2026. [Online]. Available: <https://uvadoc.uva.es/handle/10324/79502>
- [72] S. Sen and A. Raghunathan, “Approximate computing for Long Short Term Memory (LSTM) neural networks,” *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 37, no. 11, pp. 2266–2276, Nov. 2018, doi: 10.1109/TCAD.2018.2858362.
- [73] “Cálculo aproximado para redes neuronales de memoria a largo plazo (LSTM) | Revistas y publicaciones del IEEE | IEEE Xplore.” Accessed: Oct. 14, 2025. [Online]. Available: <https://ieeexplore.ieee.org/document/8417946>
- [74] “Detección de deepfakes en imágenes, audio y vídeos manipulados: un marco de aprendizaje profundo multimodal de tres etapas | Inteligencia Artificial.” Accessed: Apr. 14, 2026. [Online]. Available: <https://journal.iberamia.org/index.php/intartif/article/view/1674>
- [75] L. Nelson, H. Batra, and P. Radha, “Deepfake Detection in Manipulated Images/ Audio/ Videos: A Three-Stage Multi-Modal Deep Learning Framework,” *Inteligencia Artificial*, vol. 28, no. 76, pp. 20–39, Jun. 2025, doi: 10.4114/INTARTIF.VOL28ISS76PP20-39.
- [76] J. Iriz González, “Procesamiento de Audio con Técnicas de Inteligencia Artificial,” 2018.
- [77] D. Stranneby and W. Walker, “Spectral analysis and modulation,” *Digital Signal Processing and Applications*, pp. 131–158, 2004, doi: 10.1016/B978-075066344-1/50005-7.
- [78] “Spectral analysis and modulation - ScienceDirect.” Accessed: Oct. 24, 2025. [Online]. Available: <https://www.sciencedirect.com/science/article/abs/pii/B978075066344150>

0057?via%3Dihub

- [79] C. Asuai, A. P. Arinomor, C. T. Atumah, I. F. Kowhoro, and D. E. Ogheneochuko, “Hybrid CNN-LSTM Architectures for Deepfake Audio Detection Using Mel Frequency Cepstral Coefficients and Spectrogram Analysis,” *American Journal of Mathematical and Computer Modelling* 2025, Volume 10, Page 98, vol. 10, no. 3, pp. 98–109, Sep. 2025, doi: 10.11648/J.AJMCM.20251003.12.
- [80] “Hybrid CNN-LSTM Architectures for Deepfake Audio Detection Using Mel Frequency Cepstral Coefficients and Spectrogram Analysis , American Journal of Mathematical and Computer Modelling, Science Publishing Group.” Accessed: Oct. 24, 2025. [Online]. Available: <https://www.sciencepublishinggroup.com/article/10.11648/j.ajmcm.20251003.12>
- [81] M. Melad and A. Ashames, “Mel-Spectrograms and Data Augmentation for Spoken Digit Classification.”
- [82] D. Povedanó Alvarez, A. Lucila, S. Orozco, L. Javier, and G. Villalba, *mediante Espectrogramas y Aprendizaje por Transferencia*.
- [83] J. Picone and S. Balakrishnama, “Speech Recognition using Mel cepstrum, delta cepstrum and delta-delta features,” 1998.
- [84] A. Garetà Manzano, “Extracción de características y clasificación automática de señales de audio,” Sep. 08, 2015, *Universitat Politècnica de Catalunya*. Accessed: Apr. 15, 2026. [Online]. Available: <https://hdl.handle.net/2117/77451>
- [85] “Extracción de características y clasificación automática de señales de audio.” Accessed: Apr. 15, 2026. [Online]. Available: <https://upcommons.upc.edu/entities/publication/78d722ba-c0f4-4707-9972-ad3e616cd4a1>
- [86] “Detección de audio deepfake con características espectrales y arquitectura basada en ResNeXt - ScienceDirect.” Accessed: Apr. 15, 2026. [Online]. Available:

<https://www.sciencedirect.com/science/article/pii/S0950705125007725>

- [87] G. Tahaoglu, D. Baracchi, D. Shullani, M. Iuliani, and A. Piva, “Deepfake audio detection with spectral features and ResNeXt-based architecture,” *Knowl. Based. Syst.*, vol. 323, p. 113726, Jul. 2025, doi: 10.1016/J.KNOSYS.2025.113726.
- [88] “Un enfoque para identificar deepfakes mediante aprendizaje profundo | SpringerLink.” Accessed: Oct. 14, 2025. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-031-35501-1_57
- [89] S. S. Gedela, N. Yanda, H. Kusumanchi, S. Daki, K. Challa, and P. Gurralla, “An Approach to Identify DeepFakes Using Deep Learning,” *Lecture Notes in Networks and Systems*, vol. 716 LNNS, pp. 574–583, 2023, doi: 10.1007/978-3-031-35501-1_57.
- [90] K. Verma *et al.*, “Deepfake Audio Detection: A Comparative Study of Advanced Deep Learning Models,” *IEEE Access*, vol. 13, pp. 168447–168462, 2025, doi: 10.1109/ACCESS.2025.3611839.
- [91] “Deepfake Audio Detection: A Comparative Study of Advanced Deep Learning Models | IEEE Journals & Magazine | IEEE Xplore.” Accessed: Oct. 14, 2025. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/11172285>
- [92] R. K. Bhukya, A. Raj, and D. N. Raja, “Audio Deepfakes: Feature Extraction and Model Evaluation for Detection,” *2024 5th International Conference for Emerging Technology, INCET 2024*, 2024, doi: 10.1109/INCET61516.2024.10593405.
- [93] “Deepfakes de audio: Extracción de características y evaluación de modelos para su detección | Publicación de la conferencia IEEE | IEEE Xplore.” Accessed: Oct. 14, 2025. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/10593405>
- [94] V. Sundaram, S. Babitha, and S. Vekkot, “Leveraging Acoustic Features and Deep Neural Architectures for Audio Deepfake Detection,” *ICACC - International Conference on Advances in Computing and*

Communications, no. 2024, 2024, doi:
10.1109/ICACC63692.2024.10845392.

- [95] “Aprovechamiento de las características acústicas y las arquitecturas neuronales profundas para la detección de deepfakes de audio | Publicación de la Conferencia IEEE | IEEE Xplore.” Accessed: Oct. 14, 2025. [Online]. Available: <https://ieeexplore.ieee.org/document/10845392>
- [96] “Análisis de características acústicas para la detección de suplantación de audio basada en aprendizaje automático explicable - ScienceDirect.” Accessed: Apr. 20, 2026. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1077314224002261>
- [97] C. Bisogni, V. Loia, M. Nappi, and C. Pero, “Acoustic features analysis for explainable machine learning-based audio spoofing detection,” *Computer Vision and Image Understanding*, vol. 249, p. 104145, Dec. 2024, doi: 10.1016/J.CVIU.2024.104145.
- [98] N. Kaur, A. Dixit, and S. Kingra, “A Deep Learning Fusion Model Leveraging Spectral Features for Audio Deepfake Detection,” *Int. J. Advanced Networking and Applications*, pp. 6551–6560, 2025.
- [99] M. Chitale, A. Dhawale, M. Dubey, and S. Ghane, “A Hybrid CNN-LSTM Approach for Deepfake Audio Detection,” *2024 3rd International Conference on Artificial Intelligence for Internet of Things, AIIoT 2024*, 2024, doi: 10.1109/AIIOT58432.2024.10574576.
- [100] “A Hybrid CNN-LSTM Approach for Deepfake Audio Detection | IEEE Conference Publication | IEEE Xplore.” Accessed: Oct. 14, 2025. [Online]. Available: <https://ieeexplore.ieee.org/document/10574576>

ANEXOS

Anexo 1. Árbol de Problemas

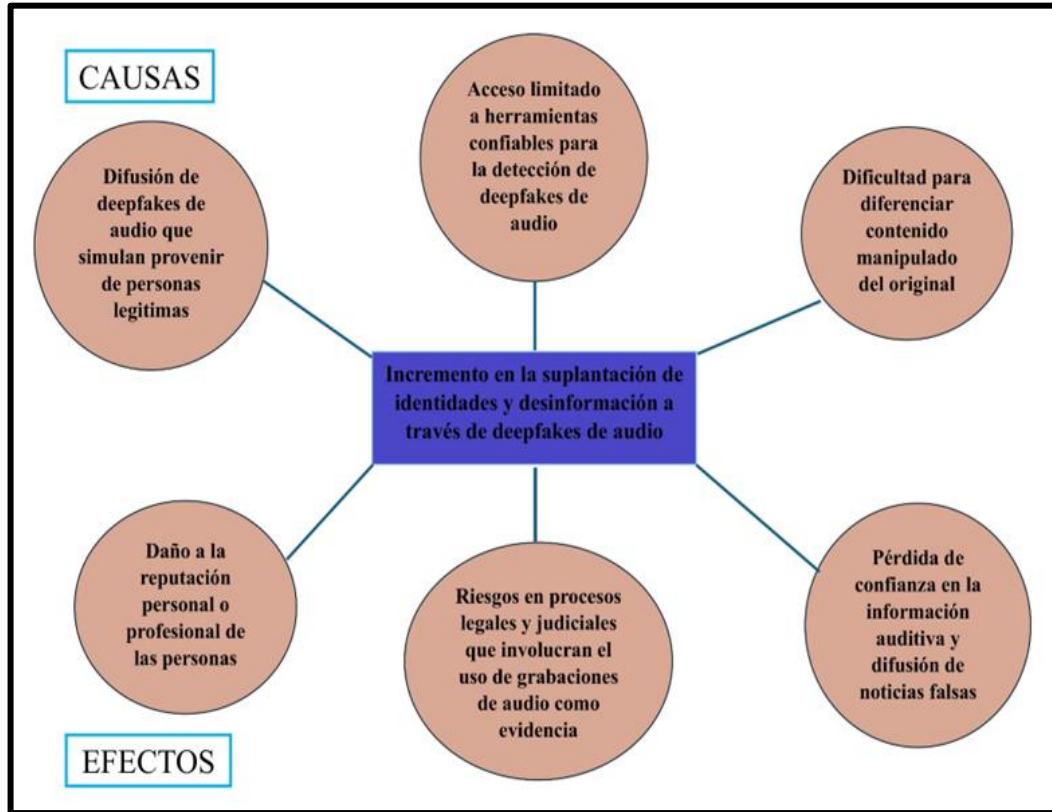


Figura 15. Árbol de Problemas

Anexo 2. Formato de ficha de análisis

Campo	Descripción
Nombre del dataset	Nombre del conjunto de datos al que pertenece la muestra
Origen	Fuente o procedencia de la muestra de audio
Formato	Formato del archivo de audio
Duración (s)	Duración de la muestra expresada en segundos
Frecuencia (Hz)	Frecuencia de muestreo de la señal de audio
Tipo (real / fake)	Clasificación de la muestra como real o deepfake
Subconjunto	Subconjunto al que pertenece la muestra dentro del dataset
Ruta	Ubicación del archivo dentro del conjunto de datos

Tabla 18. Formato de Ficha de Análisis

Anexo 3. Fragmento de Ficha de Análisis – Aishell

Nombre del dataset: AISHELL							
Origen	Nombre del Archivo	Formato	Duración (s)	Frecuencia (Hz)	Tipo	Subconjunto	Ruta
OpenSLR	BAC009S0002W0122.wav	wav	6	16000	real	training	Aishell\wav\train\S0002\BAC009S0002W0122.wav
OpenSLR	BAC009S0002W0123.wav	wav	3,87	16000	real	training	Aishell\wav\train\S0002\BAC009S0002W0123.wav
OpenSLR	BAC009S0002W0124.wav	wav	5,41	16000	real	training	Aishell\wav\train\S0002\BAC009S0002W0124.wav
OpenSLR	BAC009S0724W0121.wav	wav	4,28	16000	real	validation	Aishell\wav\dev\S0724\BAC009S0724W0121.wav
OpenSLR	BAC009S0724W0122.wav	wav	4,34	16000	real	validation	Aishell\wav\dev\S0724\BAC009S0724W0122.wav
OpenSLR	BAC009S0724W0123.wav	wav	3,24	16000	real	validation	Aishell\wav\dev\S0724\BAC009S0724W0123.wav
OpenSLR	BAC009S0764W0121.wav	wav	4,2	16000	real	testing	Aishell\wav\test\S0764\BAC009S0764W0121.wav
OpenSLR	BAC009S0764W0122.wav	wav	4,12	16000	real	testing	Aishell\wav\test\S0764\BAC009S0764W0122.wav
OpenSLR	BAC009S0764W0123.wav	wav	4	16000	real	testing	Aishell\wav\test\S0764\BAC009S0764W0123.wav

Tabla 19. Fragmento de Ficha de Análisis - Aishell

Anexo 4. Fragmento de Ficha de Análisis – Common Voice

Nombre del dataset: Common Voice							
Origen	Nombre del Archivo	Formato	Duración (s)	Frecuencia (Hz)	Tipo	Subconjunto	Ruta
Kaggle	sample-000000.mp3	mp3	4,1	48000	real	training	CommonVoice\cv-valid-train\cv-valid-train\sample-000000.mp3
Kaggle	sample-000001.mp3	mp3	7,56	48000	real	training	CommonVoice\cv-valid-train\cv-valid-train\sample-000001.mp3
Kaggle	sample-000000.mp3	mp3	5,06	48000	real	validation	CommonVoice\cv-valid-dev\cv-valid-dev\sample-000000.mp3
Kaggle	sample-000001.mp3	mp3	3,58	48000	real	validation	CommonVoice\cv-valid-dev\cv-valid-dev\sample-000001.mp3
Kaggle	sample-000001.mp3	mp3	1,94	48000	real	testing	CommonVoice\cv-valid-test\cv-valid-test\sample-000001.mp3
Kaggle	sample-000002.mp3	mp3	1,94	48000	real	testing	CommonVoice\cv-valid-test\cv-valid-test\sample-000002.mp3

Tabla 20. Fragmento de Ficha de Análisis - Common Voice

Anexo 5. Fragmento de Ficha de Análisis – LibriSpeech

Nombre del dataset: LibriSpeech							
Origen	Nombre del Archivo	Formato	Duración (s)	Frecuencia (Hz)	Tipo	Subconjunto	Ruta
OpenSLR	103-1240-0000.flac	flac	14,09	16000	real	training	LibriSpeech\train-clean-100\103\1240\103-1240-0000.flac
OpenSLR	103-1240-0001.flac	flac	15,95	16000	real	training	LibriSpeech\train-clean-100\103\1240\103-1240-0001.flac
OpenSLR	1272-128104-0000.flac	flac	5,86	16000	real	validation	LibriSpeech\dev-clean\1272\128104\1272-128104-0000.flac
OpenSLR	1272-128104-0001.flac	flac	4,82	16000	real	validation	LibriSpeech\dev-clean\1272\128104\1272-128104-0001.flac
OpenSLR	1089-134686-0001.flac	flac	3,27	16000	real	testing	LibriSpeech\test-clean\1089\134686\1089-134686-0001.flac
OpenSLR	1089-134686-0002.flac	flac	6,62	16000	real	testing	LibriSpeech\test-clean\1089\134686\1089-134686-0002.flac

Tabla 21. Fragmento de Ficha de Análisis – LibriSpeech

Anexo 6. Fragmento de Ficha de Análisis – VCTK Corpus

Nombre del dataset: VCTK Corpus							
Origen	Nombre del Archivo	Formato	Duración (s)	Frecuencia (Hz)	Tipo	Subconjunto	Ruta
Kaggle	p225_001.wav	wav	2,05	48000	real	sin_division	VCTK-Corpus\VCTK-Corpus\wav48\p225\p225_001.wav
Kaggle	p225_002.wav	wav	4,06	48000	real	sin_division	VCTK-Corpus\VCTK-Corpus\wav48\p225\p225_002.wav
Kaggle	p225_003.wav	wav	7,81	48000	real	sin_division	VCTK-Corpus\VCTK-Corpus\wav48\p225\p225_003.wav
Kaggle	p225_006.wav	wav	7,77	48000	real	sin_division	VCTK-Corpus\VCTK-Corpus\wav48\p225\p225_006.wav
Kaggle	p225_007.wav	wav	6,19	48000	real	sin_division	VCTK-Corpus\VCTK-Corpus\wav48\p225\p225_007.wav
Kaggle	p225_008.wav	wav	8,07	48000	real	sin_division	VCTK-Corpus\VCTK-Corpus\wav48\p225\p225_008.wav

Tabla 22. Fragmento de Ficha de Análisis – VCTK Corpus

Anexo 7. Fragmento de Ficha de Análisis – ASVSpooof 2019

Nombre del dataset: ASVSpooof 2019							
Origen	Nombre del Archivo	Formato	Duración (s)	Frecuencia (Hz)	Tipo	Subconjunto	Ruta
Kaggle	LA_T_1138215.flac	Flac	3,46	16000	real	Train	Downloads \LA\training\flac\LA_T_1138215.flac
Kaggle	LA_T_1271820.flac	Flac	4,4	16000	real	Train	Downloads \LA\training\flac\LA_T_1271820.flac
Kaggle	LA_D_1047731.flac	Flac	5,41	16000	real	Val	Downloads \LA\validation\flac\LA_D_1047731.flac
Kaggle	LA_E_2834763.flac	Flac	1,43	16000	fake	Test	Downloads \LA\testing\flac\LA_E_2834763.flac
Kaggle	LA_E_8877452.flac	Flac	3,98	16000	fake	Test	Downloads \LA\testing\flac\LA_E_8877452.flac
Kaggle	LA_E_6828287.flac	Flac	1,56	16000	fake	Test	Downloads \LA\testing\flac\LA_E_6828287.flac

Tabla 23. Fragmento de Ficha de Análisis – ASVSpooof 2019

Anexo 8. Fragmento de Ficha de Análisis – CVoiceFake

Nombre del dataset: CVoiceFake							
Origen	Nombre del Archivo	Formato	Duración (s)	Frecuencia (Hz)	Tipo	Subconjunto	Ruta
Kaggle	common_voice_de_38024994.mp3	mp3	4,14	16000	real	sin_division	CVoiceFake\de\Bonafide\common_voice_de_38024994.mp3
Kaggle	common_voice_de_38024995.mp3	mp3	6,3	16000	real	sin_division	CVoiceFake\de\Bonafide\common_voice_de_38024995.mp3
Kaggle	common_voice_de_38024996.mp3	mp3	3,96	16000	real	sin_division	CVoiceFake\de\Bonafide\common_voice_de_38024996.mp3
Kaggle	common_voice_de_38024998.mp3	mp3	8,39	16000	real	sin_division	CVoiceFake\de\Bonafide\common_voice_de_38024998.mp3
Kaggle	common_voice_de_38024999.mp3	mp3	6,37	16000	real	sin_division	CVoiceFake\de\Bonafide\common_voice_de_38024999.mp3
Kaggle	common_voice_de_38025000.mp3	mp3	7,45	16000	real	sin_division	CVoiceFake\de\Bonafide\common_voice_de_38025000.mp3

Tabla 24. Fragmento de Ficha de Análisis – CVoiceFake

Anexo 9. Fragmento de Ficha de Análisis – DFADD

Nombre del dataset: DFADD							
Origen	Nombre del Archivo	Formato	Duración (s)	Frecuencia (Hz)	Tipo	Subconjunto	Ruta
Hugging Face	p225_001_GradTTS.flac	flac	2,37	16000	fake	training	DFADD\DATASET_GradTTS\train\p225_001_GradTTS.flac
Hugging Face	p225_002_GradTTS.flac	flac	3,29	16000	fake	training	DFADD\DATASET_GradTTS\train\p225_002_GradTTS.flac
Hugging Face	p226_001_GradTTS.flac	flac	2,37	16000	fake	validation	DFADD\DATASET_GradTTS\valid\p226_001_GradTTS.flac
Hugging Face	p226_002_GradTTS.flac	flac	3,29	16000	fake	validation	DFADD\DATASET_GradTTS\valid\p226_002_GradTTS.flac
Hugging Face	p227_001_GradTTS.flac	flac	2,37	16000	fake	testing	DFADD\DATASET_GradTTS\test\p227_001_GradTTS.flac
Hugging Face	p227_002_GradTTS.flac	flac	3,29	16000	fake	testing	DFADD\DATASET_GradTTS\test\p227_002_GradTTS.flac

Tabla 25. Fragmento de Ficha de Análisis – DFADD

Anexo 10. Fragmento de Ficha de Análisis – Fake-or-Real

Nombre del dataset: Fake-or-Real							
Origen	Nombre del Archivo	Formato	Duración (s)	Frecuencia (Hz)	Tipo	Subconjunto	Ruta
Kaggle	file1000.mp3	mp3	1,8	16000	fake	training	Fake-or-Real\for-original\training\file1000.mp3
Kaggle	file10000.mp3	mp3	2,23	16000	fake	training	Fake-or-Real\for-original\training\file10000.mp3
Kaggle	file1.mp3	mp3	1,72	22050	fake	validation	Fake-or-Real\for-original\validation\file1.mp3
Kaggle	file100.mp3	mp3	1,12	16000	fake	validation	Fake-or-Real\for-original\validation\file100.mp3
Kaggle	file1.wav	wav	1,97	24000	fake	testing	Fake-or-Real\for-original\testing\file1.wav
Kaggle	file10.wav	wav	1,63	24000	fake	testing	Fake-or-Real\for-original\testing\file10.wav

Tabla 26. Fragmento de Ficha de Análisis – Fake-or-Real

Anexo 11. Fragmento de Ficha de Análisis – In-The-Wild

Nombre del dataset: In-The-Wild							
Origen	Nombre del Archivo	Formato	Duración (s)	Frecuencia (Hz)	Tipo	Subconjunto	Ruta
Kaggle	100.wav	wav	2,88	16000	real	training	In-The-Wild\training\real\100.wav
Kaggle	1000.wav	wav	2,34	16000	real	training	In-The-Wild\training\real\1000.wav
Kaggle	10001.wav	wav	6,33	16000	real	training	In-The-Wild\training\real\10001.wav
Kaggle	10013.wav	wav	2,79	16000	real	validation	In-The-Wild\validation\real\10013.wav
Kaggle	10024.wav	wav	2,84	16000	real	validation	In-The-Wild\validation\real\10024.wav
Kaggle	10030.wav	wav	4,33	16000	real	validation	In-The-Wild\validation\real\10030.wav
Kaggle	10004.wav	wav	8,43	16000	real	testing	In-The-Wild\testing\real\10004.wav
Kaggle	10018.wav	wav	2,08	16000	real	testing	In-The-Wild\testing\real\10018.wav
Kaggle	10038.wav	wav	3,67	16000	real	testing	In-The-Wild\testing\real\10038.wav

Tabla 27. Fragmento de Ficha de Análisis – In-The-Wild

Anexo 12. Fragmento de Ficha de Análisis – LibriSeVoc

Nombre del dataset: LibriSeVoc							
Origen	Nombre del Archivo	Formato	Duración (s)	Frecuencia (Hz)	Tipo	Subconjunto	Ruta
Kaggle	1034_121119_000010_000004_gen.wav	wav	6,15	24000	fake	sin_division	LibriSeVoc\diffwave\1034_121119_000010_000004_gen.wav
Kaggle	1034_121119_000010_000006_gen.wav	wav	7,5	24000	fake	sin_division	LibriSeVoc\diffwave\1034_121119_000010_000006_gen.wav
Kaggle	103_1241_000004_000002_gen.wav	wav	5,5	24000	fake	sin_division	LibriSeVoc\diffwave\103_1241_000004_000002_gen.wav
Kaggle	1040_133433_000071_000002_gen.wav	wav	10,73	24000	fake	sin_division	LibriSeVoc\diffwave\1040_133433_000071_000002_gen.wav
Kaggle	1040_133433_000138_000001_gen.wav	wav	6,44	24000	fake	sin_division	LibriSeVoc\diffwave\1040_133433_000138_000001_gen.wav
Kaggle	1040_133433_000164_000000_gen.wav	wav	5,47	24000	fake	sin_division	LibriSeVoc\diffwave\1040_133433_000164_000000_gen.wav

Tabla 28. Fragmento de Ficha de Análisis – LibriSeVoc

Anexo 13. Fragmento de Ficha de Análisis – MLAAD

Nombre del dataset: MLAAD							
Origen	Nombre del Archivo	Formato	Duración (s)	Frecuencia (Hz)	Tipo	Subconjunto	Ruta
Hugging Face	mlaad_00000.wav	wav	2,53	22050	fake	sin_division	MLAAD\mlaad_00000.wav
Hugging Face	mlaad_00001.wav	wav	6,35	22050	fake	sin_division	MLAAD\mlaad_00001.wav
Hugging Face	mlaad_00002.wav	wav	3,81	22050	fake	sin_division	MLAAD\mlaad_00002.wav
Hugging Face	mlaad_00004.wav	wav	9,4	22050	fake	sin_division	MLAAD\mlaad_00004.wav
Hugging Face	mlaad_00005.wav	wav	13,76	22050	fake	sin_division	MLAAD\mlaad_00005.wav
Hugging Face	mlaad_00006.wav	wav	7,4	22050	fake	sin_division	MLAAD\mlaad_00006.wav
Hugging Face	mlaad_00007.wav	wav	6,92	22050	fake	sin_division	MLAAD\mlaad_00007.wav
Hugging Face	mlaad_00008.wav	wav	7,91	22050	fake	sin_division	MLAAD\mlaad_00008.wav
Hugging Face	mlaad_00009.wav	wav	7,68	22050	fake	sin_division	MLAAD\mlaad_00009.wav

Tabla 29. Fragmento de Ficha de Análisis – MLAAD

Anexo 14. Fragmento de Ficha de Análisis – PartialSpoof

Nombre del dataset: PartialSpoof							
Origen	Nombre del Archivo	Formato	Duración (s)	Frecuencia (Hz)	Tipo	Subconjunto	Ruta
Zenodo	CON_T_0000029.wav	wav	2,93	16000	fake	training	PartialSpoof\training\con_wav\CON_T_0000029.wav
Zenodo	CON_T_0000069.wav	wav	3,61	16000	fake	training	PartialSpoof\training\con_wav\CON_T_0000069.wav
Zenodo	CON_D_0000001.wav	wav	5	16000	fake	validation	PartialSpoof\validation\con_wav\CON_D_0000001.wav
Zenodo	CON_D_0000009.wav	wav	2,35	16000	fake	validation	PartialSpoof\validation\con_wav\CON_D_0000009.wav
Zenodo	CON_E_0000368.wav	wav	3,09	16000	fake	testing	PartialSpoof\testing\con_wav\CON_E_0000368.wav
Zenodo	CON_E_0000371.wav	wav	1,73	16000	fake	testing	PartialSpoof\testing\con_wav\CON_E_0000371.wav

Tabla 30. Fragmento de Ficha de Análisis – PartialSpoof

Anexo 15. Fragmento de Ficha de Análisis – WaveFake

Nombre del dataset: WaveFake							
Origen	Nombre del Archivo	Formato	Duración (s)	Frecuencia (Hz)	Tipo	Subconjunto	Ruta
Kaggle	gen_0.wav	Wav	3,65	22050	fake	sin_division	WaveFake\common_voices_prompts_from_conformer_fastspeech2_pwg_ljspeech\gen_0.wav
Kaggle	gen_1.wav	Wav	1,23	22050	fake	sin_division	WaveFake\common_voices_prompts_from_conformer_fastspeech2_pwg_ljspeech\gen_1.wav
Kaggle	gen_10.wav	Wav	7,12	22050	fake	sin_division	WaveFake\common_voices_prompts_from_conformer_fastspeech2_pwg_ljspeech\gen_10.wav
Kaggle	gen_10000.wav	Wav	2,07	22050	fake	sin_division	WaveFake\common_voices_prompts_from_conformer_fastspeech2_pwg_ljspeech\gen_10000.wav
Kaggle	gen_10002.wav	Wav	4,05	22050	fake	sin_division	WaveFake\common_voices_prompts_from_conformer_fastspeech2_pwg_ljspeech\gen_10002.wav
Kaggle	gen_10003.wav	Wav	5,41	22050	fake	sin_division	WaveFake\common_voices_prompts_from_conformer_fastspeech2_pwg_ljspeech\gen_10003.wav

Tabla 31. Fragmento de Ficha de Análisis – WaveFake

Anexo 16. Fragmento de Ficha de Análisis – Coqui TTS

Nombre del dataset: Coqui TTS							
Origen	Nombre del Archivo	Formato	Duración (s)	Frecuencia (Hz)	Tipo	Subconjunto	Ruta
CoquiTTS	coqui_en_train_00001.wav	wav	3,62	24000	fake	train	coqui_en_train_00001.wav
CoquiTTS	coqui_en_train_00002.wav	wav	4,68	24000	fake	train	coqui_en_train_00002.wav
CoquiTTS	coqui_en_train_00003.wav	wav	5,47	24000	fake	train	coqui_en_train_00003.wav
CoquiTTS	coqui_en_val_00001.wav	wav	4,64	24000	fake	val	coqui_en_val_00001.wav
CoquiTTS	coqui_en_val_00002.wav	wav	3,76	24000	fake	val	coqui_en_val_00002.wav
CoquiTTS	coqui_en_val_00003.wav	wav	4,59	24000	fake	val	coqui_en_val_00003.wav
CoquiTTS	coqui_en_test_00001.wav	wav	3,57	24000	fake	test	coqui_en_test_00001.wav
CoquiTTS	coqui_en_test_00002.wav	wav	4,96	24000	fake	test	coqui_en_test_00002.wav
CoquiTTS	coqui_en_test_00003.wav	wav	5,98	24000	fake	test	coqui_en_test_00003.wav

Tabla 32. Fragmento de Ficha de Análisis – Coqui TTS

Anexo 17. Fragmento de Ficha de Análisis – Edge TTS

Nombre del dataset: Edge TTS							
Origen	Nombre del Archivo	Formato	Duración (s)	Frecuencia (Hz)	Tipo	Subconjunto	Ruta
edgeTTS	es_00001.mp3	mp3	5,76	24000	fake	training	training\es\es_00001.mp3
edgeTTS	es_00002.mp3	mp3	5,4	24000	fake	training	training\es\es_00002.mp3
edgeTTS	es_00003.mp3	mp3	5,52	24000	fake	training	training\es\es_00003.mp3
edgeTTS	es_00421.mp3	mp3	5,04	24000	fake	validation	validation\es\es_00421.mp3
edgeTTS	es_00422.mp3	mp3	5,54	24000	fake	validation	validation\es\es_00422.mp3
edgeTTS	es_00423.mp3	mp3	6,38	24000	fake	validation	validation\es\es_00423.mp3
edgeTTS	es_00511.mp3	mp3	6,02	24000	fake	testing	testing\es\es_00511.mp3
edgeTTS	es_00512.mp3	mp3	6,41	24000	fake	testing	testing\es\es_00512.mp3
edgeTTS	es_00513.mp3	mp3	5,45	24000	fake	testing	testing\es\es_00513.mp3

Tabla 33. Fragmento de Ficha de Análisis – Edge TTS

Anexo 18. Fragmento de Ficha de Análisis – ElevenLabs

Nombre del dataset: ElevenLabs							
Origen	Nombre del Archivo	Formato	Duración (s)	Frecuencia (Hz)	Tipo	Subconjunto	Ruta
ElevenLabs	el_0001.mp3	mp3	3,99	44100	fake	train	fake\ElevenLabs\elevenlabs\audio_output\train\en\el_0001.mp3
ElevenLabs	el_0002.mp3	mp3	3,81	44100	fake	train	fake\ElevenLabs\elevenlabs\audio_output\train\en\el_0002.mp3
ElevenLabs	el_0902.mp3	mp3	3,02	44100	fake	val	fake\ElevenLabs\elevenlabs\audio_output\val\en\el_0902.mp3
ElevenLabs	el_2401.mp3	mp3	3,72	44100	fake	test	fake\ElevenLabs\elevenlabs\audio_output\test\en\el_2401.mp3
ElevenLabs	el_2402.mp3	mp3	4,92	44100	fake	test	fake\ElevenLabs\elevenlabs\audio_output\test\en\el_2402.mp3
ElevenLabs	el_2403.mp3	mp3	4,88	44100	fake	test	fake\ElevenLabs\elevenlabs\audio_output\test\en\el_2403.mp3

Tabla 34. Fragmento de Ficha de Análisis – ElevenLabs

Anexo 19. Fragmento de Ficha de Análisis – Higgs Audio

Nombre del dataset: Higgs Audio							
Origen	Nombre del Archivo	Formato	Duración (s)	Frecuencia (Hz)	Tipo	Subconjunto	Ruta
Higgs Audio	higgs_0001.mp3	mp3	4,92	24000	fake	train	output\higgs_audio\audio\higgs_0001.mp3
Higgs Audio	higgs_0002.mp3	mp3	4,64	24000	fake	train	output\higgs_audio\audio\higgs_0002.mp3
Higgs Audio	higgs_0146.mp3	mp3	3,24	24000	fake	val	output\higgs_audio\audio\higgs_0146.mp3
Higgs Audio	higgs_0147.mp3	mp3	4	24000	fake	val	output\higgs_audio\audio\higgs_0147.mp3
Higgs Audio	higgs_0216.mp3	mp3	6,2	24000	fake	test	output\higgs_audio\audio\higgs_0216.mp3
Higgs Audio	higgs_0217.mp3	mp3	4,48	24000	fake	test	output\higgs_audio\audio\higgs_0217.mp3
Higgs Audio	higgs_0218.mp3	mp3	4,32	24000	fake	test	output\higgs_audio\audio\higgs_0218.mp3

Tabla 35. Fragmento de Ficha de Análisis – Higgs Audio

Anexo 20. Fragmento de Ficha de Análisis – Kokoro

Nombre del dataset: Kokoro							
Origen	Nombre del Archivo	Formato	Duración (s)	Frecuencia (Hz)	Tipo	Subconjunto	Ruta
Kokoro	kokoro_am_adam_0001.wav	wav	5,28	24000	fake	train	Documents\kokoro_output\kokoro_am_adam_0001.wav
Kokoro	kokoro_am_adam_0002.wav	wav	3,83	24000	fake	train	Documents\kokoro_output\kokoro_am_adam_0002.wav
Kokoro	kokoro_am_michael_0002.wav	wav	6,83	24000	fake	val	Documents\kokoro_output\kokoro_am_michael_0002.wav
Kokoro	kokoro_am_michael_0003.wav	wav	7,03	24000	fake	val	Documents\kokoro_output\kokoro_am_michael_0003.wav
Kokoro	kokoro_am_eric_0002.wav	wav	6,33	24000	fake	test	Documents\kokoro_output\kokoro_am_eric_0002.wav
Kokoro	kokoro_am_eric_0003.wav	wav	4,53	24000	fake	test	Documents\kokoro_output\kokoro_am_eric_0003.wav

Tabla 36. Fragmento de Ficha de Análisis – Kokoro

Anexo 21. Fragmento de Ficha de Análisis – Piper TTS

Nombre del dataset: Piper TTS							
Origen	Nombre del Archivo	Formato	Duración (s)	Frecuencia (Hz)	Tipo	Subconjunto	Ruta
Piper TTS	piper_amy_0001.wav	wav	4,28	22050	Fake	train	Documents\piper_output\piper_amy_0001.wav
Piper TTS	piper_amy_0002.wav	wav	4,63	22050	Fake	train	Documents\piper_output\piper_amy_0002.wav
Piper TTS	piper_tom_0002.wav	wav	5,22	44100	Fake	val	Documents\piper_output\piper_tom_0002.wav
Piper TTS	piper_tom_0003.wav	wav	3,01	44100	Fake	val	Documents\piper_output\piper_tom_0003.wav
Piper TTS	piper_davefx_0001.wav	wav	3,41	22050	Fake	test	Documents\piper_output\piper_davefx_0001.wav
Piper TTS	piper_davefx_0002.wav	wav	3,97	22050	Fake	test	Documents\piper_output\piper_davefx_0002.wav
Piper TTS	piper_davefx_0003.wav	wav	2,37	22050	Fake	test	Documents\piper_output\piper_davefx_0003.wav

Tabla 37. Fragmento de Ficha de Análisis – Piper TTS

Anexo 22. Caso N° 01 - Iniciar sesión y acceder a las funcionalidades protegidas



Figura 16. Caso de uso N° 01 - Iniciar sesión y acceder a las funcionalidades protegidas

Anexo 23. Caso N° 02 - Cargar y validar el formato del archivo del audio

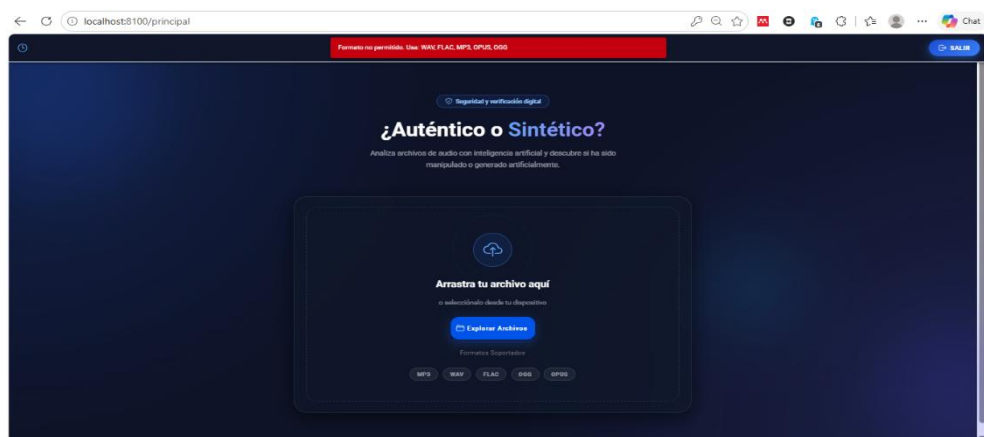


Figura 17. Caso de uso N° 02 - Cargar y validar el formato del archivo del audio

Anexo 24. Caso N° 03 - Validar la duración del archivo de audio

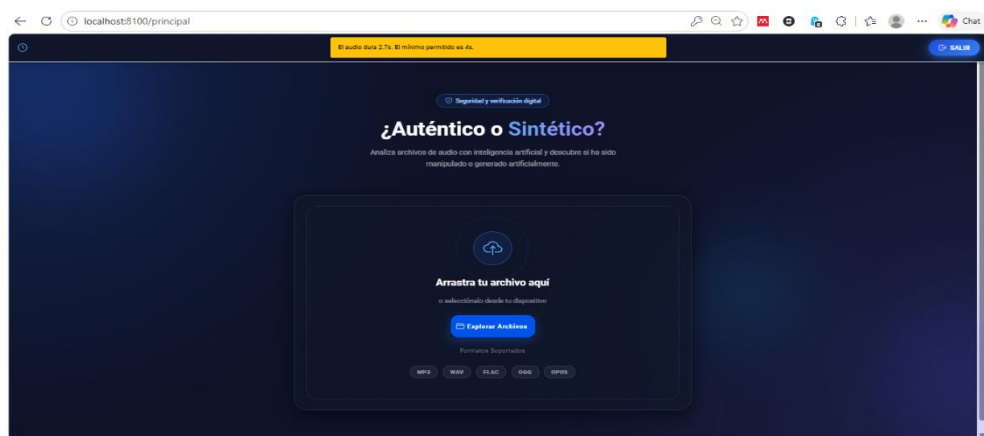


Figura 18. Caso de uso N° 03 - Validar la duración del archivo de audio

Anexo 25. Caso N° 04 - Analizar y clasificar una muestra de audio

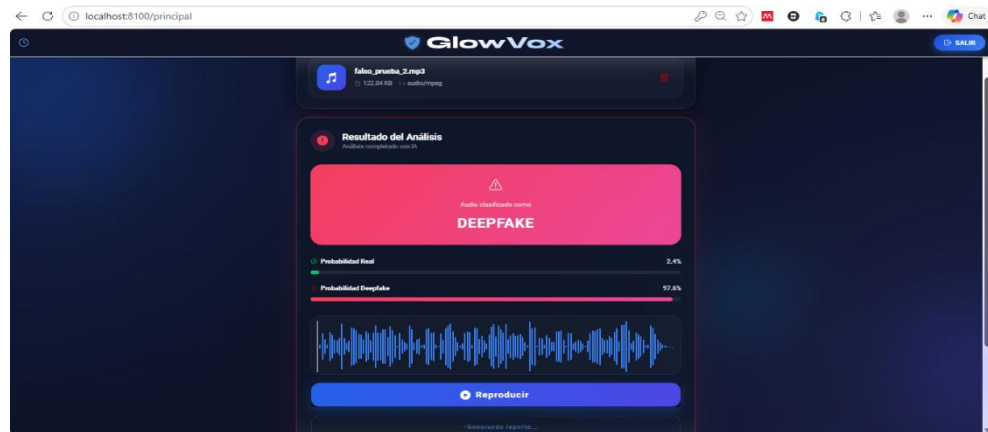


Figura 19. Caso de uso N° 04 - Analizar y clasificar una muestra de audio

Anexo 26. Caso N° 05 - Generar y descargar el reporte del análisis



Figura 20. Caso de uso N° 05 - Generar y descargar el reporte del análisis