Aumento de datos para detección de deterioro cognitivo en habla espontánea

Por:

Migan Giuseppe Galban Pineda

Tesis sometida como requisito parcial
para obtener el grado de:

MAESTRÍA EN CIENCIAS EN EL ÁREA DE CIENCIAS COMPUTACIONALES

en el

Instituto Nacional de Astrofísica,
Óptica y Electrónica

Diciembre, 2024

Tonantzintla, Puebla

Dirigida por:

Dr. Luis Villaseñor-Pineda

Dr. Manuel Montes-y-Gómez

©INAOE 2024

Derechos Reservados

El autor otorga al INAOE el permiso de reproducir y distribuir copia de esta tesis en su totalidad o en partes mencionando la fuente

Tabla de contenido

Lista de figuras

Lista de tablas

Agradecimientos

Primeramente, agradecer a Dios y a mis padres que siempre me han apoyado; a la Secretaría de Ciencia, Humanidades, Tecnología e Innovación de México y al Instituto Nacional de Astrofísica, Óptica y Electrónica por darme la oportunidad de participar en este posgrado. También quisiera agradecer a todas esos amigos y amigas que compartieron conmigo buenos y malos momentos, los que ahora solo serán bonitos recuerdos, agradezco a todos los que estuvieron pendientes de mí, alentándome a seguir adelante de diferentes maneras, ya que sin ellos todo habría sido menos llevadero. Para terminar quiero agradecer a mis tutores y sinodales por sus valiosas apreciaciones y consejos.

Resumen

El procesamiento de voz a través de técnicas de aprendizaje profundo (*deep learning*) ha alcanzado excelentes resultados en los últimos años. No obstante, es necesario contar con grandes cantidades de datos para poder trabajar con dichos modelos. La carencia o falta de datos es un reto por enfrentar en este tipo de situaciones. En particular, dentro del ámbito de las aplicaciones médicas esto se convierte en un importante cuello de botella. En este trabajo se aborda dicho problema mediante la exploración de diferentes técnicas de aumento de datos para la detección de deterioro cognitivo leve en voz espontánea, además de los procesos y análisis que esto conlleva; es por esto que el presente trabajo consideró diferentes aspectos como: (i) técnicas de extracción de características en audios, (ii) el aprovechamiento de modelos preentrenados en audio para extracción de *embeddings*, (iii) técnicas de aumento agregando datos reales de una clase similar, (iv) técnicas de aumento que modifican la señal de audio, (v) técnicas de aumento que usan la representación del elemento, y (vi) técnicas de aumento usando redes neuronales profundas generativas. Bajo este último punto, específicamente se analizaron dos variantes de la *Wasserstein Generative Adversarial Networks with Gradient Penalty* (WGAN-GP), que usan capas convolucionales unidimensionales, para la generación de datos sintéticos a partir de *embeddings* de audio.

Mediante la experimentación con la colección de datos de la competencia (*Alzheimer's Dementia Recognition through Spontaneous Speech (audio only)*) ADReSSo, que contiene audio de entrevistas diagnósticas a personas con deterioro cognitivo y sanos; se encontró que la adición de datos de pacientes con deterioro cognitivo avanzado, así como la aplicación de técnicas de aumento que modifican la señal de audio, mejoran la detección del deterioro cognitivo leve. Aunque los modelos generativos no superaron el rendimiento de muchas de estas técnicas de aumento, este trabajo representa una primera exploración del uso de redes neuronales profundas generativas para la generación de datos sintéticos en la detección del deterioro cognitivo leve.

En general, este trabajo demuestra el potencial del aumento de datos artificiales para mejorar la detección del deterioro cognitivo leve a partir del habla espontánea, sentando las bases para futuras investigaciones en el uso de técnicas de *deep learning* que ayuden en el diagnóstico del deterioro cognitivo leve, relacionado con la Enfermedad de Alzheimer.

Abstract

Speech processing through deep learning techniques has achieved excellent results in recent years. However, it is necessary to have large amounts of data to be able to work with such models. The lack of data is a challenge to be faced in this type of situation. Particularly, within the field of medical applications this becomes a major bottleneck. This work addresses this problem by exploring different data augmentation techniques for the detection of Mild Cognitive Impairment in speech, in addition to the processes and analysis involved. The present work considers different aspects such as: (i) feature extraction techniques in audios, (ii) the exploitation of pre-trained models in audio for *embeddings* extraction, (iii) augmentation techniques adding real data from a similar class, (iv) augmentation techniques modifying the audio signal, (v) augmentation techniques using element representation, and (vi) augmentation techniques using deep generative neural networks, specifically two variants of Wasserstein Generative Adversarial Networks with Gradient Penalty (WGAN-GP), which use one-dimensional convolutional layers, for synthetic data generation from audio *embeddings*.

We experimented on the ADReSSo (Alzheimer's Dementia Recognition through Spontaneous Speech (audio only)) challenge data collection, which contains audio from diagnostic interviews of cognitively impaired and healthy people, it was found

that the addition of data from patients with advanced cognitive impairment, as well as the application of augmentation techniques that modify the audio signal, improve the detection of mild cognitive impairment. Although the generative models did not outperform these augmentation techniques, this work represents a first exploration of the use of generative deep neural networks for synthetic data generation in the detection of mild cognitive impairment using voice.

Overall this work demonstrates the potential of artificial data augmentation to improve the detection of Mild Cognitive Impairment from spontaneous speech, laying the groundwork for future research in the use of Deep Learning techniques to aid in the diagnosis of Alzheimer's-related Mild Cognitive Impairment.

Capítulo 1

Introducción

La enfermedad de Alzheimer (EA) es un tipo de demencia que afecta la memoria, el pensamiento y el comportamiento. Los síntomas se desarrollan lentamente y empeoran con el tiempo, llegando a interferir con las actividades cotidianas (6). Actualmente, se estima que más de 55 millones de personas padecen de demencia a nivel mundial. Cada año se registran millones de nuevos casos, de los cuales aproximadamente el 60% se producen en países de ingresos medios y bajos. La demencia es actualmente la séptima causa principal de muerte y una de las principales causas de discapacidad y dependencia entre las personas mayores a nivel mundial, donde entre el 60 y el 70% de los casos corresponden a la EA (46).

Dada la situación previamente descrita, es fundamental detectar el deterioro cognitivo a tiempo, con el fin de que las personas diagnosticadas puedan recibir tratamiento oportuno que ralentice el avance de la enfermedad. El deterioro cognitivo leve MCI por sus siglas en Inglés (*Mild Cognitive Impairment*) es una etapa temprana de diferentes tipos de demencia, incluyendo la EA. La detección del MCI permitirá al paciente tomar decisiones bien informadas para atender su condición de una forma oportuna. De esta manera, el o la paciente estará en posición de cambiar su estilo de vida (alimento, ejercicio, etc.) lo que llevará a mejorar la calidad de ésta. Se ha demostrado que diversos tipos de demencia, incluida la EA, afectan el habla y

el lenguaje (63). También se ha observado que la presencia de pausas y la duración de los silencios al generar el habla en personas con EA puede ser un indicador claro de la pérdida de capacidad cognitiva (52). Esto sugiere que el habla espontánea es una fuente valiosa para la detección temprana del deterioro cognitivo. Diferentes investigaciones han mostrado la eficacia de las técnicas automáticas para este propósito. Un ejemplo es el *Interdisciplinary Longitudinal Study on Adult Development and Aging* (ILSE) (64), que recopila datos y realiza un seguimiento a adultos mayores en dos centros urbanos de Alemania. Otro ejemplo es la competencia *Alzheimer's Dementia Recognition through Spontaneous Speech* (ADReSS), promovido por el *Usher Institute of The Edinburgh Medical School* (59), presentado en diversas conferencias (2; 3). En esta competencia se analizan audios grabados de entrevistas donde los pacientes describen la imagen del "robo de galletas" del examen *Boston Diagnostic Aphasia*.

Las descripciones de imágenes son comúnmente utilizadas para observar déficits en el lenguaje, provocados por el envejecimiento normal, la demencia o accidentes cerebrovasculares; la imagen del "robo de galletas" se utiliza para evaluar la gravedad de la afasia en los pacientes, un trastorno que afecta directamente el lenguaje, tanto oral como escrito (13), la presencia de este trastorno está relacionado con el deterioro cognitivo, asociado a demencias como el Alzheimer.

El objetivo es encontrar métodos automáticos que apoyen a los profesionales de la salud, siendo el ideal aquel que permita estimar, a partir del habla espontánea del paciente, el puntaje correspondiente en el *Mini Mental Status Examination* (MMSE) (26) con el cual se valora el deterioro cognitivo de una persona con un puntaje de 1 a 30; donde una persona con deterioro cognitivo avanzado CI (*Cognitive Impairment*), va de un puntaje de 1 a 23; de 24 a 28 puntos, aproximadamente, están las personas con deterioro cognitivo leve MCI (*Mild Cognitive Impairment*) y alrededor de los 26 a 30 puntos, se encuentra el rango de las personas sin deterioro cognitivo o cognitivamente normal CN (*Cognitively Normal*) por sus siglas en Inglés.

Dada la experiencia que se tiene actualmente con los modelos de *deep learning* utilizados como método de aumento de datos, éstos arrojan mejores resultados que los métodos tradicionales (57), existe un gran interés en aplicarlos a esta tarea. No obstante, el entrenamiento de estos modelos exige grandes cantidades de muestras con el fin de que logren abstraer de manera eficaz las características que se requieren (65). Si bien existen estrategias que ayudan a mitigar esta problemática, como el transfer learning (utilizar modelos pre-entrenados), el aprendizaje semi-supervisado (aprovechar datos no etiquetados) y auto-supervisado (generar tareas de aprendizaje sin etiquetas), estas suelen aplicarse en etapas distintas del proceso y podrían ser exploradas como complementos en trabajos futuros. En esta tesis, se priorizó el aumento de datos como estrategia principal para abordar la limitación de datos, permitiendo enfocar esfuerzos en este método.

Gracias a la técnica de aumento de datos que modifican la señal de audio, se robustece el modelo de clasificación entrenado, incrementando los escenarios posibles. En el caso del audio, es posible considerar los cambios en un escenario como: ruido ambiente, tonos diferentes de voz y variaciones de volumen por diferentes circunstancias, entre otros. Además de la técnica anterior, en este proyecto también se evalúan técnicas de aumento agregando datos reales de una clase similar, técnicas de aumento que usan la representación del elemento, incluyendo el uso de redes neuronales profundas generativas, específicamente dos variantes de las *Wasserstein Generative Adversarial Networks with Gradient Penalty* (WGAN-GP), que usan capas convolucionales unidimensionales, para la generación de datos sintéticos a partir de *embeddings* de audio, con el fin de mejorar la detección de deterioro cognitivo leve en habla espontánea a partir de grabaciones de audio.

Cabe mencionar que se realizaron pruebas preliminares con un modelo generativo basado en modelos de difusión y transformers, denominado "TorToise" (14),

con el objetivo de clonar las voces de los pacientes y generar elementos de audio artificial que pudieran servir como aumento de datos. Los resultados obtenidos en dichas pruebas no fueron satisfactorios para el propósito de esta tesis, por lo que esta opción fue descartada. Las dificultades encontradas en la generación directa de audio, motivaron la búsqueda de alternativas; esta experiencia previa fundamentó la presente propuesta, que explora el aumento de datos a partir de la manipulación de representaciones del audio y embeddings neuronales.

1.1 Planteamiento del problema

Para la tarea de detección de deterioro cognitivo leve en habla espontánea existen muy pocas colecciones de datos disponibles, y las que hay suelen ser limitadas, contando con solo unas pocas decenas de grabaciones etiquetadas. Aunque las redes neuronales profundas han mostrado un mejor desempeño que las técnicas de aprendizaje de máquina tradicionales en la mayoría de sus aplicaciones, es bien sabido que estos modelos requieren grandes cantidades de datos para funcionar de manera óptima. Por ello, es necesario generar más datos mediante diversas técnicas de aumento. Estas pueden incluir métodos que realizan variaciones directamente en los audios, así como enfoques más avanzados, que realizan el aumento a nivel de su representación o *embeddings*.

1.1.1 Preguntas de investigación

¿Qué tan efectivas son las distintas técnicas de extracción de características en audio para la detección del deterioro cognitivo leve en habla espontánea?

¿Qué técnicas de aumento de datos son más efectivas para la detección del deterioro cognitivo leve en habla espontánea: las que modifican la señal de audio

directamente o las que trabajan sobre las características extraídas del audio?

Dentro de las técnicas que trabajan sobre las características extraídas del audio, ¿Cuáles son las más efectivas, las convencionales como SMOTE o las basadas en Redes Generativas Adversarias (GAN)?

1.2 Objetivos

1.2.1 Objetivo general

Diseñar y evaluar un método de aumento de datos basado en redes generativas adversarias para la tarea de detección de deterioro cognitivo leve en habla espontánea.

1.2.2 Objetivos específicos

Los objetivos específicos se listan a continuación:

- Implementar y evaluar varias técnicas de extracción de características acústicas en la detección del deterioro cognitivo en voz.
- Implementar y evaluar técnicas de aumento de datos aplicados directamente al audio para mejorar la detección del deterioro cognitivo en voz.
- Implementar y evaluar técnicas de aumento de datos aplicados sobre distintas representaciones del audio para mejorar la detección del deterioro cognitivo en voz.
- Proponer un método basado en redes generativas adversarias para la generación de datos artificiales a partir de un dataset de habla espontánea, orientados a la detección de deterioro cognitivo leve.

- Comparar las distintas técnicas de aumento de datos implementadas, con el fin de identificar la que más aporte tiene a la detección del deterioro cognitivo en voz.

1.3 Contribuciones

Las contribuciones de esta tesis se centran en la exploración y evaluación de diferentes técnicas de aumento de datos, haciendo énfasis en aquellas que se aplican sobre las distintas representaciones vectoriales de los audios.

A continuación, se detallan las principales contribuciones de este trabajo:

- Se contrastaron técnicas convencionales de aumento de datos en audio y métodos avanzados de generación de datos sintéticos basados en *deep learning*, específicamente utilizando GANs, aplicados a la detección del deterioro cognitivo leve, evaluando características independientes del contexto temático. Si bien las técnicas basadas en deep learning mejoraron la clasificación, este análisis estableció que no superan a las técnicas clásicas en este contexto, sugiriendo la necesidad de explorar ajustes en el uso de redes generativas o enfocarse en el uso de técnicas más sencillas que funcionan con tan pocos datos.

- Se propuso la aplicación de las GANs para la generación de datos sintéticos basados en *embeddings* neuronales que representan el audio, en forma de vectores estadísticos unidimensionales, en lugar de espectrogramas (representación del audio en imágenes), como se hace en otros estudios, donde se usan GANs para generar datos relacionados al audio.

- Se realizaron pruebas para comprobar si la inclusión de datos de pacientes con deterioro cognitivo más avanzado CI por sus siglas en Inglés *Cognitive Impairment*, podrían compartir similitudes con las de pacientes con MCI y mejorar su

clasificación. Los resultados mostraron una mejora significativa, confirmando la validez de esta estrategia y sugiriendo que la información acústica en etapas más avanzadas de deterioro cognitivo puede ser relevante para la detección temprana de MCI.

1.4 Alcance y limitaciones

La presente investigación se enfocó en la evaluación de técnicas de aumento de datos para la detección temprana del deterioro cognitivo leve a través del análisis del habla. Si bien el estudio exploró diversas metodologías y se obtuvieron resultados prometedores, es importante delimitar su alcance y reconocer las limitaciones inherentes al proyecto. A continuación, se detallan las características y restricciones que enmarcan la investigación.

- Se centró en la evaluación de diferentes técnicas de aumento de datos para la detección del deterioro cognitivo leve a partir del análisis de la voz, utilizando el conjunto de datos del DementiaBank para el reto ADReSSo de Interspeech 2021.
- Se delimitó el estudio a la clasificación de MCI vs. CN, utilizando un subconjunto del conjunto de datos original (SubADReSSo), que incluye solo a pacientes con puntuaciones MMSE entre 24 y 30.
- Se exploraron exclusivamente las características acústicas de la voz, sin considerar el contenido lingüístico de las grabaciones. Esto permite evaluar el potencial de los marcadores acústicos para la detección de MCI aprovechando la independencia del contexto en el diálogo.

Por otro lado, cabe mencionar que el presente estudio se realizó en grabaciones del idioma Inglés, debido al conjunto de datos utilizado. A pesar de que se están

trabajando con características intrínsecas de la voz, que debería ser independientes de los idiomas, para verificar que el estudio es independiente del idioma, es necesario realizar pruebas con otros conjuntos de datos para comprobar el rendimiento de los métodos evaluados.

Finalmente, es importante señalar que este trabajo no busca reemplazar el diagnóstico médico ni las herramientas clínicas existentes para la detección del deterioro cognitivo leve. La idea es explorar y desarrollar un sistema de apoyo basado en el análisis de voz que, en un futuro, pueda servir como una herramienta de alerta temprana para la detección de deterioro cognitivo, complementando la evaluación médica y brindando información adicional para un diagnóstico médico más preciso.

1.5 Organización de tesis

Los capítulos se estructuran de la siguiente manera:

- **Capítulo 2: Marco teórico.** Se presentan los fundamentos teóricos relacionados con el Alzheimer y el deterioro cognitivo, técnicas de procesamiento de voz, técnicas de aumento de datos tradicionales y usando modelos de *deep learning*.
- **Capítulo 3: Trabajo relacionado.** Se realiza una revisión de la literatura sobre métodos existentes en la detección del deterioro cognitivo a través del análisis de voz y el uso de técnicas de aumento de datos.
- **Capítulo 4: Evaluación de Representaciones Acústicas usadas en Detección de MCI.** Se describe detalladamente la metodología propuesta, para la selección de datos, preprocesamiento, extracción y evaluación de características, esto con el fin de dar un punto de referencia inicial o *baseline*, con el cual comparar el aumento de datos y establecer las técnicas base a utilizar.

- **Capítulo 5: Análisis Comparativo de Estrategias de Aumento de Datos Acústicos para la Detección de MCI.** Se presenta la implementación de modelos y técnicas de aumento de datos, en conjunto con los resultados obtenidos, acompañados de un análisis y discusión sobre su significado y relevancia.

- **Capítulo 5: Conclusiones y Trabajo futuro.** Se resumen las conclusiones principales del trabajo y se proponen líneas de investigación futuras.

CAPÍTULO 2

MARCO TEÓRICO

Este capítulo establece las bases teóricas que sustentan la presente investigación sobre la detección del deterioro cognitivo leve a través del análisis de la voz. Se exploran los fundamentos de la enfermedad de Alzheimer (EA) y su relación con el MCI, destacando la importancia de la detección temprana para una intervención oportuna y una mejor calidad de vida para los pacientes. Se abordan las técnicas de procesamiento de voz, centrándonos en la extracción de características acústicas relevantes para la tarea. Se describen diferentes métodos, desde los clásicos Coeficientes Cepstrales de Frecuencias Mel (MFCC) y los coeficientes Chroma, el conjunto de características eGeMAPS, hasta las representaciones de *embeddings* neuronales más modernas, como *wav2vec* y *wav2vec 2.0*. Además, se revisan algunas técnicas de aumento de datos, cruciales para abordar la escasez de datos en el ámbito acústico. Se detallan métodos que modifican directamente la señal de audio, como *Gain Transition, Pitch Shift, Polarity Inversion, Shift* y *Time Stretch*, junto con técnicas aplicadas a la representación del audio, como SMOTE y, de manera innovadora, las Redes Generativas Adversarias (GANs), incluyendo las variantes condicionales (cGAN) y Wasserstein con penalización de gradiente (WGAN-GP). Este capítulo proporciona el marco conceptual esencial para comprender el desarrollo y los resultados de la investigación presentada en los capítulos posteriores, destacando la novedad de aplicar GANs directamente sobre *embeddings* de audio para la detección

del deterioro cognitivo.

2.1 La enfermedad de Alzheimer y el deterioro cognitivo leve

La enfermedad de Alzheimer es una condición neurodegenerativa compleja con múltiples factores de influencia, incluyendo la predisposición genética, el estilo de vida y la presencia de otras enfermedades. Actualmente, la demencia afecta a más de 55 millones de personas a nivel mundial. Cada año, se registran millones de nuevos casos, de los cuales el 60% se producen en países de ingresos bajos y medianos (46). La demencia es actualmente la séptima causa principal de muerte y una de las principales causas de discapacidad y dependencia entre las personas mayores a nivel mundial, donde entre el 60 y el 70% de los casos es de Alzheimer (46). Desafortunadamente, no se tiene una cura para esta enfermedad. De ahí la importancia de determinar esta enfermedad en etapas tempranas. La persona diagnosticada puede recibir información oportuna para ralentizar el avance de la enfermedad, y además podrá planificar el futuro para tomar decisiones sobre el cuidado posterior al agravamiento de su enfermedad. Con ello, las personas enfermas tendrían una mejor calidad de vida, con más tiempo para compartir momentos especiales y disfrutar de relacionarse con sus seres queridos.

El MCI es detectable en pruebas neuropsicológicas que pueden evaluar diferentes dominios cognitivos como lenguaje verbal, oral, atención, memoria y habilidades de coordinación visual (54); el MCI se trata de un pequeño deterioro que no afecta significativamente la vida diaria, este se considera un síntoma temprano de demencia, incluyendo la EA (37). Además, las personas con MCI experimentan un deterioro cognitivo más acelerado que el envejecimiento normal, especialmente en áreas como la memoria y la velocidad de procesamiento (37).

Existen pruebas de que los cambios en el habla pueden anticipar el inicio clínico de la EA en varios años. Los cambios en el lenguaje oral pueden manifestarse en aspectos como el tono de voz, la velocidad del habla, la recuperación de palabras y la fluidez del discurso. Las personas mayores con la enfermedad de Alzheimer o con deterioro cognitivo leve, experimentan estos cambios en el lenguaje, incluyendo la producción de pausas más largas y disfluencias en la conversación (23) (repeticiones, prolongaciones de sonidos, bloqueos, muletillas, o cambios en la velocidad y ritmo de las palabras). De igual forma, los parámetros de prosodia del habla se han mostrado efectivos para detectar EA y MCI, ya que reflejan patrones en el ritmo, entonación y tono de voz (30).

2.2 Extracción de Características en la voz

La extracción de características acústicas relevantes es fundamental para la detección del deterioro cognitivo leve en habla espontánea. En esta sección se describirán los métodos utilizados para extraer características que capturan diferentes aspectos del sonido, como el contenido espectral, la dinámica temporal y la melodía. Estas técnicas incluyen los Coeficientes Cepstrales de Frecuencias Mel (MFCC) y sus deltas, los coeficientes Chroma y los parámetros acústicos del conjunto eGeMAPS, así como el proceso de extracción de *embeddings* neuronales usando los modelos *wav2vec* y *wav2vec 2.0*.

2.2.1 MFCC: Coeficientes Cepstrales de Frecuencias Mel

En el campo del análisis de audio, la transformación de señales complejas en representaciones concisas y relevantes es crucial. Los Coeficientes Cepstrales de Frecuencias Mel (MFCC), junto con sus Deltas, proporcionan una representación numérica compacta y significativa del sonido, preservando información esencial sobre su con-

tenido espectral y su evolución temporal. Esta representación[1], se presenta como una herramienta ideal para alimentar algoritmos de aprendizaje automático. En el marco de esta tesis, los MFCC se emplearán como una de las herramientas para analizar las características acústicas del habla en el contexto de la detección del deterioro cognitivo.

La extracción de los Coeficientes Cepstrales de Frecuencias Mel se lleva a cabo mediante un proceso que comprende tres etapas.

1. **Segmentación temporal:** Se divide el audio en secciones de corta duración denominadas "ventanas". A cada una de estas ventanas se le aplica la Transformada Rápida de Fourier (FFT) para obtener su espectro de frecuencias. La longitud de las ventanas y el solapamiento entre ellas son parámetros ajustables que permiten adaptar el análisis a las particularidades del sonido.

2. **Mapeo perceptivo del espectro de frecuencias utilizando la escala Mel:** Esta escala, que simula la percepción no lineal de las frecuencias por parte del oído humano, otorga mayor relevancia a las bandas frecuenciales más significativas para nuestra audición.

3. **Transformada Coseno Discreta de tipo II (DCT-II):** Se aplica la DCT-II al espectro de frecuencias mapeado en la escala Mel. Esta transformada representa una señal como una suma ponderada de funciones coseno. Esencialmente, "codifica" la señal al describirla en términos de componentes cosenoidales. En pocas palabras La DCT-II descompone una señal en una serie de ondas coseno con diferentes frecuencias, donde cada coeficiente DCT-II indica cuánto de cada onda coseno está presente en la señal original. Esta transformación matemática permite condensar la información espectral en un número reducido de coeficientes, extrayendo la "esencia" del sonido en cada ventana y descartando detalles irrelevantes.

[1] Obtenida a través de la librería `Librosa` (lib)

Para una mejor comprensión de la manera en que se extraen los coeficientes MFCC se ejemplifica la Figura 2.1; donde se observa el desplazamiento de las ventanas de 25ms, cada 10ms, para obtener los coeficientes MFCC por cada ventana y finalmente, representar la distribución temporal de estos MFCC en estadísticos funcionales (promedio, mediana, desviación estándar, asimetría y curtosis).

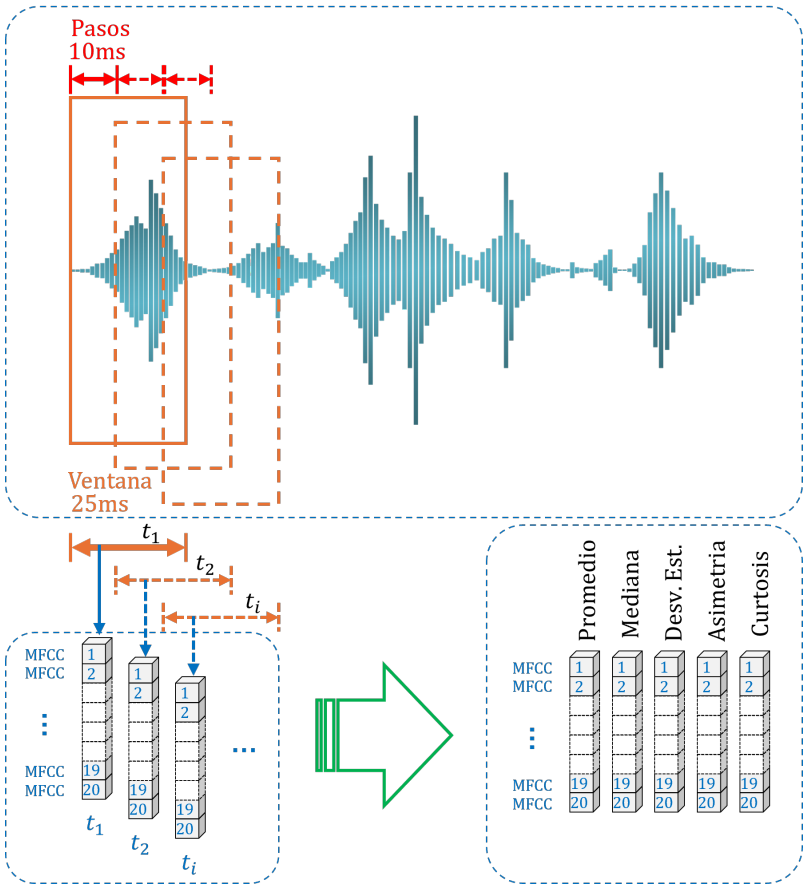


Figura 2.1: Representación Gráfica de la extracción de los 20 coeficientes MFCC como estadísticos funcionales.

Además de los MFCC, se pueden calcular los **Delta MFCC** para reflejar la dinámica temporal del sonido. Estos coeficientes representan la tasa de cambio de los MFCC a lo largo del tiempo. Los Delta MFCC proporcionan información valiosa sobre la evolución de las características acústicas a lo largo del tiempo, complementando la información espectral inherente a los MFCC (35).

2.2.2 Coeficientes Chroma

Los coeficientes Chroma actúan como un analizador musical que captura la "esencia melódica" del sonido, más allá de las notas exactas. En lugar de representar la intensidad de cada frecuencia individual, los Chroma se enfocan en la distribución de energía en las 12 notas de la escala musical (semitonos), creando una especie de "ADN melódico" del sonido. Este análisis melódico se hace de la siguiente manera:

1. **Segmentación temporal:** El audio se divide en pequeñas "ventanas" de tiempo, en este caso, de 50 milisegundos con un avance de 10 ms.

2. **Mapeo del sonido a la escala musical:** Se calcula el espectro de frecuencia de cada ventana y se "mapea" a una escala de semitonos utilizando filtros triangulares. Este es un filtro que resalta las notas musicales y atenúa las frecuencias que no encajan en la escala.

3. **Generación del perfil melódico:** Para cada semitono, se suma la energía de las frecuencias que caen dentro de su rango, creando un vector de 12 valores que representan la "intensidad" de cada nota en la escala.

Para un mejor entendimiento de los coeficientes Chroma, se propone un breve ejemplo; imagine que está escuchando una canción en la que se toca un acorde de Do Mayor. Los Chroma mostrarían una mayor intensidad en las notas Do, Mi y Sol (son las notas que componen el acorde de Do Mayor), sin importar si se tocan en octavas graves o agudas. Se podría decir que el Chroma identifica las 12 notas básicas que encontramos, en cada una de las secciones de un piano, como las vemos en la Figura 2.2

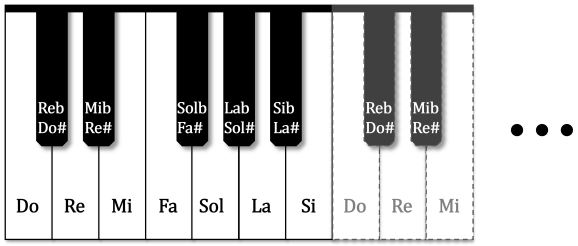


Figura 2.2: 12 Semitonos capturados por Chroma, representados en la sección de un Piano.

En el contexto del deterioro cognitivo, los Chroma podrían ser útiles para detectar cambios sutiles en la prosodia del habla. Por ejemplo, alteraciones en la entonación o la monotonía de la voz podrían reflejarse en cambios en el "ADN melódico" capturado por estos coeficientes.

2.2.3 eGeMAPS

El conjunto extendido de parámetros acústicos minimalistas de Geneva (eGeMAPS) es una herramienta valiosa para caracterizar la voz en el contexto del deterioro cognitivo[2]. Este conjunto, compuesto por 88 parámetros, se basa en la premisa de que los cambios emocionales y cognitivos se reflejan en la producción del habla. eGeMAPS ofrece una representación acústica completa que abarca:

Prosodia: Parámetros como el tono, el ritmo y las pausas.

Calidad de la Voz: Características como el *jitter* (variación en la frecuencia del tono, como un temblor en la voz) y el *shimmer* (variación en la amplitud del tono, como un vibrato irregular), que reflejan la estabilidad y tensión de las cuerdas vocales.

Características Espectrales: Incluyen formantes (resonancias del tracto vocal, estas son como las "huellas dactilares" de la boca, nariz y garganta en conjunto), la

[2]Obtenida mediante la librería `opensmile`

relación armónico-ruido y la pendiente espectral, que brindan información sobre el timbre de la voz y la distribución de energía en el espectro. Los formantes 2 y 3 son especialmente importantes para distinguir sonidos del habla.

MFCC 1-4: Coeficientes cepstrales que capturan la forma espectral global, complementando la información de otros parámetros.

Flujo Espectral: Mide la velocidad de cambio en el espectro, sensible a transiciones abruptas.

Ancho de banda de los formantes 2-3: Complementa la información de los formantes, ofreciendo una descripción más completa del tracto vocal. eGeMAPS aplica funciones estadísticas a estos parámetros para resumir su comportamiento temporal, facilitando la detección de patrones asociados al deterioro cognitivo (25).

2.2.4 Wav2vec

Este modelo fue creado para mejorar el reconocimiento automático de voz, mediante el entrenamiento no supervisado, para entrenar modelos de reconocimiento de voz mucho más eficientes, especialmente en situaciones con pocos datos etiquetados. Lo que hace el modelo es comprimir la señal de audio en representaciones compactas y luego intenta predecir la siguiente muestra de audio, diferenciándola de otras falsas. Este entrenamiento contrastivo le permite aprender la estructura temporal del audio y generar representaciones contextualizadas muy ricas. El modelo consta de dos etapas, primero codifica el audio crudo utilizando una red convolucional, luego una red de contexto, que combina estas representaciones para capturar la información contextual temporal.

Red de Codificación (Extractor de Características)

La red de codificación se puede imaginar como un embudo que transforma la señal de audio cruda en una representación más compacta y útil. Este "embudo" está formado por cinco capas convolucionales, cada una con filtros específicos (tamaños de kernel) que se deslizan sobre la señal con cierto paso (*strides*), en la Figura 2.3 se puede ver representada la señal de audio a la entrada de la capa de extracción de características como una **"X"** y la salida de esta capa, con la letra **"Z"**.

Los tamaños de kernel determinan el tamaño de la ventana que la red "observa" en cada paso para extraer características. Los pasos indican cuánto se mueve la ventana en cada capa. Un paso mayor significa que la red avanza más rápido, sacrificando información detallada por una visión más general. En esencia, esta red de codificación comprime la señal de audio de 16kHz (16.000 muestras por segundo); al procesar dichas muestras, genera representaciones de salida que se actualizan cada 10 ms, donde cada una de estas representaciones, capturan la información que hay en 30 ms del audio original, esto podríamos verlo a grandes rasgos, como que se analizan los datos en una ventana de 30ms y se va desplazando la ventana con un stride o paso de 10ms, por lo cual la representación de un segundo de audio estará dada por 100 datos; lo anterior sirve como base para la siguiente etapa del modelo, la red de contexto.

Red de Contexto

Se puede pensar en la Red de Contexto como un "detective" que examina las pistas generadas por la Red de Codificación para entender el contexto del audio. Esta red recibe la secuencia de representaciones de baja frecuencia y las combina utilizando nueve capas convolucionales. Cada capa tiene un tamaño de kernel de tres, lo que significa que analiza grupos de tres representaciones consecutivas para buscar patrones y relaciones. El paso de uno asegura un análisis detallado, moviéndose de

una representación a la siguiente sin saltos. Esta configuración, con nueve capas y un paso pequeño, permite a la red "ver" un amplio contexto temporal de aproximadamente 210 ms, uniendo las piezas del rompecabezas proporcionado por la Red de Codificación. El resultado es un "tensor contextualizado" que contiene información mucho más rica para el posterior análisis y la predicción de la siguiente muestra de audio, dicho tensor, se ve representado por la capa con la letra **"C"** en la Figura 2.3.

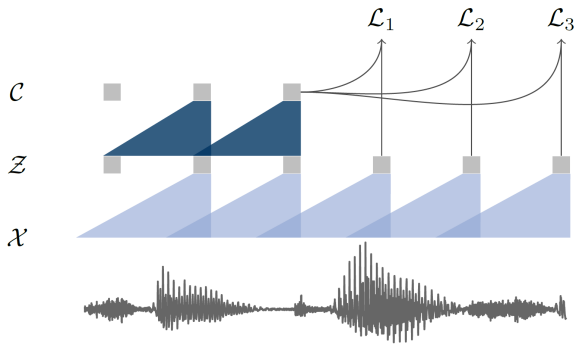


Figura 2.3: Estructura del modelo *wav2vec*, tomado de (55).

Para este proyecto se usó la variante del modelo "wav2vec large" con dos transformaciones lineales adicionales y una red de contexto de doce capas de kernel crecientes. Además, se agregan *skip connections* para mejorar la convergencia, alcanzando un campo receptivo final de unos 810 ms. Se elige esta arquitectura dado su entrenamiento con un campo receptivo de 810ms, puesto que lo que quisiéramos que se analice con el modelo *wav2vec* para nuestra tarea, es la forma en como se dicen las palabras y no solamente los fonemas.

Proceso de Entrenamiento

- **Entrada de Audio Crudo:** La señal de audio sin procesar se introduce al modelo.
- **Codificación Inicial:** Una red de codificación, compuesta por capas convolu-

cionales, comprime la señal en una representación de baja frecuencia que captura la información esencial del sonido.

- **Contextualización:** La red de contexto amplía la visión analizando múltiples pasos temporales de la representación codificada. Esto crea una representación contextualizada que entiende el flujo temporal del audio.
- **Predicción de Futuras Representaciones:** El modelo se entrena para predecir la siguiente muestra de audio, diferenciando la muestra real de otras falsas. Esta tarea de discriminación, guiada por la pérdida contrastiva, obliga al modelo a comprender patrones complejos en el audio.
- **Preentrenamiento en Datos Sin Etiquetar:** El modelo se preentrena en grandes cantidades de datos de audio sin etiquetar para aprender representaciones generales de audio.
- **Ajuste Fino con Datos Etiquetados:** Después del preentrenamiento, las representaciones aprendidas se utilizan como características de entrada para un modelo acústico que se ajusta finamente con datos de audio transcritos.

2.2.5 Wav2vec 2.0

El modelo *wav2vec 2.0* introduce un codificador de características con bloques convolucionales temporales, utiliza normalización por capas y activación GELU (*Gaussian Error Linear Unit*) por sus siglas en Inglés, además, integra una red transformer como red de contexto.

Red de Codificación (Extractor de Características)

La red de codificación en *wav2vec 2.0* actúa como un filtro avanzado que transforma la señal de audio cruda en representaciones latentes ricas y compactas, en la Figura

2.4 se puede ver representada la señal de audio a la entrada de la capa de extracción de características como una **"X"** y la salida de esta capa, con la letra **"Z"**. A diferencia de su predecesor, esta red está diseñada con una mayor complejidad y eficiencia, utilizando capas convolucionales más profundas y funciones de activación avanzadas. La red de codificación de *wav2vec 2.0* está formada por siete bloques de convoluciones temporales. Estas convoluciones están diseñadas para extraer características detalladas de la señal de audio. En este modelo se tienen tamaños de *kernel* de (10, 3, 3, 3, 3, 2, 2) y *strides* de (5, 2, 2, 2, 2, 2, 2). Después de cada convolución, las características son normalizadas mediante *Layer Normalization* y se les aplica la función de activación GELU. Esta combinación mejora la estabilidad del entrenamiento y la capacidad de modelado no lineal de la red. La red de codificación transforma la señal de audio de alta frecuencia en representaciones latentes de baja frecuencia. Estas representaciones generan representaciones de salida que se actualizan cada 20 ms, donde cada una captura la información que hay en 25 ms del audio original. En otras palabras, se analizan los datos en una ventana de 25ms y se va desplazando la ventana con un *stride* de 20ms, por lo cual la representación de un segundo de audio estará dada por 50 datos. Este embudo de características sirve como base para la red de contexto *Transformer*, la cual se encargará de contextualizar estas representaciones a lo largo de la secuencia de audio.

Red de Contexto

Es una red *Transformer* que toma las representaciones latentes del codificador de características y las transforma en representaciones contextualizadas y su salida, se ve representada con la letra **"C"** en la Figura 2.4.

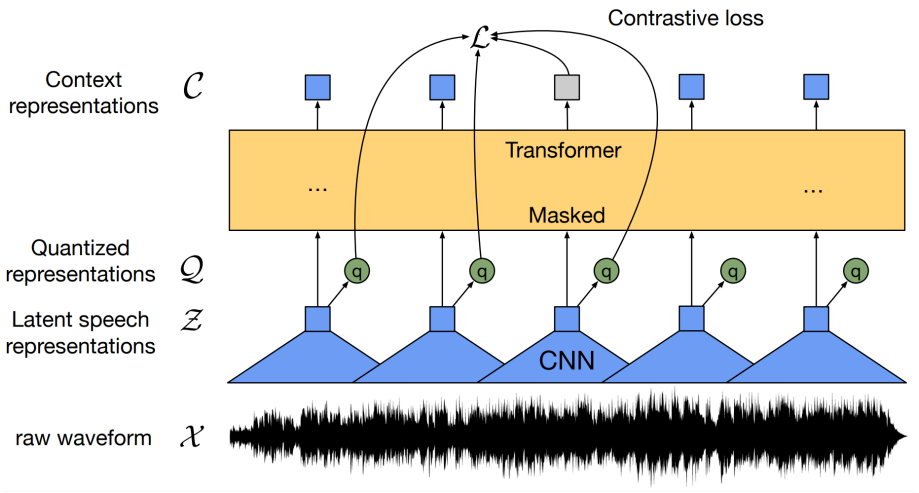


Figura 2.4: Estructura del modelo *wav2vec 2.0*, tomado de (8).

La red de contexto consiste en varias capas *Transformer* que procesan las representaciones latentes. La configuración básica (Base) tiene 12 capas *Transformer*, mientras que la configuración más grande (Large) tiene 24 capas; para tener una dimensión de salida de 768 para Base y 1024 para Large, dada en 8 cabezas de atención para BASE y 16 para Large. A pesar de que en nuestro caso, no se hace uso de la capa de contexto, donde se encuentra la diferencia entre el modelo Base y el Large, se eligió la configuración Large, dado que se entrenó para analizar muestras con una longitud superior.

Entrenamiento

El entrenamiento de *wav2vec 2.0* se realiza en dos etapas principales, preentrenamiento autosupervisado y ajuste fino supervisado.

- **Pre-Entrenamiento Auto-Supervisado:** En esta etapa, se enmascaran porciones del audio latente para que el modelo prediga las representaciones enmascaradas basándose en el contexto. El modelo se entrena para distinguir la representación latente verdadera, entre otros distractores, utilizando una pérdida

contrastiva y una pérdida de diversidad para asegurar el uso equitativo de las entradas del *codebook*.

- **Ajuste Fino Supervisado:** Después del preentrenamiento, se ajusta el modelo con datos etiquetados añadiendo una capa de proyección lineal y optimizando con una pérdida de Clasificación Temporal Conexionista (CTC). Este ajuste fino permite al modelo mapear las representaciones contextualizadas al espacio de etiquetas de audio, mejorando el rendimiento en tareas de reconocimiento de voz.

2.3 Técnicas para Aumento de Datos Acústicos que Modifican la Señal

En el ámbito del procesamiento de audio y el aprendizaje automático, la disponibilidad de datos es crucial para entrenar modelos efectivos y robustos. Sin embargo, en muchos casos, obtener una cantidad suficiente de datos puede ser un desafío, especialmente en aplicaciones médicas como la detección del deterioro cognitivo. Para abordar este problema, se utilizan técnicas de aumento de datos, las cuales permiten incrementar la cantidad y la diversidad del conjunto de datos disponibles mediante la aplicación de diversas transformaciones. A continuación, se describen cinco métodos de aumento de datos en audio, que consisten en la modificación de la señal.

2.3.1 Gain Transition (Transición de Ganancia)

La técnica de *Gain Transition* consiste en aumentar o disminuir gradualmente el volumen durante un periodo de tiempo aleatorio, también conocida como fundido de entrada y salida. Esta técnica se aplica de manera logarítmica, lo cual es natural para el oído humano.

El procedimiento funciona seleccionando dos niveles de ganancia: una ganancia inicial y una ganancia final. Luego, se elige un rango de tiempo durante el cual se realizará la transición entre estos dos niveles de ganancia. Esta transición puede empezar antes de que comience el audio o finalizar después de que el audio termine, lo que significa que el audio de salida puede comenzar o finalizar en medio de una transición. La ganancia permanece constante en el nivel inicial hasta que comienza la transición, luego cambia gradualmente hasta alcanzar la ganancia final, la cual se mantiene constante hasta el final del audio (32).

2.3.2 Pitch Shift (Desplazamiento de Tono)

La técnica de *Pitch Shift* consiste en cambiar el tono del sonido hacia arriba o hacia abajo sin alterar el tempo. En términos técnicos, esta técnica implica primero la expansión o contracción del tiempo (mediante el *vocoding* de fase) seguido de un nuevo muestreo. Es importante tener en cuenta que el *vocoding* de fase puede degradar la calidad del audio al "difuminar" los sonidos transitorios, alterar el timbre de los sonidos armónicos y distorsionar las modulaciones de tono. Esto puede resultar en una pérdida de nitidez, claridad o naturalidad en el audio transformado (32).

2.3.3 Polarity Inversion (Inversión de Polaridad)

La técnica de *Polarity Inversion* consiste en invertir las muestras de audio, "volteándolas de cabeza" y cambiando su polaridad. Esto se logra multiplicando la forma de onda por -1, de modo que los valores negativos se convierten en positivos y viceversa. El resultado sonará igual al original cuando se reproduce de forma aislada, pero puede diferir cuando se mezcla con otras fuentes de audio.

Esta técnica de inversión de forma de onda se utiliza para la cancelación de audio o para obtener la diferencia entre dos formas de onda. En el contexto del

aumento de datos de audio, esta transformación puede ser útil al entrenar modelos de aprendizaje automático sensibles a la fase (32).

2.3.4 Shift (Desplazamiento)

La técnica de *Shift* consiste en desplazar las muestras de audio hacia adelante o hacia atrás. Este desplazamiento puede realizarse con o sin repetición (rollover), es decir, las muestras movidas más allá del final del audio pueden volver a insertarse al comienzo o simplemente desaparecer. Esta técnica es útil para alterar la temporalidad del audio sin modificar otras características como el tono o el tempo (32).

2.3.5 Time Stretch (Estiramiento de Tiempo)

La técnica de *Time Stretch* consiste en cambiar la velocidad o la duración de la señal de audio sin alterar el tono. En términos técnicos, se emplea el *vocoding* de fase, un método que puede degradar la calidad del audio al "difuminar" los sonidos transitorios, alterar el timbre de los sonidos armónicos y distorsionar las modulaciones de tono. Esto puede resultar en una pérdida de nitidez, claridad o naturalidad en el audio transformado, especialmente cuando la tasa de cambio es extrema (32).

2.4 Técnicas de Aumento de Datos Aplicadas a su Representación

En aprendizaje automático, el desequilibrio de clases es un problema común que surge cuando una clase (la clase mayoritaria) está sobrerrepresentada en el conjunto

de datos en comparación con otra clase (la clase minoritaria). Esto puede sesgar el entrenamiento del modelo y perjudicar su capacidad para clasificar correctamente la clase minoritaria, que a menudo es la de mayor interés.

2.4.1 SMOTE (Synthetic Minority Over-sampling Technique)

Es una técnica diseñada para abordar el problema del desbalance de clases, mediante la creación de ejemplos sintéticos de la clase minoritaria, equilibrando así la distribución de clases en el conjunto de datos. SMOTE, no se limita a duplicar los ejemplos existentes de la clase minoritaria, (lo que podría llevar a un sobreajuste y una pobre generalización), sino que, crea nuevos ejemplos "sintéticos" interpolando entre los ejemplos existentes de la clase minoritaria (18). El proceso de creación de ejemplos sintéticos con SMOTE se puede resumir en los siguientes pasos:

- **Identificación de vecinos:** Para cada ejemplo de la clase minoritaria, SMOTE identifica sus k vecinos más cercanos en el espacio de características (normalmente $k = 5$).
- **Interpolación lineal:** Se selecciona aleatoriamente uno de los vecinos identificados en el paso anterior. Se crea un nuevo ejemplo sintético interpolando linealmente entre el ejemplo original y el vecino seleccionado. La interpolación se realiza en el espacio de características, creando un nuevo punto a lo largo del segmento de línea que conecta los dos ejemplos.
- **Repetición:** Los pasos 1 y 2 se repiten hasta que se genera el número deseado de ejemplos sintéticos. El grado de sobre muestreo es un parámetro configurable, lo que permite controlar la cantidad de ejemplos sintéticos que se añaden al conjunto de datos.

2.4.2 Red Generativa Adversaria (GAN)

Las GANs consisten en dos redes neuronales que compiten entre sí: un generador **G** que produce datos sintéticos a partir de ruido y un discriminador **D** que distingue entre datos reales y generados. La dinámica adversarial entre el generador y el discriminador se captura en la función objetivo minimax de la GAN, donde el generador busca minimizar esta función, engañando al discriminador para que clasifique los datos sintéticos como reales, mientras que el discriminador busca maximizarla, identificando correctamente tanto los datos reales como los sintéticos. Esto se ve representado en la ecuación 2.1.

$$\min_G \max_D \mathop{\mathbb{E}}_{x \sim p_{\text{data}}(x)} [\log D(x)] + \mathop{\mathbb{E}}_{z \sim p_z(z)} [\log(1 - D(G(z)))] \tag{2.1}$$

En la primera parte de la ecuación, el discriminador D trata de maximizar su capacidad para asignar correctamente la probabilidad de que un dato x proviene de la distribución real $p_{\text{data}}(x)$. Es decir, el discriminador intenta identificar los datos reales. En la segunda parte de la ecuación, el generador G, a partir de una variable aleatoria z extraída de una distribución de ruido $p_z(z)$, genera una muestra $G(z)$. El discriminador D intenta maximizar la probabilidad de clasificar correctamente estas muestras generadas como falsas (es decir, que no provienen de los datos reales) (28). A continuación se ilustra el modelo básico de una GAN.

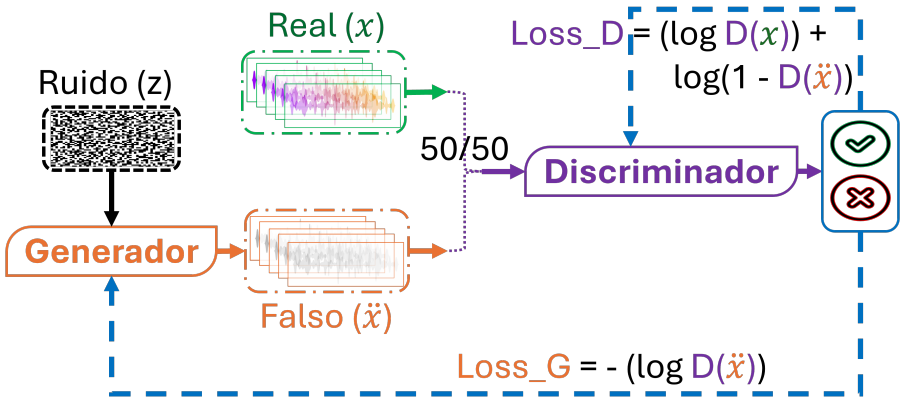


Figura 2.5: Modelo básico de una Red Generativa Adversaria (GAN)

Sin embargo, las GANs sufren problemas de inestabilidad en el entrenamiento y, a menudo, conducen a gradientes que desaparecen o explotan, dado que G no debe entrenarse demasiado sin actualizar D. Si esto sucede, G puede colapsar muchos valores de z al mismo valor de x, perdiendo la diversidad necesaria para modelar $p_{\text{data}}(x)$, a esto se le llama el "escenario Helvetica" (28).

2.4.3 Red Generativa Adversaria Condicional (cGAN)

Las Redes Generativas Adversarias Condicionales (cGAN) se basan en la estructura de las GAN tradicionales, pero incorporan un elemento clave: el condicionamiento. Este condicionamiento, representado por una variable "y", introduce información adicional que guía tanto al generador como al discriminador durante el proceso de entrenamiento. En una cGAN, tanto el generador como el discriminador reciben como entrada la variable de condicionamiento "y". El generador, además de "y", recibe un vector de ruido aleatorio "z", y su objetivo es generar datos sintéticos que sean indistinguibles de los datos reales, pero que además cumplan con la condición impuesta por "y". El discriminador, por su parte, recibe como entrada un dato, que puede ser real o generado por el generador, junto con la variable "y", y debe determinar si el dato es real o falso, teniendo en cuenta la condición impuesta por

"y". El entrenamiento de una cGAN es similar al de una GAN común, solo que la función de pérdida de la cGAN se modifica para incorporar el condicionamiento, de manera que tanto el generador como el discriminador sean penalizados si no cumplen con la condición impuesta por "y".

$$\min_G \max_D \mathop{\mathbb{E}}_{x\sim p_{\text{data}}(x)}[\log D(x|y)] + \mathop{\mathbb{E}}_{z\sim p_z(z)}[\log(1 - D(G(z|y)))] \tag{2.2}$$

En esencia, las cGAN aprenden una distribución de probabilidad condicional, lo que les permite generar datos sintéticos que no solo se asemejan a los datos reales, sino que también cumplen con las condiciones específicas impuestas por la variable y, que representa la información adicional que se incluye en el proceso de generación y discriminación. En lugar de simplemente generar datos aleatorios x desde el ruido z, se condiciona tanto la entrada del generador $G(z|y)$ como la entrada del discriminador $D(x|y)$, permitiendo que el modelo tenga más información sobre la clase o característica de los datos que debe generar o clasificar (44). A continuación se ilustra el modelo básico de una cGAN.

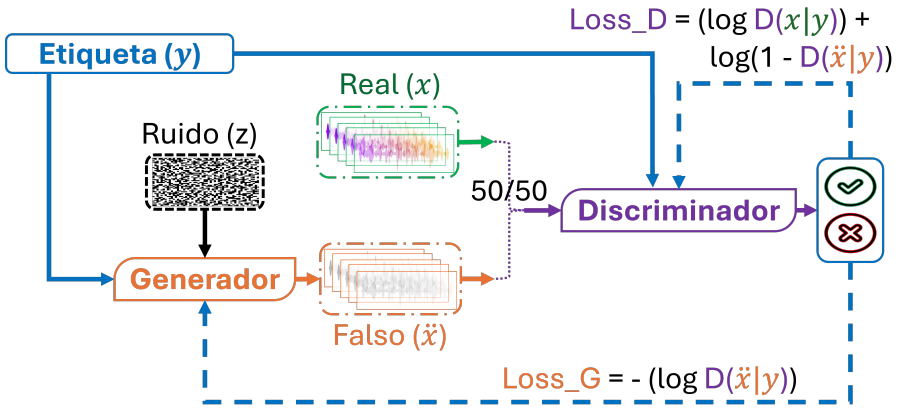


Figura 2.6: Modelo básico de una Red Generativa Adversaria Condicional (cGAN)

2.4.4 Wasserstein GAN (WGAN)

La WGAN aborda el problema de inestabilidad en el entrenamiento de la GAN utilizando la distancia de Wasserstein (Earth-Mover Distance) en lugar de la divergencia de Jensen-Shannon. Esto resulta en una función de valor con mejores propiedades teóricas para la convergencia (7).

Fórmula del valor de WGAN:

$$\min_{G} \max_{D \in \mathcal{D}} \mathop{\mathbb{E}}_{\boldsymbol{x} \sim \mathbb{P}_r} [D(\boldsymbol{x})] - \mathop{\mathbb{E}}_{\tilde{\boldsymbol{x}} \sim \mathbb{P}_g} [D(\tilde{\boldsymbol{x}})] \tag{2.3}$$

En este contexto, $\mathcal{D}$ representa el conjunto de funciones que cumplen la condición 1-Lipschitz [3] (una restricción en la rapidez con la que puede cambiar la salida del discriminador con el fin de evitar cambios abruptos) y $\mathbb{P}_g$ es la distribución de probabilidad del modelo, que se define de forma implícita mediante la ecuación $\tilde{x} = G(z)$, donde z sigue la distribución $p(z)$. Si se utiliza un discriminador óptimo (denominado "crítico" en este estudio, ya que su función no es clasificar), la minimización de la función de valor con respecto a los parámetros del generador implica la minimización de la distancia de Wasserstein entre la distribución real $\mathbb{P}_r$ y la distribución del modelo $\mathbb{P}_g$, representada como $W(\mathbb{P}_r, \mathbb{P}_g)$.

D es el conjunto de funciones 1-Lipschitz. Para garantizar la propiedad Lipschitz, se utiliza el recorte de pesos (*weight clipping*), pero esto puede llevar a comportamientos no deseados.

El recorte de pesos en WGAN puede causar problemas de optimización, como gradientes que explotan o desaparecen, y una capacidad de uso subóptima del discriminador, en este modelo el discriminador es llamado crítico (29).

[3] Las funciones que cumplen con esta condición, son las que gráficamente al trazar una línea recta entre dos puntos de la gráfica (una secante), su pendiente no sobrepasa un valor absoluto mayor a 1.

Solución al Weight Clipping: WGAN con Penalidad de gradiente (WGAN-GP)

Para superar estos problemas, se propone la WGAN-GP (Wasserstein GAN con Penalización de Gradiente). En lugar de recortar los pesos, se penaliza el valor de la norma del gradiente del crítico respecto a su entrada, asegurando así la propiedad Lipschitz de manera más efectiva.

Fórmula de la pérdida con penalización de gradiente:

$$L = \underbrace{\mathop{\mathbb{E}}_{\tilde{x}\sim\mathbb{P}_g}[D(\tilde{x})] - \mathop{\mathbb{E}}_{x\sim\mathbb{P}_r}[D(x)]}_{\text{Original critic loss}} + \lambda \underbrace{\mathop{\mathbb{E}}_{\hat{x}\sim\mathbb{P}_{\hat{x}}}\left[(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2\right]}_{\text{Gradient penalty}} \tag{2.4}$$

En la anterior ecuación, la pérdida del crítico original, $D(x)$ es la salida del discriminador para datos reales $x \sim P_r$, y $D(\tilde{x})$ es la salida para datos generados $\tilde{x} \sim P_g$. El objetivo es maximizar la diferencia entre las puntuaciones dadas a los datos reales y a los generados, de modo que el discriminador pueda distinguir correctamente entre ambos. Su complemento, la penalidad del gradiente, fue diseñado para forzar que el gradiente de D respecto a las muestras $\hat{x}$ tenga una norma cercana a 1, cumpliendo la condición 1-Lipschitz. Ahora, λ es un coeficiente de regularización que controla el peso de esta penalización (29).

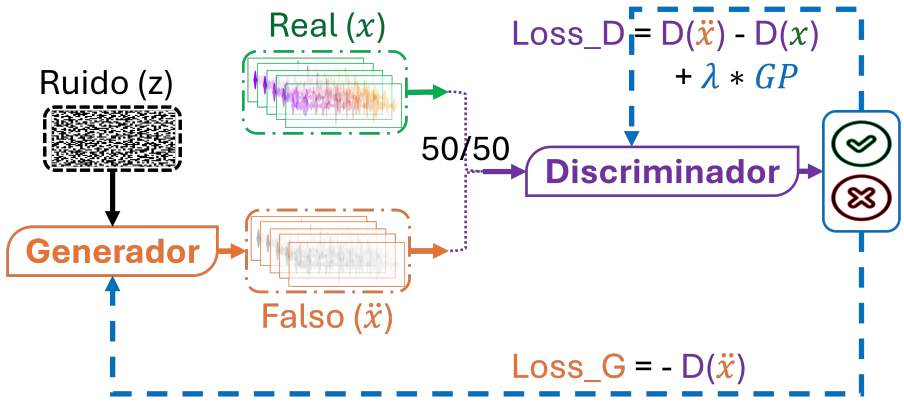


Figura 2.7: Modelo básico de una GAN Wasserstein con Penalidad de Gradiente

Implementación de WGAN-GP

El algoritmo de WGAN-GP se implementa de la siguiente manera:

Actualización del crítico: Para cada iteración del crítico, se calcula la pérdida del crítico con la penalización de gradiente, para luego actualizar los parámetros del crítico usando Adam.

Actualización del generador: Se calcula la pérdida del generador basada en la salida del crítico, para luego actualizar los parámetros del generador usando Adam.

La WGAN-GP mejora la estabilidad y el rendimiento de las GANs al reemplazar el *weight clipping* con una penalización de gradiente, logrando entrenamientos más estables y eficientes, y mejorando la calidad de las muestras generadas, por lo cual sera la Red Generativa Adversaria que utilizaremos en este proyecto, dada la cantidad de datos tan escasa y la necesidad de una pronta convergencia del modelo.

CAPÍTULO 3

TRABAJO RELACIONADO

La investigación sobre el deterioro cognitivo a partir del habla espontánea ha generado diversos conjuntos de datos en múltiples idiomas, como las de Calzà et al. (17) y Beltrami et al. (12), con corpus de habla en italiano, Chien et al. (20) utilizando audios en mandarín, Toth et al. (58) con húngaro y Konig et al. (33) en idioma francés, por mencionar algunos; estos estudios comparten un enfoque común, la extracción de características del habla, seguida de la clasificación mediante algoritmos de aprendizaje automático. Sería interesante una exploración de la existencia de marcadores biológicos acústicos universales, mediante el análisis combinado de estos conjuntos de datos multilingües. Sin embargo, este trabajo se limita al análisis del conjunto de datos ADReSSo (39) que se encuentra en inglés y al cual tenemos acceso, aunque con la perspectiva de generar una base sólida para futuras investigaciones multilingües.

Con el fin de obtener un panorama general de las investigaciones realizadas alrededor de la detección del deterioro cognitivo en habla espontánea, se presentan trabajos realizados para la competencia ADReSSo (39), donde se usan audios de pacientes describiendo una imagen, para computacionalmente, clasificarlos con y sin deterioro cognitivo, además de estimar su puntaje en una prueba cognitiva como lo es el *Mini Mental Status Examination*.

Basados en el *baseline* la competencia ADReSSo y la bibliografía producida en el año 2021, se puede inferir una estructura a seguir en cada proyecto realizado, que consta de tres partes, la primera, la extracción de características del dataset de grabaciones de voz, la segunda, el entrenamiento y predicción de Alzheimer's Dementia (AD) con diferentes clasificadores, y por último, la evaluación de resultados. En la presentación de ADReSSo, el *baseline* se generó extrayendo el conjunto de características eGeMaps, el cual es comúnmente utilizado en extracción de características para tareas relacionadas al reconocimiento de emociones en voz.

3.1 Investigaciones que usan el dataset ADReSSo 2021

A continuación, se presenta una Tabla comparativa de algunos trabajos relacionados que presentaron sus resultados donde involucran procesamiento de datos meramente acústico, sin transcripciones y análisis de las entrevistas. Cabe aclarar que, aunque algunos de los trabajos mencionados emplean características derivadas tanto de la transcripción como del análisis acústico para comprender el examen diagnóstico de afasia realizado a los pacientes; en la Tabla 3.1 solo se han considerado aquellos trabajos que presentan sus resultados utilizando características acústicas, esto con el objetivo de tener una idea del desempeño de los trabajos que tienen un enfoque similar al de esta tesis.

Autor	Preproces	Característica	Clasificación	Acc	F1
Luz (39)	Norm, Filter	eGeMaps	Decision Trees	64.79	-
Balagopalan (10)	-	MFCC + DNN	SVM	67.61	70.89
Chen (19)	-	Sets Features [1]	Vot May. Reg. Log.	67.61	-
Perez-Toro (49)	-	X-Vectores	SVM(RBF)	67.61	67.00
Pappagari (48)	-	SpeechBrain(Enc/Dec)	Reg Logistica	71.80	71.50
Pan (47)	-	wav2vec2.0	Tree Bagger(10)	74.65	74.52
Gauder (27)	Segmnt 5seg	wav2vec2.0	DNN	78.90	-

Tabla 3.1: Detección del deterioro cognitivo en la voz mediante características acústicas

Balagopalan et al. (10), extrajeron 168 características, obtenidas a partir de los estadísticos de promedio, varianza, asimetría, y curtosis de los 42 primeros Coeficientes Cepstrales de Frecuencia Mel (MFCC). Se usaron los clasificadores SVM, árboles de decisión, redes neuronales y regresión logística, dando como mejor resultado con SVM con una exactitud de 67.61 puntos y con árboles de decisión un F1-score de 70.89 puntos.

Chen et al. (19), utilizaron modelos de regresión logística para la clasificación de AD. Se utilizaron los 13 primeros Coeficientes de MFCC, el conjunto de GeMAPS y eGeMaps, el conjunto ComParE-2016 y el conjunto Paraling. Realizando votación mayoritaria de todos los clasificadores de la modalidad de audio obtuvo un accuracy del 67.61%.

Perez-Toro et al. (49), extrajeron X-Vectores (*Embeddings* obtenidas con una red TDNN), características de prosodia basadas en las tasas de habla y energía, y los contornos de frecuencia fundamental (F0) y *Embeddings* basados en el modelo emocional PAD (Placer, Incitación y Dominancia) donde se utilizó un clasificador SVM con kernel RBF, que obtuvo un accuracy de 67.61 y un F1 de 67 utilizando una fusión de X-Vectores y Dominancia del PAD.

Pappagari et al. (48), aplicaron la extracción de características con "SB Enc/Dec", un encoder-decoder de SpeechBrain, conformado por un codificador con arquitectura de redes neuronales convolucionales recurrentes profundas seguido de un LSTM bidireccional y una capa totalmente conectada para obtener la representación acústica, se usaron clasificadores de regresión logística y XGBoost. Con esta combinación obtuvo un accuracy de 71.80 y F1 de CN-72 y AD-71.

Pan et al. (47), trabajaron con un modelo multimodal que usa transcripciones y características acústicas, extraídas con *wav2vec2.0*. Se promediaron los datos correspondientes a la temporalidad de la salida para *wav2vec2.0*, dentro de los experimentos realizados, se observa que se obtuvo sobre el conjunto de test de ADReSSo un Accuracy de 74.65 y un F1 de CN-76.32 AD-72.73. Se usó el clasificador *Tree Bagger* con 10 estimadores.

Gauder et al. (27), segmentaron las grabaciones en fragmentos de 5 segundos con solapamiento de 1 segundo entre segmentos. Para la extracción de características, se usó eGeMaps, *Trill Embeddings* no semánticos usado en tareas de reconocimiento emocional y de identificación de hablantes), *Allosaurus* (Modelo de reconocimiento de fonemas universal) y *wav2vec2.0*. Se entrenó una red neuronal que incluye capas convolucionales de una dimensión, la cual toma los vectores de características y genera puntajes de cada segmento para finalmente promediar los puntajes de los segmentos de cada audio y dar su clasificación. El mejor modelo fue el que usa *wav2vec2.0* con un accuracy de 78.90, no se dio información del F1 Score.

Rohanian et al. (53), llevaron a cabo en un proyecto que trabaja con modelos BiLSTM y BERT con fusión multimodal, aquí se utilizaron características léxicas, acústicas, difluencias y pausas; para extraer las características acústicas se usó el software COVAREP con el cual se obtuvieron características relacionadas a la prosodia, calidad de voz y espectrales. Su mejor resultado se alcanzó con el modelo BiLSTM que usa capas highway donde se obtuvo un accuracy de 84% y un RMSE de 4.26.

Wang et al. (61) proponen una arquitectura multimodal denominada C-Attention Network, que consiste en tres redes independientes, C-Attention-Acoustic Network, C-Attention-FT Network y C-Attention-Embedding Network, que procesa características acústicas, lingüísticas o *embeddings* de audio y lenguaje, respectivamente. El modelo C-Attention-Unified con características lingüísticas y X-vectors fue el mejor, con un accuracy de 80.28%.

Zhu et al. (69) utilizan *Wav2vec* para extraer información semántica y no semántica del habla y BERT para el análisis semántico proponiendo modelos que mantengan la información no semántica crucial para mejorar la detección de demencia. Las características no semánticas del habla son identificadas mediante tokens en blanco producidos por *Wav2vec*, y convertidas en puntuaciones (puntos y comas) para su análisis con BERT. El modelo extendido WavBERT alcanzó una precisión del 83.1% en la tarea de clasificación. Ilias et al. (31) trabajaron en un modelo que combina capas de atención, métodos de adaptación de dominio mediante transporte óptimo que adapta las distribuciones de las características de diferentes dominios (audio y texto) y técnicas de fusión multimodal. En cuanto a características acústicas se utilizaron eGeMAPS, Espectrogramas Log-Mel y X-vectors. Este modelo alcanzó una precisión (Accuracy) del 85.35%, y una puntuación F1 del 85.27.

Mirheidari et al. (43) realizaron un estudio para mejorar la clasificación del deterioro cognitivo mediante el aumento de características extraídas de conversaciones entre pacientes y un agente virtual inteligente. Se tomaron características de análisis conversacional, acústicas, léxicas, coeficientes cepstrales de MEL, entre otras. Para la generación de características sintéticas se utilizó un autoencoder variacional y como clasificadores se hicieron pruebas con Regresión Logística y Redes Neuronales Profundas. En este estudio demuestra que el aumento de datos utilizando un VAE mejora el rendimiento de los clasificadores tanto en la regresión logística como en la red neuronal profunda.

Agbavor et al. (4) presentan un sistema de inteligencia artificial que usa grabaciones de voz directamente para la detección y evaluación de la enfermedad de Alzheimer. El modelo usa *data2vec*, un modelo pre-entrenado que funciona con voz, visión y texto. El sistema fue evaluado con una red neuronal, se obtuvo un F1 Score de 72.8 y una Accuracy de 72.7% en una validación cruzada de 10 pliegues.

Altinok et al. (5) presenta un modelo multimodal que fusiona información textual (transcripciones) y acústica (espectrogramas) mediante cross-attention para la detección de demencia. Se usan brevemente ViT (Vision Transformer) y RoBERTa como extractores de características, junto a un clasificador fully connected, logrando un 90.01% de Accuracy en el dataset ADReSSo.

Olachea-Hernandez et al. (45) proponen dos caracterizaciones basadas en el análisis del lenguaje para detectar la enfermedad de Alzheimer (EA) a través de alteraciones en el habla. La primera caracterización se centra en aspectos superficiales del lenguaje, como la estructura gramatical y la fluidez oral, mientras que la segunda utiliza un modelo preentrenado de lenguaje (BERT) para identificar anomalías en el uso del lenguaje. Los resultados muestran un rendimiento comparable al estado del arte, con un 87% de Accuracy al combinar ambas caracterizaciones.

Balamurali B.T et al. (11) evalúan el rendimiento de tres modelos de lenguaje, ChatGPT-3.5, ChatGPT-4 y Bard en la detección de AD a partir de transcripciones. Emplean un enfoque con dos tipos de preguntas, una directa y otra que forza al razonamiento del modelo. Bard destaca en la identificación de AD con 71 de F1 score, mientras que GPT-4 se muestra más efectivo en identificar CN con un 62 de F1 score.

3.2 Trabajos Relacionados en Aumento de Datos Acústicos

Si bien el aumento de datos es una técnica ampliamente utilizada en el procesamiento del lenguaje natural, especialmente para la transcripción y análisis del lenguaje, su aplicación al análisis de audio para la detección del deterioro cognitivo es aún limitada. La mayoría de las investigaciones para la tarea de interés, se centran en el aumento de datos a nivel de transcripción, como se evidencia en los trabajos de Liu et al. (38), Lin et al. (36) y Balagopalan et al. (9).

En el ámbito del audio, los trabajos existentes abordan tareas como el reconocimiento de emociones, generación de audios clínicos y la clonación de voz. Por ejemplo, en (57) utilizan una red generativa adversaria condicional para generar espectrogramas de Mel para el reconocimiento de emociones. En (56) se trabajó en aumento de audios clínicos (sonidos comunes en cirugía), utilizando una Style-GAN, que surgió de la generación de rostros, en este caso se usa para generar espectrogramas. En Li et al. (34), se utilizó una cGAN con módulos de atención, para hacer cambios de voces entre un hablante y otro, además de ser capaz de modificar la voz para simular el canto y por último se encontró un proyecto de código abierto para clonación de voz que usa módulos de tipo *transformer*, para abstraer las características de la voz, pudiendo "clonar la voz" a partir de ejemplos de audio y genera nuevos audios con solo darle otra entrada de texto (14).

Esmaeilpour et al. (24) trabajaron el aumento de datos con la (WCCGAN) Weighted Cycle-Consistent Generative Adversarial Network por sus siglas en inglés, aplicada a espectrogramas de sonidos ambientales para la clasificación de estos, utilizando características extraídas mediante SURF (Speeded-Up Robust Features) que es un algoritmo de visión por computadora que se utiliza para detectar y describir puntos de interés en imágenes. y evaluándolas con Random Forest.

Madhu Kumaraswamy (40) utilizaron una WaveGAN para el aumento de datos aplicada a sonidos ambientales para su clasificación, utilizando espectrogramas Mel como características y evaluando el rendimiento con una red neuronal convolucional (CNN).

Mertes et al. (42) emplearon una WaveGAN junto con un algoritmo evolutivo para el aumento de datos de audio para la clasificación de paisajes sonoros, utilizando características espectrales y evaluando el rendimiento con una Máquina de Soporte Vectorial (SVM)

Qian et al. (50) investigaron el uso de GANs y (cGANs) Conditional GANs por sus siglas en inglés, para el aumento de datos aplicado a espectrogramas de voz para el reconocimiento de voz robusto al ruido, utilizando características espectrales y evaluando el rendimiento con una (VDCNN) Very Deep Convolutional Networks por sus siglas en inglés.

Ramesh et al. (51) desarrollaron CoughGAN, una variante de WaveGAN, para generar toses sintéticas aplicadas a datos de audio de toses para la clasificación de enfermedades respiratorias, utilizando características espectrales de corta duración y evaluando el rendimiento con SVM y Random Forest.

Wang et al. (60) propusieron xGAN, un modelo basado en GAN con LSTM jerárquico, para el aumento de datos de texto aplicado a diálogos del Internet de las Cosas (IoT) para mejorar el reconocimiento de voz, utilizando un modelo de lenguaje N-gram para la evaluación.

Wang et al. (62) utilizaron cGANs y (VAEs) Variational Autoencoders por sus siglas en inglés, para generar embeddings de hablante aplicados a datos de embeddings para la verificación de hablante basada en PLDA (Probabilistic Linear Discriminant Analysis), utilizando i-vectors, x-vectors y r-vectors como embeddings y evaluando el rendimiento con PLDA.

Zhang et al. (67) presentaron Snore-GANs, GANs semi-supervisadas condicionales (scGANs), para el aumento de datos aplicado a sonidos de ronquidos para la clasificación automática de estos, utilizando descriptores de bajo nivel, (BoAW) Bag of Audio Words por sus siglas en inglés y evaluando el rendimiento con SVM y (GRU-RNN) Gated Recurrent Unit con Recurrent Neural Network por sus siglas en inglés.

Zhao et al. (68) utilizaron una (AC-GAN) Auxiliary Classifier GAN por sus siglas en inglés, para generar vectores de características espectrales CQT-TFR de muestras de voz, aplicadas para contrarrestar ataques de suplantación de identidad por repetición en la verificación del hablante, evaluando el rendimiento con sistemas basados en GMM (Gaussian Mixture Model), DNN y (LCNN) Lookup-based Convolutional Neural Network por sus siglas en inglés.

Yella y Rajan (66) exploraron el aumento de datos utilizando GAN para el diagnóstico de COVID-19 basado en sonido. Emplearon WaveGAN, una variante de GAN, para sintetizar datos de audio de tos y los evaluaron con una CNN para clasificar estas toses como positivas o negativas.

Es importante destacar que, en su mayoría, estas técnicas se basan en la generación de espectrogramas (representaciones visuales del audio) y su posterior transformación a audio. A diferencia de estos enfoques, este trabajo se centra en el aumento de datos a nivel de representaciones vectoriales del audio, sin generar imágenes intermedias.

3.3 Discusión

Diversas investigaciones, como las presentadas en la competencia ADReSSo, han abordado la extracción y clasificación de características acústicas mediante una variedad de enfoques metodológicos. Entre los métodos utilizados, se destacan las

técnicas clásicas, como los coeficientes cepstrales de Mel y también se aprecia la implementación de modelos más avanzados, como *wav2vec2.0*, que permite la captura de representaciones profundas del audio.

El análisis comparativo de estos trabajos evidencia que la incorporación de modelos basados en redes neuronales profundas, junto con la extracción de características, incrementa el accuracy en la clasificación del deterioro cognitivo. No obstante, es importante señalar que hasta la fecha no se han desarrollado investigaciones que aborden de manera directa la mejora en la detección de deterioro cognitivo leve *usando aumento de datos*, un reto considerable debido a la limitada cantidad de datos disponibles.

Considerando la escasez de datos y la complejidad inherente a la obtención de muestras de voz adecuadas, el presente trabajo se centra en la exploración de técnicas de aumento de datos utilizando meramente las características acústicas de los mismos. Puesto que permite el desarrollo de herramientas de diagnóstico que no dependen de un contexto específico, como lo es el examen de Boston Aphasia.

Una distinción clave de este trabajo es la aplicación de técnicas de aumento de datos, incluyendo GANs, directamente sobre representaciones vectoriales del audio, sin recurrir a la generación de espectrogramas. A nuestro entender, no se han encontrado en la literatura trabajos que utilicen GANs de esta manera para la detección del deterioro cognitivo leve.

CAPÍTULO 4

EVALUACIÓN DE REPRESENTACIONES ACÚSTICAS USADAS EN DETECCIÓN DE MCI

Este capítulo se centra en la evaluación de diversas representaciones acústicas y su eficacia para la detección del Deterioro Cognitivo Leve (MCI, por sus siglas en Inglés). Comenzamos analizando el conjunto de datos ADReSSo 2021, poniendo especial énfasis en el problema del desbalance entre clases que este presenta. Exploraremos la extracción de características acústicas, tanto clásicas como las basadas en *embeddings*. En el caso de las clásicas, utilizaremos MFCCs, Chroma y eGeMAPS, evaluando su rendimiento con algoritmos de clasificación como kNN, SVM, Random Forest y ADA Boost. Analizaremos además el impacto de la segmentación de los audios en la eficiencia de la clasificación, buscando la granularidad temporal óptima para la representación de las señales de habla. Finalmente, se compararán los resultados de las representaciones clásicas con los obtenidos usando *embeddings* generados a partir de los modelos *wav2vec* y *wav2vec 2.0*, con el fin de determinar qué enfoque ofrece un mejor rendimiento para la tarea de detección de MCI. La meta es establecer un *baseline* para la posterior aplicación de técnicas de aumento de datos, que serán exploradas en el siguiente capítulo.

4.1 Dataset ADReSSo 2021

Para este proyecto se hará uso del dataset del *Dementia bank* propuesto para el reto ADReSSo promovido en Interspech 2021 (39), en esta competencia se buscaba analizar el habla espontánea mediante audios grabados de entrevistas, a partir de la descripción por parte del paciente de la imagen del “robo de galletas" del examen *Boston Diagnostic Aphasia*. Este dataset también incluye una evaluación del paciente con la puntuación del Mini Mental Status Examination (MMSE) hecha por un profesional de la salud. La finalidad de esta competencia es trabajar en un método computacional que permita clasificar entre pacientes con deterioro cognitivo y pacientes sanos, estimar el puntaje MMSE, así como predecir el riesgo de desarrollar deterioro a través del tiempo para cada paciente. Para el conjunto de entrenamiento (*Train)*, se cuenta con 87 grabaciones en audio de personas con Alzheimer y 79 audios de personas de control. Las personas con Alzheimer presentaron un MMSE entre 1 y 28; a este subconjunto le llamaremos AD por sus siglas en Inglés Alzheimer's Disease. Las personas de control, al cual identificaremos como CN por sus siglas en Inglés Cognitive Normal, poseen puntajes de MMSE entre 26 y 30 puntos. Para el conjunto de *Test* se tienen 35 grabaciones de la clase AD y 36 de CN.

Dado que la enfermedad de Alzheimer es una condición actualmente incurable, como se mencionó anteriormente, identificar las etapas preclínicas de esta enfermedad es vital para maximizar la duración y calidad de vida autónoma de las personas que la padecen. Es por ello importante destacar que se realizaron experimentos con un subconjunto del conjunto original. Nuestro interés es distinguir entre los individuos con deterioro cognitivo leve e individuos de control. Para distinguir el subconjunto con puntajes inferiores a 24 puntos de MMSE, indicando un deterioro cognitivo avanzado, lo llamaremos CI (*Cognitive Impairment*). De esta forma el conjunto de pacientes con la enfermedad de Alzheimer incluye al conjunto CI.

El conjunto de individuos con MCI está conformado por aquellos individuos

del conjunto original pertenecientes a la clase AD y cuyos puntajes de MMSE se encuentran arriba de 24 puntos; dado que según la literatura el MCI se divide en deterioro cognitivo leve temprano (*Early MCI*) y deterioro cognitivo leve tardío (*Late MCI*); estas clases se encuentran comprendidas en un rango de puntaje MMSE de 27.69 ± 1.45 para *Early MCI* y 26.55 ± 1.60 para *Late MCI*; entonces, se podría decir que el puntaje de MMSE para MCI se encuentra en un rango de 29.14 y 24.95 puntos de MMSE (37); que comparando estos datos con los de las Figuras 4.1 y 4.2, es más o menos el rango observable para MCI, entendiendo esta clase como el punto donde empieza a aparecer el deterioro cognitivo, la cual es una zona donde se difumina el punto de corte para un puntaje MMSE entre una persona cognitivamente normal y una persona con deterioro cognitivo.

El otro subconjunto contiene a los individuos de control del conjunto original, cuyos puntajes de MMSE también están arriba de los 24 puntos. Es importante notar, que la frontera que separa estos dos subconjuntos es difusa, por lo cual, esta delimitación agrega complejidad al análisis debido a la ambigüedad diagnóstica en este rango, donde las diferencias entre individuos sanos y aquellos con inicios de Alzheimer pueden ser mínimas, como se evidencia en casos donde sujetos sanos obtienen 24 puntos en el MMSE y personas con Alzheimer alcanzan los 30 puntos. Para dar a entender mejor la naturaleza de los datos se muestran las Figuras 4.1 y 4.2, en la cuales se observa visualmente la distribución y el solapamiento entre la clase MCI y CN, tanto en el conjunto de Train como en el conjunto de Test.

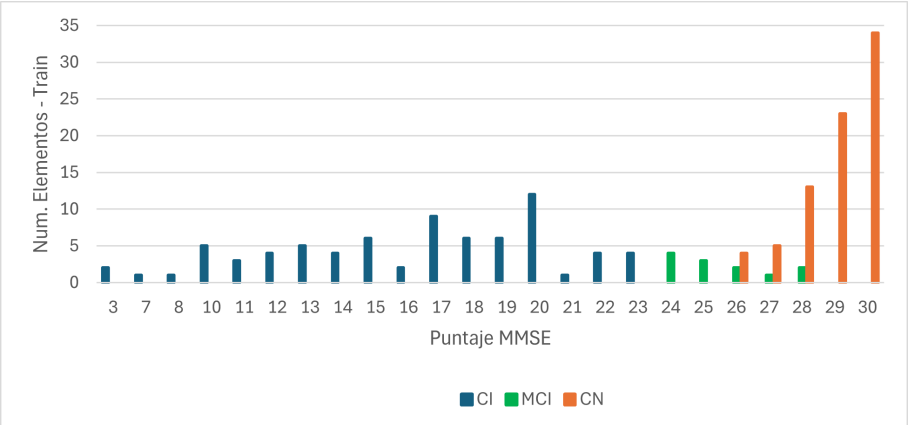


Figura 4.1: Distribución de puntajes MMSE en Train de ADReSSo

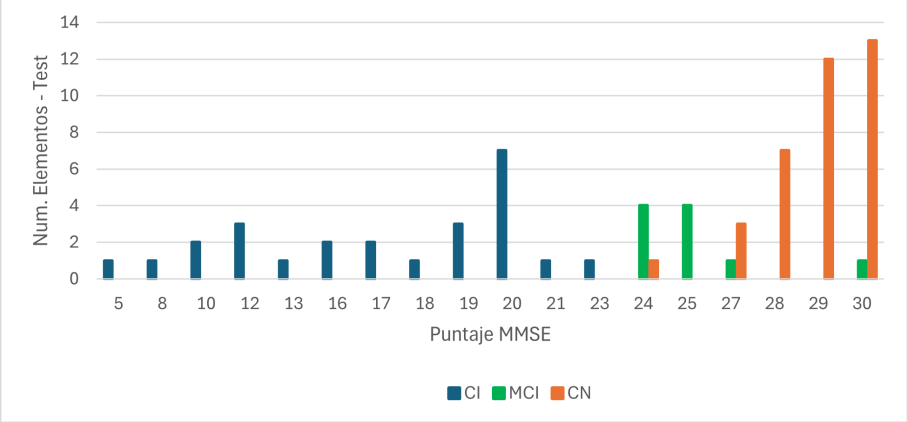


Figura 4.2: Distribución de puntajes MMSE en Test de ADReSSo

Además, otra complejidad en este caso, es el desbalance del dataset dada la poca disponibilidad de elementos con MCI. Para diferenciar este subconjunto de datos, el cual comprende los pacientes que tienen puntuaciones de MMSE entre 24 y 30, lo llamaremos *SubADReSSo*. Las Tablas 4.1 y 4.2 presentan las distribuciones de datos de estos dos conjuntos. Las Tablas resumen la cantidad de datos de cada dataset (tanto en porcentaje como en número total de instancias).

Conjunto	AD %	CN %	AD	CN	Total
Train	52.41	47.59	87	79	166
Test	49.30	50.70	35	36	71

Tabla 4.1: Tabla del Dataset ADReSSo (MMSE de 1 a 30)

Conjunto	MCI %	CN %	MCI	CN	Total
Train	13.19	86.81	12	79	91
Test	21.74	78.26	10	36	46

Tabla 4.2: Tabla del Dataset SubADReSSo (MMSE de 24 a 30)

Como se observa en la Tabla 4.2 existe un fuerte desbalance. Debido a este desbalance entre clases, los resultados se presentarán usando el *F1 Score* por clase.

Identificación de los hablantes

Los audios de este dataset incluyen también las intervenciones del entrevistador; dado que lo que se desea es obtener un diagnóstico para el paciente, lo ideal es trabajar únicamente con la voz del paciente. Para ello, se aplicó un proceso de diarización (identificación de los hablantes), con el fin de eliminar las intervenciones del entrevistador, las cuales podrían llegar a alterar los resultados alcanzados. Es por ello que entre los elementos que entrega la competencia ADReSSo 2021 (39) se incluyen las marcas de tiempo para la diarización, tanto para el conjunto de Train como el de Test, con ellas se realizó un preprocesamiento de los datos, con el fin de eliminar los segmentos donde habla el profesional.

4.2 Experimentos con Representaciones Clásicas

Esta sección evalúa representaciones acústicas clásicas para detectar deterioro cognitivo leve en habla espontánea. Se analizan MFCC, sus deltas, Chroma y eGeMAPS, ampliamente utilizadas en la literatura y seleccionadas por su potencial para capturar información relevante relacionada con la prosodia, el contenido espectral y los rasgos emocionales del habla, aspectos que se ven afectados por el deterioro cognitivo. Estas características serán evaluadas con diversos algoritmos de clasificación para identificar la combinación óptima, entre conjunto de características y clasificador. También se estudia el impacto de la segmentación del audio en la extracción de características, buscando mitigar la pérdida de información temporal y aumentar los datos de entrenamiento. Estos resultados servirán de base para la posterior exploración de técnicas de aumento de datos y la generación de datos sintéticos con GANs.

4.2.1 Extracción de Características

A partir de los trabajos de participantes en el reto se identificaron aquellas técnicas más prometedoras. Entre ellas, están: (i) el conjunto de características eGeMAPS propuesto para extraer rasgos emocionales; (ii) los primeros 20 coeficientes cepstrales en las frecuencias de Mel, que se orientan a las características auditivas dentro del rango del habla humana; (iii) los deltas de los primeros 20 coeficientes MFCC, que buscan observar la dinámica temporal de estos coeficientes y (iv) los coeficientes Chroma orientados a describir la curva melódica del audio. Los últimos tres conjuntos de características se trabajaron calculando los estadísticos funcionales[1] de las

[1]Se calcularon el promedio, desviación estándar, mediana, curtosis y asimetría de cada coeficiente obtenido para la serie temporal de datos que representa cada audio.

series obtenidas al calcular los coeficientes por segmentos temporales[2], puesto que el conjunto de eGeMaps, arroja las 88 características procesadas e independientes de la variable temporal del audio. Para una mejor comprensión se ilustran los pasos de obtención de las características clásicas en la siguiente Figura.

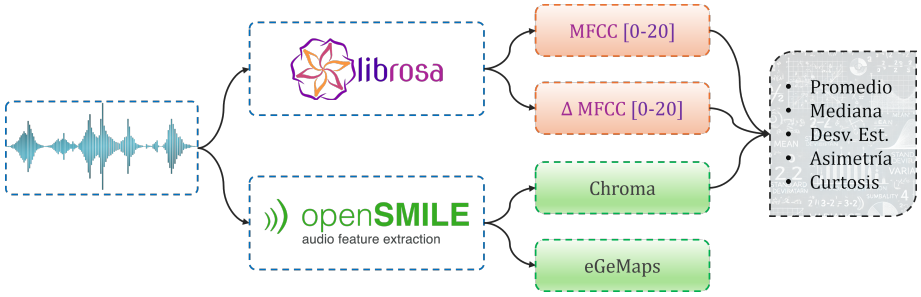


Figura 4.3: Modelo de extracción de características clásicas.

Para una mayor claridad de los datos extraídos, la Tabla 4.3 muestra la cantidad de características obtenidas con cada una de los métodos clásicos de extracción.

Método	**Cantidad**
MFCC [0-20]	100
Δ MFCC [0-20]	100
Chroma	60
eGeMaps	88

Tabla 4.3: Tabla con la cantidad de características extraída por métodos.

Una de las razones por las que se escogieron los conjuntos de eGeMAPS y MFCC, fue debido a su amplio uso en la bibliografía, y se incluyó en este análisis las características Chroma, puesto que estos coeficientes caracterizan los semitonos con deformación de octava, reconociendo acordes y tonalidades, con esto se podrían encontrar posibles alteraciones del timbre de voz en los pacientes enfermos. En la

[2]Para el cálculo del conjunto de eGeMaps y Chroma, se utilizó la librería de **OpenSmile** y para los coeficientes de MFCC y sus Deltas se usó la librería **Librosa** para Python.

siguiente sección se realiza un análisis más profundo acerca de las características de la voz que se afectan por el deterioro cognitivo y representaciones matemáticas asociadas.

Características acústicas de la voz afectadas por deterioro cognitivo

La Enfermedad de Alzheimer (EA) presenta déficits lingüísticos que se manifiestan desde sus primeras etapas, lo que los convierte en una herramienta prometedora para su diagnóstico temprano. Durante la prolongada fase asintomática y preclínica de esta enfermedad, diversos parámetros del habla y la voz, estrechamente relacionados con las funciones cognitivas, pueden anticipar la aparición de síntomas clínicos característicos de la demencia. En este contexto, ha surgido un creciente interés en los deterioros del lenguaje que se presentan desde etapas iniciales, incluso en fases prodrómicas de la EA, debido a la correlación comprobada entre medidas acústicas y características patológicas en la voz de los pacientes (41).

Los estudios en este campo han identificado una variedad de características del habla afectadas por la EA, lo que permite agrupar estos cambios en categorías específicas que incluyen parámetros acústicos, suprasegmentales o prosódicos, articulatorios y perceptuales. Cada una de estas áreas aporta información relevante que podría facilitar la identificación de la EA en etapas tempranas y su diferenciación de otros tipos de deterioro cognitivo.

Se observan alteraciones en la duración del habla y la fonación, una disminución en la relación armónico-ruido (HNR) y variaciones en la intensidad de la voz, que puede discriminar entre distintos grupos clínicos. Además, el análisis del tono muestra patrones específicos de disminución o aumento en la frecuencia fundamental (F0) según el género, y la presencia de una "entonación plana", característica de personas con EA. Cambios en los formantes, como las distorsiones en F2 y F3, y en las características espectrales, como el centroide espectral y los coeficientes cepstrales

de frecuencia de Mel, son también indicativos de la EA (41).

En el ámbito prosódico, el ritmo o la prosodia del habla presentan una disminución, caracterizada por variaciones en la velocidad, las pausas y los acentos en personas con EA. En los aspectos articulatorios, se observa un incremento en los segmentos sin voz durante la articulación de palabras, reflejo de un déficit en la coordinación vocal (41).

Por último, en las características perceptuales, la EA se asocia con la presbifonía, alteración de la voz común en el envejecimiento, junto con una mayor incidencia de vacilaciones y tartamudeos, y una variabilidad en la duración de las sílabas. Indicadores como el *shimmer* reflejan mayores fluctuaciones en la amplitud de la voz, mientras que los cambios en el espectro, como la energía espectral, la variabilidad, la asimetría y la curtosis, ofrecen parámetros adicionales para el análisis de esta patología (41).

MFCC (Coeficientes Cepstrales de Frecuencias Mel) y sus Deltas

Bisogni et al. (15), habla de que los MFCC utilizan un conjunto de filtros que simulan la manera en que el oído humano procesa el sonido. Esto posibilita la obtención del espectro de potencia, el cual, a su vez, sirve para representar las propiedades del tracto vocal y las sutilezas de la voz, útiles para evidenciar distorsiones en las características espectrales y la presencia de una "entonación plana".

Proceso: Se divide el audio en ventanas de 25ms con un solapamiento de 10ms. Para cada ventana, se calcula la Transformada Rápida de Fourier (FFT). Luego, se aplica la escala Mel al espectro de frecuencias para imitar la percepción auditiva humana. Finalmente, se aplica la Transformada Coseno Discreta (DCT) para obtener los MFCC.

Representación Final: Se extraen 20 coeficientes MFCC por ventana. Para

obtener una representación fija para cada audio, se calcula el promedio, la desviación estándar, la mediana, la curtosis y la asimetría de los 20 coeficientes a lo largo del tiempo. Esto resulta en un vector de 100 características (20 MFCC * 5 estadísticos).

Delta MFCC

Los Deltas de MFCC capturan la variación temporal de los MFCC, lo que proporciona información sobre la dinámica del habla, por lo cual podrían capturar variaciones espectrales resultado indirecto de cambios en F0, otra característica importante para la detección del deterioro cognitivo como veíamos anteriormente. Además, en (22) se calculan las derivadas de primer y segundo orden de los MFCC, para capturar información temporal y dinámica de la señal. Lo cual podría relacionarse con la prosodia, la variabilidad de la voz y el control vocal.

Proceso: Se calculan las derivadas de primer orden de los MFCC a lo largo del tiempo.

Representación Final: Se calculan 5 estadísticos promedio, desviación estándar, mediana, curtosis y asimetría para los Deltas de MFCC, resultando en otro vector de 100 características.

Chroma

Las características Chroma representan la distribución de la energía en las 12 notas de la escala musical (semitonos). Esto proporciona una representación del "ADN melódico" del audio y puede ser útil para detectar cambios sutiles en la prosodia, como la entonación o la monotonía, que pueden estar presentes en el deterioro cognitivo. En (15) se utilizan características Chroma para capturar la composición armónica del sonido, lo que se relaciona con la percepción del tono. Se calcula la media, mediana y desviación estándar de las características Chroma para obtener

una descripción estadística del marco armónico.

En otro estudio (21), se utilizan 6 bandas de Chroma en lugar de las 12 típicas en música. Esto debido a que agregan información espectral en bandas de tono, capturan variaciones en timbre (a través de la distribución de energía en bandas de tono), melodía correlacionada, patrones rítmicos, acento y entonación, que son prominentes entre los dialectos y para nuestro caso estas características también ayudan a la identificación de deterioro cognitivo.

Proceso: Se divide el audio en ventanas de 64ms con un solapamiento de 10ms. Para cada ventana se calcula la FFT y se mapea a la escala de semitonos. Se suma la energía en cada semitono para obtener un vector de 12 valores.

Representación Final: Se calculan los mismos 5 estadísticos (promedio, desviación estándar, mediana, curtosis y asimetría) para cada una de las 12 características Chroma a lo largo del tiempo, resultando en un vector de 60 características (12 Chroma * 5 estadísticos).

eGeMaps

El conjunto extendido de parámetros acústicos minimalistas de Geneva (eGeMAPS) es un conjunto extendido de 88 parámetros acústicos diseñado para capturar una amplia gama de características del habla, incluyendo la prosodia, la calidad de la voz y el contenido espectral. En el contexto del deterioro cognitivo, se espera que eGeMAPS capture una variedad de cambios sutiles en el habla, proporcionando una representación más completa del deterioro, puesto que muchas de las características que se mencionan en la sección anterior, están incluidas en este conjunto de características. Dado que este conjunto es muy amplio, para una descripción detallada, se recomienda consultar el apéndice, Extracción Clásica de Características Acústicas/eGeMaps (25).

Proceso: Se utiliza la herramienta openSMILE para extraer las 88 características de eGeMAPS para cada audio.

Representación Final: Se obtiene un vector de 88 características por audio, sin necesidad de calcular estadísticos adicionales. Este conjunto ya incluye estadísticas de bajo nivel como promedio, desviación estándar, percentiles, entre otros, calculados sobre ventanas o segmentos de la señal por openSMILE.

4.2.2 Resultados de Referencia

En esta primera parte de la evaluación de representaciones acústicas, se llevó a cabo una comparación de las caracterizaciones anteriores mediante el uso de diversos algoritmos de clasificación, entre estos se encuentran: vecinos más cercanos (kNN) con k igual a 5, 10 y 30, máquinas de vectores de soporte (SVM) con kernel Lineal y Polinomial de grado 20 y 30, con el parámetro `class_weight='balanced'`, Random Forest con el parámetro `n_estimator= 100` y `random_state=0` para que sus resultados sean deterministas, al igual que el ADA Boost, pero este último con un `n_estimator= 50`.

En las siguientes dos Figuras se presenta el desempeño de estos clasificadores empleando las técnicas de extracción de características previamente descritas.

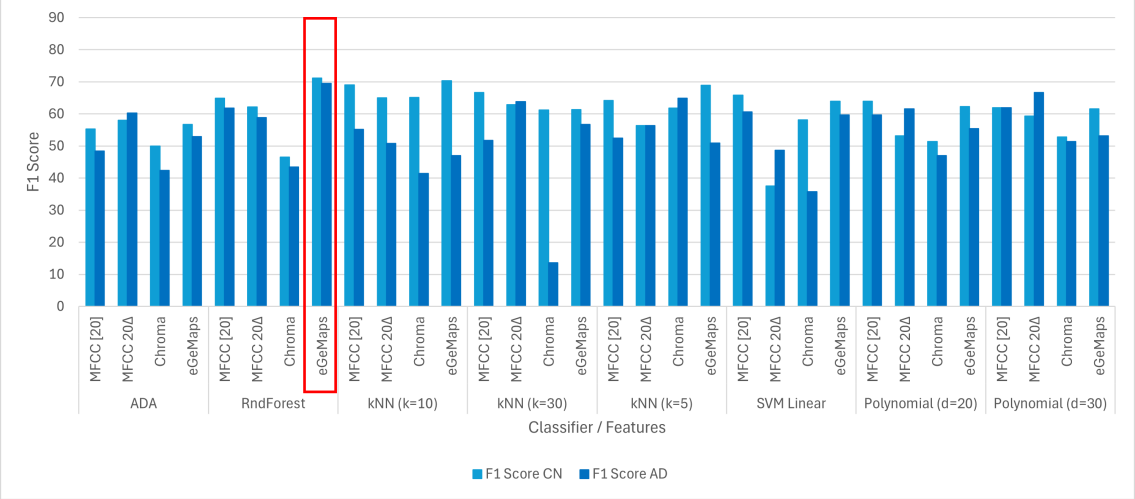


Figura 4.4: Resultados en F1 Score por clase, evaluando el Test ADReSSo. Todos los clasificadores fueron entrenados con el Train ADReSSo. Se resalta en rojo la combinación que obtuvo mejor desempeño, en este caso, el conjunto de características eGeMaps en la sección del clasificador Random Forest.

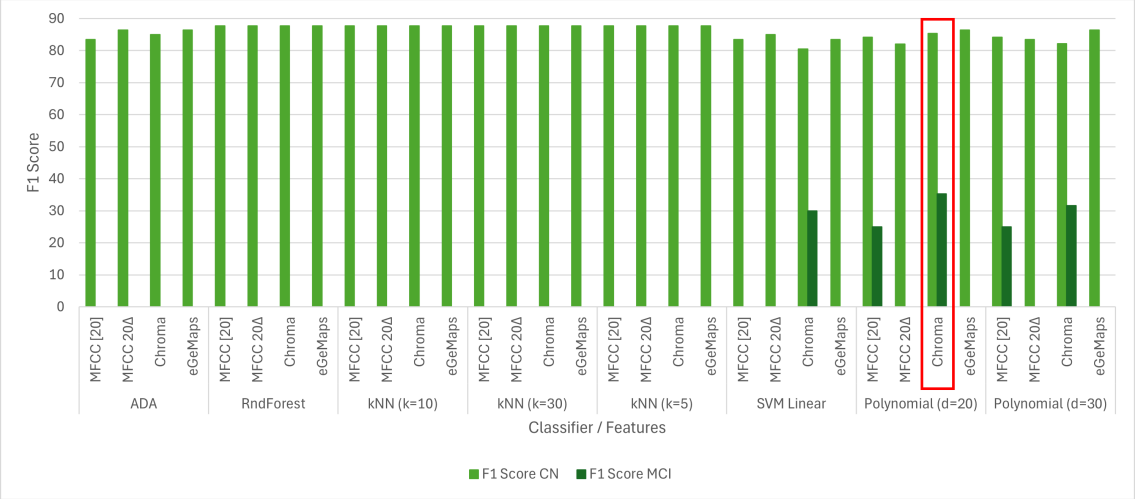


Figura 4.5: Resultados en F1 Score por clase, evaluando el Test subADReSSo. Los diferentes clasificadores fueron entrenados con el Train subADReSSo. Se resalta en rojo la combinación que obtuvo mejor desempeño, en este caso, el conjunto de características Chroma en la sección del clasificador SVM con kernel Polinomial de grado 20.

En la Figura 4.4 se puede observar como el mejor puntaje para ambas clases

es la combinación del conjunto eGeMaps con el clasificador *Random Forest*, pero, como se puede observar en la Figura 4.5, el experimento que se realiza utilizando el conjunto de datos subADReSSo, que es el de nuestro interés, se obtiene un puntaje nulo; pero, analizando ambas gráficas, se puede observar que al usar las características de 20 coeficientes de MFCC en la Figura 4.5, donde se está clasificando nuestro conjunto de interés MCI, con los clasificadores polinomiales; esta vez, si observamos ambos experimentos, representados en las Figuras 4.4 y 4.5, se obtienen puntajes más balanceados usando 20 coeficientes de MFCC y clasificador polinomial de grado 30. Dados los resultados de este experimento, los siguientes experimentos con representaciones clásicas utilizarán los parámetros elegidos. Para una descripción detallada de los experimentos realizados, se recomienda consultar el apéndice, la Tabla A.1 correspondiente a la Figura 4.4 y A.2 correspondiente a la Figura 4.5.

Sin embargo, el análisis de señales de audio presenta el desafío de la variabilidad en la duración. La extracción de características clásicas, basada en estadísticos globales, puede resultar en una pérdida de información al generalizar drásticamente la señal. Para abordar esta problemática, se propone una estrategia de segmentación arbitraria de los audios en fragmentos de 3, 5 y 10 segundos, con el fin de observar su comportamiento. Esta segmentación tiene dos objetivos principales. En primer lugar, aumentar la cantidad de datos para el entrenamiento de la GAN que se evaluara como uno de los métodos de aumento en esta tesis. En segundo lugar, trabajar con elementos del mismo tamaño, conservando información temporal detallada que podría perderse con una representación global.

La predicción se realizará por segmentos, sin embargo, la clasificación final de cada instancia de audio se basará en un conteo de los segmentos que han sido clasificados como positivos para Alzheimer y los que no, donde se irán probando umbrales para tomar el que mejor comportamiento obtenga.

Para una mejor comprensión del experimento se propone un ejemplo en la Figura 4.6, donde, primero se segmenta el audio cada X cantidad de segundos y se representa con la técnica de extracción clásica elegida, para posteriormente pasar cada uno de esos segmentos por el clasificador, el cual dará su aprobación o desaprobación, es decir, si el segmento pertenece a la categoría de AD o no; para posteriormente pasar estos segmentos en grupo por cada audio, por una función de agregación, que para efectos del ejemplo diría que si la agregación es menor o igual al 60% el audio pertenece a la clase de interés, mayor a esto será de la clase contraria; esto se hará con cada audio de la clase de Test del dataset seleccionado, para dar la clasificación final de cada audio en AD o CN, según el porcentaje de agregación evaluado.

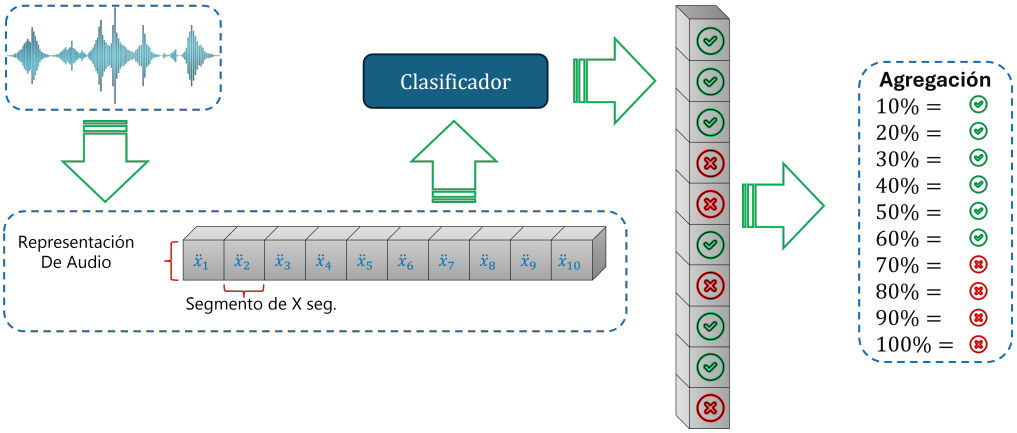
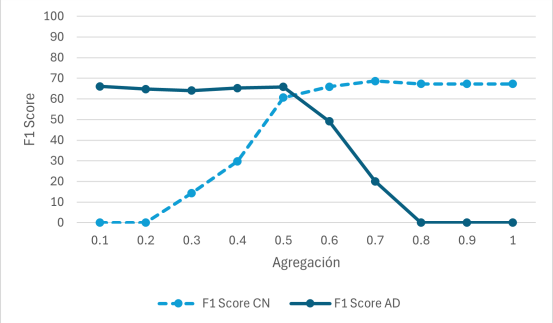
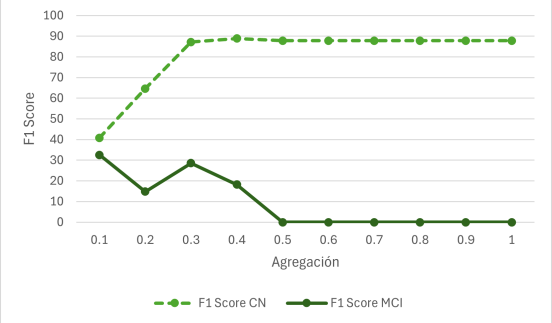


Figura 4.6: Ejemplo de clasificación por agregación de audios segmentados en segundos.

Se decidio segmentar los audios en fragmentos de 3, 5 y 10 segundos y dado el análisis de la sección anterior, se extraen los 20 primeros coeficientes de MFCC [0-20]. Posteriormente, se emplea un clasificador SVM con kernel polinomial de grado 30 para clasificar cada segmento. De acuerdo con la cantidad de segmentos que sean clasificados como positivos para Alzheimer, se emitirá una clasificación final para cada archivo de audio, evaluando porcentajes de agregación de dichos segmentos de 10 a 100%.

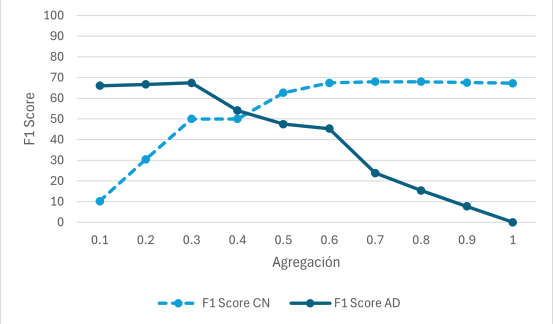


(a) Evaluación del Test ADReSSo utilizando como entrenamiento el Train ADReSSo.

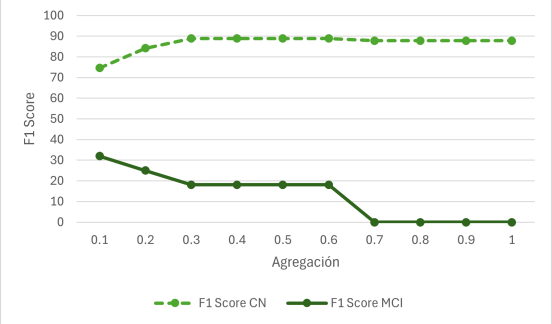


(b) Evaluación del Test SubADReSSo utilizando como entrenamiento el Train SubADReSSo.

Figura 4.7: Resultados en F1 Score por clase, comparando la evaluación del Test ADReSSo y el Test SubADReSSo, con audios segmentados a 3 segundos, utilizando los primeros 20 coeficientes de MFCC y clasificador SVM con Kernel Polinomial.

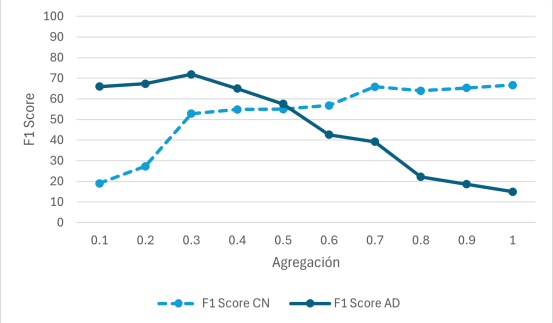


(a) Evaluación del Test ADReSSo utilizando como entrenamiento el Train ADReSSo.

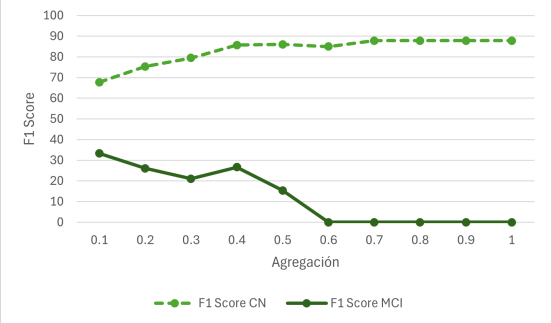


(b) Evaluación del Test SubADReSSo utilizando como entrenamiento el Train SubADReSSo.

Figura 4.8: Resultados en F1 Score por clase, comparando la evaluación del Test ADReSSo y el Test SubADReSSo, con audios segmentados a 5 segundos, utilizando los primeros 20 coeficientes de MFCC y clasificador SVM con Kernel Polinomial.



(a) Evaluación del Test ADReSSo utilizando como entrenamiento el Train ADReSSo.



(b) Evaluación del Test SubADReSSo utilizando como entrenamiento el Train SubADReSSo

Figura 4.9: Resultados en F1 Score por clase, comparando la evaluación del Test ADReSSo y el Test SubADReSSo, con audios segmentados a 10 segundos, utilizando los primeros 20 coeficientes de MFCC y clasificador SVM con Kernel Polinomial.

Como se observó en las Figuras 4.7a, 4.8a, 4.9a, en la columna izquierda, los puntos donde se alcanza un F1-Score balanceado (donde las curvas correspondientes a las clases CN y AD se cruzan) no coinciden con los puntos de agregación que proporcionan el mejor F1-Score en las Figuras 4.7b, 4.8b, 4.9b, de la columna derecha, correspondientes a las clases CN y MCI. Además, los puntajes máximos obtenidos en estas últimas figuras no superan a aquellos observados en la comparación inicial con los audios completos de las Figuras 4.4 y 4.5. En otras palabras, los porcentajes de agregación óptimos para ambos conjuntos de datos (ADReSSo y SubADReSSo) difieren, lo que invalida la segmentación temporal como una mejora. Por consiguiente, se decide continuar trabajando con el audio completo para la extracción de características, tal como se planteó inicialmente.

Para una mejor visualización de los datos plasmados anteriormente en las Figuras 4.7a, 4.8a, 4.9a, se generó la Tabla 4.4 recopilatoria con los mejores puntajes obtenidos.

Segmento	% Agg	ADReSSo			SubADReSSo		
		F1 CN	F1 AD	Accuracy	F1 CN	F1 MCI	Accuracy
No	100	**61.97**	61.97	61.97	84.21	**25.00**	73.91
3seg	50	60.61	**65.79**	**63.38**	87.8	0.00	78.26
5seg	40	50.00	54.05	52.11	**88.89**	18.18	**80.43**
10seg	50	55.07	57.53	56.34	86.08	15.38	76.09

Tabla 4.4: Comparación de los mejores resultados para tratar el audio en segmentos o completo, donde se resaltan a su vez los mejores puntajes por columna.

En la Tabla 4.4, en la primera columna podemos observar el tamaño del segmento tomado en el preprocesamiento de cada audio, en la segunda columna el porcentaje de agregación donde se observa un balance entre el puntaje de F1 Score entre las dos clases evaluadas, ya sea CN vs AD o CN vs MCI, ahora, en las columnas tres, cuatro y cinco se encuentran las métricas obtenidas para el dataset ADReSSo y en las columnas seis, siete y ocho, las métricas obtenidas para el dataset SubADReSSo. Donde se confirma que se obtiene un mejor desempeño cuando **NO** se realiza segmentación del audio. Para una descripción detallada de los experimentos realizados, se recomienda consultar el apéndice A, las Tablas A.3, A.4, A.5, correspondientes a las Figuras 4.7a, 4.8a y 4.9a respectivamente; también podrá encontrar las Tablas A.6, A.7, A.8 que corresponden a las Figuras 4.7b, 4.8b y 4.9b, respectivamente.

4.3 Experimentos con Representaciones Acústicas Basadas en Embeddings

Unos de los objetivos de este trabajo es desarrollar un método para generar datos sintéticos que representen audio y que mejoren la clasificación entre pacientes con

deterioro cognitivo leve y pacientes sanos. Habiendo establecido un *baseline* con representaciones acústicas clásicas, esta sección se adentra en la evaluación de representaciones acústicas más contemporáneas, basadas en *embeddings* neuronales. Estos *embeddings*, extraídos mediante modelos pre-entrenados, ofrecen una alternativa a las características clásicas. Exploraremos la efectividad de *embeddings* generados por *wav2vec* y *wav2vec2.0*, comparando su desempeño en la detección de MCI con los resultados previamente obtenidos. Este análisis permitirá determinar si estas representaciones avanzadas ofrecen una mejora significativa en la tarea de clasificación.

4.3.1 Wav2vec y wav2vec 2.0 como extractor de características

Dadas las características mencionadas en el Marco Teórico sobre *wav2vec*, además de ser entrenado con datos a 16 kHz, tiene la opción de tomar los datos de la capa de contexto (C) o la capa de características (Z); cada una de estas capas produce una salida de 512 neuronas, equivalente a un vector de 512 características abstraídas por cada 10 ms de audio. En nuestro caso, tomaremos los datos extraídos por la capa de características. La diferencia más notable entre estos dos modelos en su etapa de extracción de características es que la del modelo *wav2vec 2.0* es más profunda, e implementa una red convolucional temporal, pero, al final se obtienen la misma cantidad de características, la representación en 512 características por los datos que representan la temporalidad de cada audio.

Para generar un vector resumen que refleje la distribución temporal de los datos de cada característica abstraída del audio, se representa esta vez mediante varios estadísticos [3]. Para ilustrar la idea planteada anteriormente, se presenta la siguiente Figura, donde podemos ver que el audio ingresa completo al modelo *wav2vec*, luego, es descompuesto en 512 características para cada 10ms analizados, para posteri-

[3]Promedio, mediana, desviación estándar, valor máximo, valor mínimo, asimetría, curtosis, cuartil 1 y cuartil 3

ormente, esa distribución temporal representarla en estadísticos funcionales. En resumen, la temporalidad de cada audio se ve representada por nueve estadísticos, donde cada estadístico tiene 512 características extraídas por el modelo *wav2vec* con el que se esté trabajando, para así representar cada audio en una matriz de 9x512.

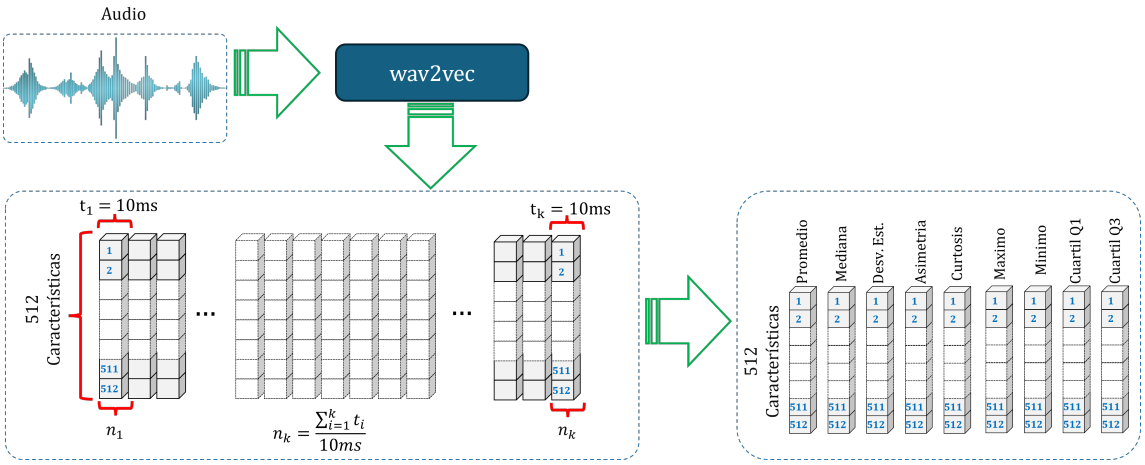


Figura 4.10: Extracción de características con los modelos *wav2vec* y *wav2vec 2.0*.

A continuación se especifican las versiones del modelo *wav2vec2.0* que se utilizaron en nuestros experimentos:

Modelo wav2vec2-large-960h-lv60

El modelo *wav2vec2-large-960h-lv60* ha sido preentrenado utilizando un extenso conjunto de datos no etiquetados que comprende 60,000 horas de audio (*Libri-Light* o LV-60k). Posteriormente, fue ajustado (fine-tuned) empleando 960 horas de datos etiquetados del corpus *LibriSpeech*. Este proceso de entrenamiento dual, que combina aprendizaje no supervisado y supervisado, permite al modelo alcanzar un alto nivel de precisión en tareas de transcripción de voz a texto. El modelo está optimizado para aplicaciones donde se requiere una gran exactitud en el reconocimiento del habla en contextos diversos.

Modelo wav2vec2-large-lv60

El modelo *wav2vec2-large-lv60* comparte con el anterior su fase de preentrenamiento en el conjunto de datos LV-60k. Sin embargo, a diferencia del primero, no ha sido sometido a un ajuste posterior con datos etiquetados. Este modelo es particularmente útil en escenarios donde se desea aprovechar las capacidades de preentrenamiento sin la necesidad de una especialización en un conjunto de datos etiquetados específicos. La ausencia de *fine-tuning* lo hace más generalista, pero a costa de una precisión ligeramente inferior en comparación con modelos ajustados.

Modelo wav2vec2-large-960h-lv60-self

Finalmente, el modelo *wav2vec2-large-960h-lv60-self* se distingue por un proceso de ajuste especial denominado *self-training*, donde se utiliza una técnica de pseudoetiquetado sobre las mismas 960 horas de datos etiquetados del *LibriSpeech*. Este enfoque permite refinar aún más la capacidad del modelo para reconocer patrones en datos de audio, incrementando su precisión en tareas de ASR. Es relevante destacar que este proceso de *self-training* es un paso adicional que busca mejorar el rendimiento del modelo en comparación con el ajuste estándar.

4.3.2 Resultados de referencia

Con el objetivo de seleccionar el modelo más adecuado como extractor de características para nuestro generador de datos a partir de *embeddings*, evaluamos distintas versiones de los modelos *wav2vec* y *wav2vec 2.0*. Esta evaluación se llevó a cabo utilizando diversos clasificadores, con el propósito de comparar sus desempeños y determinar cuál de ellos proporciona los mejores resultados para nuestra tarea en específico.

A continuación se presentan las Figuras que muestran, mediante diagramas de barras, los mejores resultados obtenidos en la evaluación de los diferentes modelos de *wav2vec* utilizando diversos clasificadores. En este capítulo para la clasificación, se ejecutaron diversas pruebas con diferentes tipos de algoritmos de clasificación. Entre los clasificadores tradicionales que se aplicaron fueron: vecinos más cercanos (kNN) con k igual a 5, 10 y 30, máquinas de vectores de soporte (SVM) con kernel Lineal y Polinomial de grado 20 y 30, con el parámetro `class_weight='balanced'`, Random Forest con el parámetro `n_estimator= 100` y `random_state=0` para que sus resultados sean deterministas, al igual que el ADA Boost, pero este último con un `n_estimator= 50`.

En las siguientes gráficas se observa el comportamiento de los clasificadores tanto en el dataset ADReSSo como del SubADReSSo. Recordemos que se usan representaciones de la distribución temporal de los datos que emplea un conjunto de estadísticos, incluyendo el promedio, la mediana, la desviación estándar, el valor máximo, el valor mínimo, la asimetría, la curtosis, y los cuartiles 1 y 3. Dando así matrices de datos de 512 características por cada uno de los 9 estadísticos, que representan cada audio.

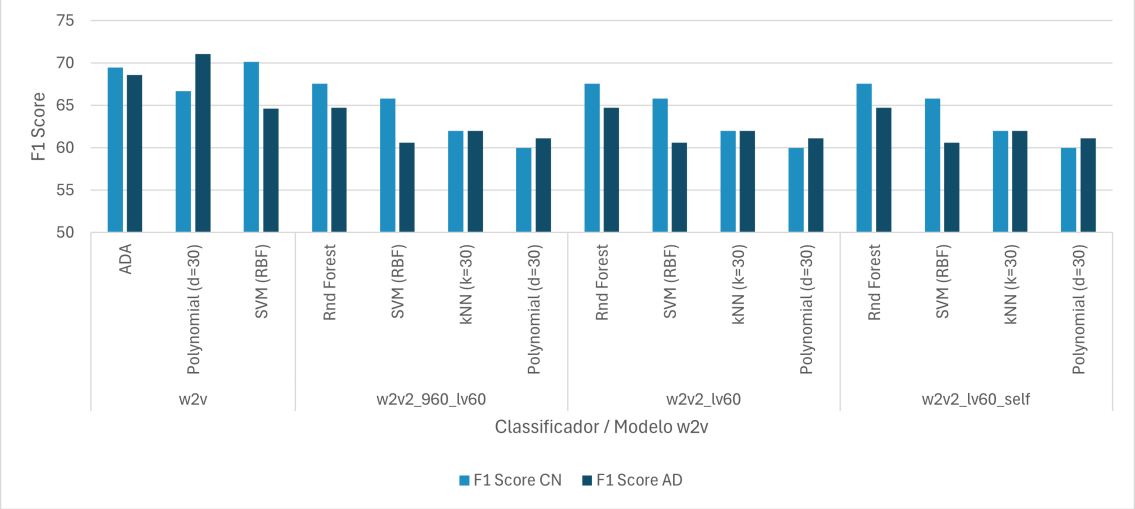


Figura 4.11: Resultados en F1 Score por clase, evaluando el Test ADReSSo, usando diferentes clasificadores entrenados con el Train ADReSSo, utilizando modelos de *wav2vec* para extracción de características con 9 estadísticos.

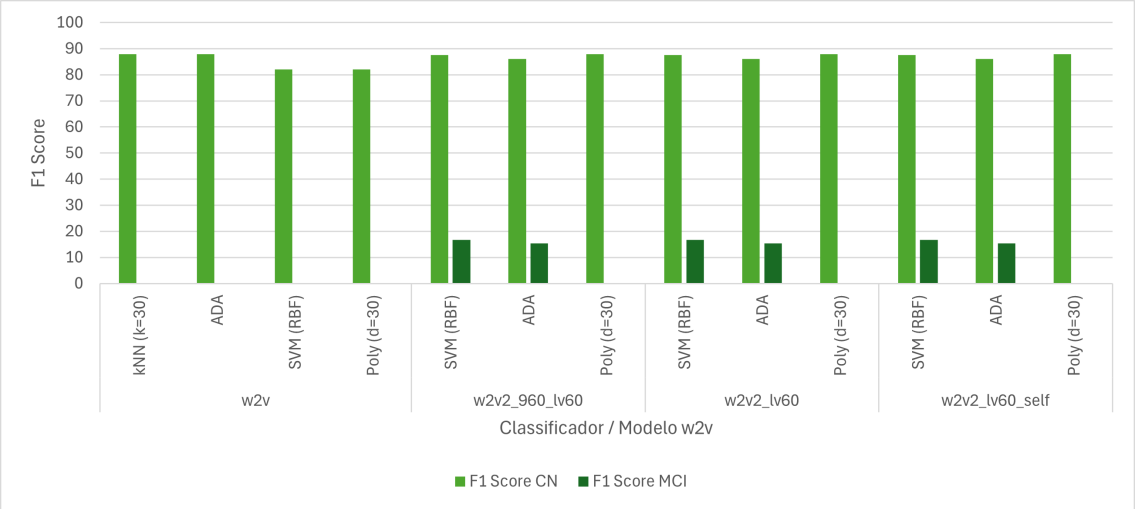


Figura 4.12: Resultados en F1 Score por clase, evaluando el Test SubADReSSo, usando diferentes clasificadores entrenados con el Train SubADReSSo, utilizando modelos de *wav2vec* para extracción de características con 9 estadísticos.

En las Figuras 4.11 y 4.12 se observa la recopilación de los mejores puntajes de F1 Score para cada clase evaluada, luego de obtener los resultados de desempeño

de diferentes modelos de *wav2vec* evaluados con diferentes clasificadores entrenados con el conjunto ADReSSo y SubADReSSo respectivamente; de estas dos Figuras podemos resaltar que los todos los modelos de *wav2vec 2.0* evaluados pudieron representar la clase de MCI, pero, si observamos ambas gráficas, una como complemento de la otra, se encuentra que en ambas la coincidencia de clasificador es el SVM con kernel RBF. Para una descripción detallada de los resultados obtenidos, se recomienda consultar el apéndice A, la Tabla A.16 correspondiente a la Figura 4.11 y A.17 correspondiente a la Figura 4.12.

4.4 Discusión

Dados los resultados de referencia presentados en la subsección 4.2.2 y 4.3.2, se propone un breve análisis, que se referencia con la Tabla 4.5, donde se observa que los puntajes de CN mejoran moderadamente cuando se emplean modelos *wav2vec* como extractores de características, aplicado al conjunto ADReSSo como en el subADReSSo. Al utilizar los 20 primeros coeficientes de MFCC, la mejora es mínima al trabajar con el conjunto de datos ADReSSo. Sin embargo, al utilizar el conjunto de datos SubAdresso para la detección de MCI, la mejora alcanza casi el doble, evidenciando una diferencia significativa en el rendimiento.

Características	**Num. Caract.**	**Clasificador**	**F1 Score CN**	**F1 Score AD**	**F1 Score CN**	**F1 Score MCI**
MFCC [20]	20*5	Polynomial	61.97	61.97	84.21	25.00
w2v2_960_lv60	512*9	SVM (RBF)	65.79	60.61	87.50	16.67
w2v2_lv60	512*9	SVM (RBF)	65.79	60.61	87.50	16.67
w2v2_lv60_self	512*9	SVM (RBF)	65.79	60.61	87.50	16.67

Tabla 4.5: Comparación de características y clasificadores elegidos para experimentos de referencia.

A partir de los resultados obtenidos, se propone un análisis comparativo con

trabajos relacionados en la detección de deterioro cognitivo. La comparación se enfoca en la tarea general de detección de deterioro cognitivo, utilizando los mejores resultados obtenidos para esta tarea, dado que no se encontraron trabajos que emplearan la misma segmentación del dataset para la detección de deterioro cognitivo leve.

Autor	**Preproces**	**Característica**	**Clasificación**	**Acc**	**F1**
Luz (39)	Norm, Filter	eGeMaps	Decision Trees	64.79	-
Balagopalan (10)	-	MFCC + DNN	SVM	67.61	70.89
Chen (19)	-	Sets Features [4]	Vot May. Reg. Log.	67.61	-
Perez-Toro (49)	-	X-Vectores	SVM(RBF)	67.61	67.00
Pappagari (48)	-	SpeechBrain(Enc/Dec)	Reg Logistica	71.80	71.50
Pan (47)	-	wav2vec2.0	Tree Bagger(10)	74.65	74.52
Gauder (27)	Segmnt 5seg	wav2vec2.0	DNN	78.90	-
Este trabajo	-	eGeMaps	RndForest	70.42	70.40
Este trabajo	-	wav2vec	ADA Boost	69.01	69.01

Tabla 4.6: Comparación de trabajos relacionados con el propuesto, evaluando detección del deterioro cognitivo mediante características acústicas.

Como se observa en la Tabla 4.6, los resultados del presente estudio no superan el estado del arte en la detección de deterioro cognitivo. No obstante, dichos resultados son competitivos. Además, es importante destacar que los trabajos relacionados se centran principalmente en la optimización de la detección de deterioro cognitivo en general, mientras que nuestro objetivo principal es investigar el impacto de diferentes técnicas de aumento de datos en la detección específica del deterioro cognitivo leve.

La comparación de representaciones en este capítulo fue fundamental para el presente trabajo, ya que los estudios existentes con el dataset ADReSSo se cen-

tran en la detección general de deterioro cognitivo, mientras que esta investigación se enfoca específicamente en la detección de deterioro cognitivo leve, utilizando un subconjunto del dataset original (SubADReSSo). Si bien el Capítulo 3 revisa trabajos relacionados, ninguno de ellos aborda la tarea específica de detección de MCI con este subconjunto. Por lo tanto, era crucial establecer un baseline propio para las diferentes representaciones acústicas.

CAPÍTULO 5

ANÁLISIS COMPARATIVO DE ESTRATEGIAS DE AUMENTO DE DATOS ACÚSTICOS PARA LA DETECCIÓN DE MCI

En este capítulo, se exploran diversas técnicas de aumento de datos aplicadas a la detección del deterioro cognitivo en habla espontánea. Se evaluarán métodos de adición de elementos reales de una clase similar, como la incorporación de datos de pacientes con deterioro cognitivo avanzado; métodos que modifican directamente la señal, como los que se exploraron en la sección anterior; métodos sencillos que trabajan el aumento sobre la representación de los datos, como la técnica SMOTE; y otros métodos que aplican modelos de Deep Learning para la generación de datos sintéticos, específicamente Redes Generativas Adversarias, para la generación de datos sintéticos basados en *embeddings* extraídos con *wav2vec* y *wav2vec 2.0*. Esto con el fin de evaluar el impacto de las diferentes maneras de aumentar datos, con la intención de balancear el dataset y mejorar la clasificación de pacientes con deterioro cognitivo leve versus pacientes sanos.

5.1 Aumento de Datos Acústicos sobre Representaciones Clásicas

Esta sección se centra en la exploración de técnicas de aumento de datos para mejorar la clasificación de pacientes con deterioro cognitivo leve, utilizando tres tipos de técnicas sobre representaciones clásicas, para este caso los primeros 20 coeficientes de MFCC. La primer técnica consiste en la adición de elementos reales de una clase similar, para este caso la incorporación de datos de pacientes con deterioro cognitivo avanzado; el segundo método son aquellos que modifican directamente la señal, como los que se exploraron en la sección anterior y por último la aplicación de la técnica SMOTE sobre la representación de los datos.

Como idea general para cada técnica de aumento realizada, se agregaron elementos uno a uno y de manera aleatoria para la sección de técnicas que modifican la señal de audio, el experimento se repitió 10 veces; en cada uno de los experimentos, se tomaron las métricas de F1 Score por clase y su macro, el Accuracy, Recall y Precisión, para al final, promediar estos resultados y obtener los resultados mostrados en el documento. La métrica principal de evaluación utilizada en este capítulo es el F1 Score por clase, dada la importancia de identificar correctamente ambas clases en el contexto clínico.

5.1.1 Adición de Elementos Reales

Como primer experimento de aumento de datos con Representaciones Clásicas, se tomaron los elementos de la clase CI (*Cognitive Impairment*), pertenecientes al Train del dataset ADReSSo (recordemos que estos son los elementos con un puntaje MMSE inferior a 24), sin ninguna alteración, para agregarlos uno a uno a la clase MCI de entrenamiento y observa el comportamiento de clasificación del dataset SubADReSSo, en especial, la clase de nuestro interés MCI. Para hacer que los resultados de esta

prueba fueran lo más estable posible, el procedimiento de adición se realizó, tomando elementos de la clase CI aleatoriamente y añadiéndolos uno a uno, se tomó las métricas en cada una de las adiciones; este procedimiento se ejecutó 10 veces, para al final, promediar las métricas. Para una descripción detallada de los resultados obtenidos, se recomienda consultar, la Tabla A.9 en el apéndice A.

Con este experimento se pudo apreciar que la adición de elementos de pacientes con un deterioro cognitivo más avanzado, permite mejorar la clasificación de la clase MCI; dando así que a mayor cantidad de elementos de CI se agregan, mejor es la clasificación de MCI, donde se pudo observar que al balancear el conjunto de Train se obtiene un F1 Score de la clase de interés de 50.09 y al terminar de agregar el restante de los datos se llegó a obtener un puntaje de 55.17. Por lo cual en las siguientes secciones, las técnicas de aumento de datos que modifican directamente la señal de audio como *PitchShift*; *TimeStretch*, *Shift*, *GainTransition* y *PolarityInversion*, serán aplicadas sobre los elementos de clase CI, pertenecientes al Train del dataset ADReSSo.

5.1.2 Técnica SMOTE aplicada a la Representación Clásica

Para abordar el desbalance de clases presente en el conjunto de datos SubADReSSo, se exploró la técnica de aumento de datos SMOTE (Synthetic Minority Over-sampling Technique) (18). SMOTE es un algoritmo que genera datos sintéticos de la clase minoritaria (en este caso, MCI) a partir de los datos existentes. Para generar los datos, se utilizó el Train del conjunto ADReSSo dado que la clase CI de este conjunto aporta a la clasificación de la clase MCI; como en este conjunto tenemos 87 elementos de AD y 79 elementos de la clase CN, pero, la función `SMOTE` de la librería `imblearn.over_sampling` requiere las dos clases para balancear el conjunto, se retiran aleatoriamente 9 elementos de la clase AD, quedando así 78 elementos de AD y 79 elementos de CN. Posteriormente, este dataset se procesó con la función

SMOTE, de la cual se extrae el elemento que se agrega como aumento al conjunto SubADReSSo y realizar la clasificación.

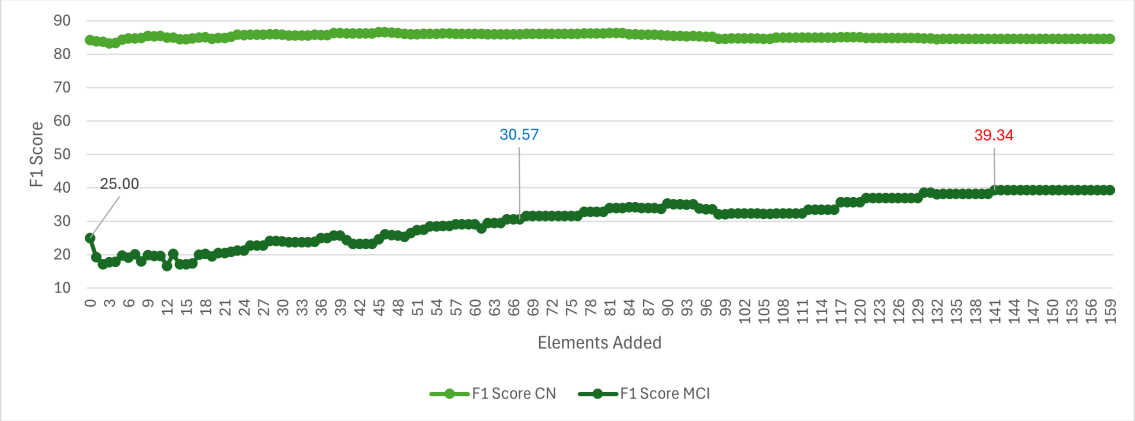


Figura 5.1: Evaluación del F1 Score en la clasificación entre CN y MCI, añadiendo datos sintéticos con SMOTE a la clase MCI, usando MFCC [0-20] hasta duplicar el conjunto de interés. Donde el puntaje se representa con un código de colores, en color negro el puntaje inicial sin aumento; en color azul, el balance del Train y en color rojo se representa el mejor puntaje, en este caso antes de duplicar la clase de interés con 141 elementos añadidos.

Para evaluar el impacto de SMOTE en la clasificación de MCI, se realizaron experimentos agregando instancias sintéticas al conjunto de entrenamiento del dataset SubADReSSo. La Figura 5.1 muestra la evolución del F1 Score de la clase MCI a medida que se agregan instancias sintéticas. Inicialmente, el conjunto de entrenamiento contenía 12 instancias de MCI y 79 instancias de CN, lo que resulta en un F1 Score inicial de 25 puntos para la clase MCI. Se pudo observar que la técnica del SMOTE a medida que se incrementan las instancias sintéticas agregadas al conjunto de entrenamiento, va mejorando la clasificación de MCI. También se observa que el comportamiento del F1 Score decrece en un principio y poco a poco crece hasta sobrepasar el puntaje inicial, cuando se nivela la clase MCI con la clase CN, se obtiene una mejora de 5 puntos, llegando a 30.57; dada la tendencia creciente observada se optó por hacer un experimento más, duplicando la clase de interés.

Luego de nivelar la clase MCI se hicieron pruebas de *oversampling* como se observa en la Figura 5.1, duplicando la cantidad de datos en el conjunto MCI, hasta (79+79) 158 datos, con lo que se encontró que continúa mejorando el F1 Score de MCI hasta casi 10 puntos más, cuando llega a 141 elementos añadidos, se pasa de un puntaje de 30.57 hasta un puntaje de 39.34; cabe resaltar que se observa una estabilización de los puntajes a partir de esta cantidad de elementos y la clase CN se vio mínimamente afectada. Para una descripción detallada de los experimentos realizados, se recomienda consultar la Tabla A.10 en el apéndice A.

5.1.3 Técnicas que Modifican Directamente la Señal

Luego de analizar la anterior prueba, se explora que consecuencia puede llegar a traer la implementación del aumento de datos que se usan de manera convencional para audio. El aumento de datos para este trabajo fue aplicado mediante técnicas que modifican la señal como: *PitchShift*, el cual realiza un cambio de tono en el audio; *TimeStretch*, el cual genera una ralentización del audio entre dos puntos; *Shift*, con el que se genera un retraso o adelanto en el tiempo; *GainTransition*: quien aplica una ganancia variante entre un punto máximo y mínimo en dB a lo largo del audio, *PolarityInversion*: con el cual se invierte la polaridad del audio.

Se realizaron 5 experimentos uno para cada técnica, se agregó una a una cada instancia de la clase AD del Train ADReSSo modificada con cada técnica de aumento, en total 87 instancias fueron agregadas al Train SubADReSSo para cada experimento, para al final evaluar el conjunto de Test SubADReSSo y obtener los resultados en diferentes métricas; para hacer que cada experimento fuera más estable o certero, cada experimento se repitió 10 veces, cada vez los elementos de aumento se agregaron aleatoriamente, para al final promediar los resultados.

La definición de parámetros para cada método de aumento de datos se realizó de

forma empírica, evaluando el resultado de manera manual con el propósito de obtener audios cuyas voces se mantuvieran lo más cercanas a la realidad. Los parámetros seleccionados fueron los siguientes:

1. **PitchShift:** min_semitones=-3, max_semitones=3.
2. **TimeStretch:** min_rate=0.8, max_rate=1.20.
3. **Shift:** min_fraction=-0.5, max_fraction=0.5, rollover=True.
4. **GainTransition:** min_gain_in_db= -13.0, max_gain_in_db= 13.0, min_duration= 10, max_duration= 20, duration_unit= "seconds".
5. **Polaritysion:** Solo se define su activación.

Con estos parámetros se incrementó el conjunto de entrenamiento original de audios hasta agregar todos los datos disponibles que fueron modificados. Es decir, dado que lo que se deseaba era observar qué técnica era la más adecuada para esta tarea, se crearon 5 conjuntos de entrenamiento. Cada uno de ellos conformado por los audios originales y aquellos obtenidos al aplicarles una de las 5 técnicas de aumento propuestas.

Para el aumento con la técnica *Time Stretch*, en la Tabla 5.1 se puede apreciar que el balanceo del conjunto de Train SubADReSSo con 67 elementos añadidos, coincide con el mejor puntaje de aumento de datos para la técnica de *Time Stretch*, pero, esto no pasa siempre, como lo veremos en los siguientes experimentos.

Usando la técnica de aumento *Shift*, en la Tabla 5.1 se puede apreciar que el mayor puntaje se observa con 39 elementos añadidos, con un F1 Score de 42.94 para la clase MCI, que no coincide con el puntaje de aumento de datos obtenidos al balancear el dataset Train SubADReSSo que llega a ser de 37.28.

Para la técnica de aumento *Pitch Shift*, en la Tabla 5.1 se puede apreciar que el mayor puntaje se observa cuando se añaden 38 elementos, con un F1 Score de 38.44

para la clase MCI, que no coincide con el puntaje de aumento de datos obtenidos al balancear el dataset Train SubADReSSo que es de 33.18.

Usando la técnica de aumento *Polarity Inversion*, en la Tabla 5.1 se puede observar que el mayor puntaje se observa cuando se agregan 44 elementos, con un F1 Score de 40.69 para la clase MCI, que no coincide con el puntaje de aumento de datos obtenidos al balancear el dataset Train SubADReSSo que es de 37.72.

Por último, usando la técnica de aumento *Gain*, en la Tabla 5.1 se puede observar que el mayor puntaje se observa cuando se agregan 27 elementos, con un F1 Score de 37.45 para la clase MCI, que no coincide con el puntaje de aumento de datos obtenidos al balancear el dataset Train SubADReSSo que es de 34.59. Para una descripción detallada de los experimentos realizados, se recomienda consultar, las Tablas A.11, A.12, A.13, A.14 y A.15 en el Apéndice A.

5.1.4 Comparación de técnicas de aumento basadas en Representaciones Clásicas

Como resumen de los resultados obtenidos por las diferentes técnicas de aumento de datos utilizando representaciones clásicas, se plasmaran los datos que consideramos más importantes en las siguientes Tablas, para hacer una comparación más sencilla visualmente de las diferentes técnicas.

ElemAdd	Balance %	F1 Score CN	F1 Score MCI	Accuracy	F1 Score AVG	Augment
0	15.19	84.21	25.00	73.91	54.61	Original
75	110.13	79.37	**55.17**	71.74	**67.27**	CI
67	100.00	82.63	37.38	72.83	60.01	Time Stretch
27	49.37	85.79	37.45	**76.96**	61.62	Gain
38	63.29	78.74	38.44	68.48	58.59	Pitch
44	70.89	79.09	40.69	69.13	59.89	Inversion
39	64.56	83.12	42.94	74.13	63.03	Shift
87	125.32	**85.81**	33.97	76.74	59.89	Smote
141	193.67	84.55	39.34	75.43	61.95	Smote

Tabla 5.1: Resultados destacables al hacer aumento de datos para el Train SubADReSSo, con las técnicas de adición de elementos reales, modificación de la señal y SMOTE utilizando MFCC [0-20] como representación.

ElemAdd	Balance %	F1 Score CN	F1 Score MCI	Accuracy	F1 Score AVG	Augment
0	15.19	84.21	25.00	73.91	54.61	Original
67	100.00	78.02	**50.09**	69.57	**64.05**	CI
67	100.00	82.63	37.38	72.83	60.01	Time Stretch
67	100.00	77.08	34.59	66.09	55.84	Gain
67	100.00	73.81	33.18	62.39	53.50	Pitch
67	100.00	76.01	37.72	65.43	56.87	Inversion
67	100.00	77.72	37.28	67.17	57.50	Shift
67	100.00	**84.73**	32.26	75.22	58.49	Smote
146	200.00	84.55	39.34	**75.43**	61.95	Smote

Tabla 5.2: Resultados al balancear la clase MCI con la clase CN en Train SubADReSSo, con las técnicas de adición de elementos reales, modificación de la señal y SMOTE utilizando MFCC [0-20] como representación.

En las Tablas 5.1 y 5.2 se encuentra en la primera columna (ElemAdd) la cantidad de elementos agregados de la clase CI del Train ADReSSo, al conjunto Train SubADReSSo, en la segunda columna se encuentra el porcentaje de balance que se tiene de la clase MCI del conjunto Train SubADReSSo con respecto a la clase CN, por ejemplo para la primera fila encontramos que se han agregado 0 instancias, pero hay un balance del 15.19%, esto es por la clase de interés MCI del conjunto

SubADReSSo parte de 12 instancias mientras que la clase CN tiene 79 instancias. En las columnas 3, 4, 5 y 6 se encuentran las métricas obtenidas de F1 Score para la clase CN, MCI del Test SubADReSSo, además, del Accuracy y el Macro F1 Score, respectivamente. En la última columna podemos observar los diferentes tipos de aumentos aplicados directamente a los audios de la clase CI del Train ADReSSo, en la primera fila encontramos los datos, sin aplicar ningún aumento, en las filas 2, 3, 4, 5 y 6 encontramos las técnicas de *Pitch, Time Stretch, Gain, Inversion* y *Shift*, por último, en la fila 7 encontramos el *Ítem* CI que es una prueba realizada para observar cómo se comportaba al realizar el aumento de datos con datos reales de la clase CI tomados desde el Train ADReSSo sin aplicarle ninguna técnica a estos datos.

De estas Tablas podemos destacar que al realizar el aumento de datos que modifica directamente la señal, es conveniente, pero a su vez, la integración de datos de la clase CI, ósea, personas con un deterioro cognitivo avanzando, también contribuye a la mejora de la clasificación de los datos MCI. También se puede observar que no se coincide en una cantidad de datos específica, para la cantidad de datos agregados, necesarios para obtener el mejor resultado de clasificación y todo depende la técnica utilizada para realizar el aumento de datos. Por consiguiente, el balancear el dataset, para nuestro caso, no siempre es el mejor resultado de clasificación. Esto puede estar dado que hay elementos que contribuyen más que otros y de cierta forma reaccionan mejor o peor a las técnicas de aumento aplicadas realizadas. Otro punto que podemos considerar, es que, cuando se utilizó la técnica *Smote*, tanto en los resultados con dataset balanceado, como en la adición del doble de los datos de la clase MCI luego de estar balanceado, no se lograron resultados que sobresalieran, pero dado que esta técnica demostró una tendencia creciente, se hizo un *oversampling* sobre el entrenamiento de la clase de interés y en ese momento, se obtuvieron mejores resultados, aunque no sería válido compararlos con los experimentos anteriores, dado que a la clase MCI se le agregaron más datos, pero, a su vez este experimento permitió

ver el potencial de la técnica *Smote* como aumento de datos.

5.2 Aumento de Datos Acústicos Basados en Embeddings Neuronales

Esta sección se centra en la exploración de técnicas de aumento de datos para mejorar la clasificación de pacientes con deterioro cognitivo leve, utilizando tres tipos de técnicas sobre representaciones obtenidas de redes neuronales profundas, para este caso se utilizó wav2vec y wav2vec 2.0. La primer técnica consiste en la adición de elementos reales de una clase similar, donde se incorporan datos de pacientes con deterioro cognitivo avanzado. El segundo método utiliza una WGAN-GP y una cWGAN-GP como modelo de generación de datos, como los que se exploraron en la sección anterior y por último la aplicación de la técnica SMOTE sobre la representación de los datos.

Como idea general para cada técnica de aumento realizada, se agregaron elementos uno a uno y de manera aleatoria para la sección de técnicas con modelos de *Deep Learning*, se generó dato a dato, hasta llegar a la cantidad de 87 elementos; el experimento se repitió 10 veces, en cada uno de los experimentos, se tomaron las métricas de F1 Score por clase y su macro, el Accuracy, Recall y Precisión, para al final, promediar estos resultados y obtener los resultados mostrados en el documento. La métrica principal de evaluación utilizada en este capítulo es el F1 Score por clase, dada la importancia de identificar correctamente ambas clases en el contexto clínico.

5.2.1 Adición de Elementos Reales

Para iniciar con los experimentos de aumentos de datos basados en *embeddings*, se tomaron los elementos de la clase CI, pertenecientes al Train del dataset ADReSSo,

sin ninguna alteración, para agregarlos a la clase MCI de entrenamiento y observa el comportamiento de clasificación del dataset SubADReSSo, en especial, la clase de interés: MCI. En la siguiente Tabla A.18 se encuentran la comparación del mejor puntaje obtenido en la evaluación de características para cada modelo wav2vec enfrentado a los puntajes obtenidos con el mismo clasificador y modelo cuando se realiza el aumento de datos, agregando la clase CI. Para una descripción detallada de los experimentos realizados, se recomienda consultar la Tabla A.18 en el apéndice A.

Cuando se evalúan los *embeddings* los cuales la distribución de los datos de su temporalidad fueron representados usando nueve estadísticos para la representación de la distribución temporal y clasificados con un SVM con kernel RBF, se observó, que los modelos wav2vec 2.0 tienen el mismo comportamiento, el aumento de datos en la clase MCI tiene un puntaje de 38.46 y en la clase CN de 75.76. Ahora, analizando el modelo wav2vec primario, se encuentra que la clasificación de la clase MCI es de 34.78 y la clase CN de 78.26. En general, se observa una mejora en la detección de MCI, cercanas a los 22 puntos en F1 Score para los modelos wav2vec2.0 y para el modelo wav2vec primario fue de casi 35 puntos.

También se observa que la clasificación ejecutada con SVM con kernel Polinomial, el puntaje de F1 Score es levemente superior en MCI para los modelos wav2vec 2.0 con 38.71 puntos, pero en CN es de 68.85; mientras que con el modelo wav2vec primario, pasa algo similar, obteniendo un puntaje de 34.48 puntos para MCI y 69.84 para CN. Lo cual refleja una diferencia apreciable en la clasificación de los elementos de CN.

Dado que parece haber un aporte significativo de la clase CI en la clasificación de MCI, y considerando las capacidades del modelo WGAN-GP con capas convolucionales 1D, se propone trabajar con la WGAN-GP de capas convolucionales y convolucionales transpuestas de 1D con el modelo wav2vec2.0 que obtuvo un mayor

puntaje de F1 Score para MCI en comparación con el modelo wav2vec primario.

5.2.2 Técnica SMOTE aplicada a Embeddings Neuronales

En esta sección de técnicas de aumento de datos basados en *embeddings*, también se exploró la técnica de aumento de datos SMOTE, dado el desbalance de clases presente en el conjunto de datos SubADReSSo. Recordemos que SMOTE es un algoritmo que genera datos sintéticos de la clase minoritaria (en este caso, MCI) a partir de los datos existentes. Para generar los datos, se utilizó el Train del conjunto ADReSSo dado que la clase CI de este conjunto aporta a la clasificación de la clase MCI; de la misma forma que la sección anterior en la que se trabajó SMOTE, con la aclaración que en este caso como tenemos una matriz de datos de 512 características por los 9 canales que equivalen a los nueve estadísticos que representan la temporalidad de los datos. Lo que se hizo fue concatenar los canales uno tras otro, para obtener un solo canal de una dimensión. Es importante destacar que, durante los experimentos de aumento de datos, el clasificador que mostró mayor sensibilidad a los datos generados fue el SVM con kernel polinomial de grado 30.

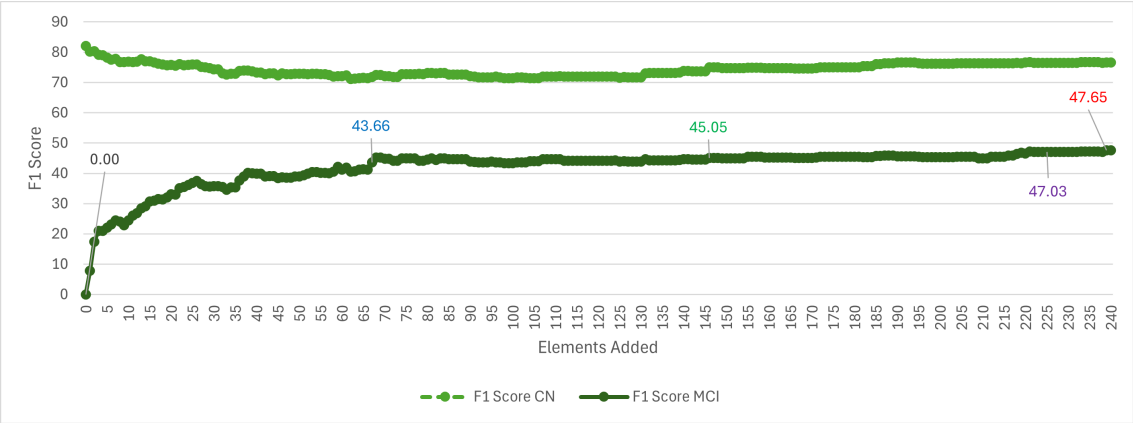


Figura 5.2: Evaluación del F1 Score en la clasificación entre CN y MCI, añadiendo datos sintéticos con SMOTE a la clase MCI. El puntaje se representa con un código de colores, en color negro el puntaje inicial sin aumento; en color azul, el balance del Train y en color rojo se representa el mejor puntaje, en este caso con 239 elementos añadidos. Además, se agregó el color verde, para indicar el punto donde se duplica el conjunto de interés y el morado donde se triplica.

En la Figura 5.2 se pudo observar que la técnica del SMOTE a medida que se incrementan las instancias sintéticas agregadas al conjunto de entrenamiento MCI, va mejorando la clasificación de esta clase. También se observa que el comportamiento del F1 Score crece rápidamente hasta que se nivela la clase MCI con la clase CN, pasando de 0 a un puntaje de 44.17, a partir de ahí, se hicieron pruebas haciendo *oversampling* de hasta 3 veces la clase MCI balanceada y se encontró que la mejora desacelera, pero sigue mejorando tanto el F1 Score de MCI y ahora la clase CN también mejora llegando a un puntaje de 47.20 (el mayor puntaje obtenido con aumento de datos usando de *embeddings*). Para una descripción detallada de los experimentos realizados, se recomienda consultar la Tabla A.19 en el apéndice A.

5.2.3 Generando Embeddings con WGAN-GP y cWGAN-GP

Las Redes Generativas Adversaria (GAN) son modelos compuestos por dos redes neuronales que compiten entre sí: un generador y un discriminador. Ahora se hablará

de la implementación de la GAN con Penalización de Gradiente para generar datos sintéticos de audio para la clase de interés MCI basado en datos de AD.

Implementación de la Red Generativa Adversaria con Penalización de Gradiente (WGAN-GP)

En este caso solo se trabajará el Dataset ADReSSo, puesto que una GAN como modelo de Deep Learning requiere la mayor cantidad de datos disponibles y si usáramos el conjunto de Train SubADReSSo lo más probable es que al implementar una cGAN se sesgara hacia los el conjunto de los pacientes sanos o en caso de hacer una GAN que genere una sola clase (pacientes enfermos), sería muy probable que nuestro conjunto de apenas 12 ejemplos de MCI fuera insuficiente para su entrenamiento.

A continuación, se realiza una descripción breve de cada una de las partes de los modelos implementados para generación de datos de nuestro interés, basado en el modelo WGAN-GP. Las dos redes implementadas son las WGAN-GP y cWGAN-GP, la primera corresponde a una WGAN-GP estándar y la segunda a su versión condicional, con una estructura de redes convolucionales y convoluciones transpuestas unidimensionales, lo que permitiría un análisis más detallado e interrelacionado de los datos por parte de las redes generadora y discriminadora.

Dentro las siguientes secciones se explicará como se implementaron para las diferentes versiones de WGAN-GP y cWGAN-GP, para esta última se agrega la Figura 5.3 con el fin de ilustrar el modelo. Para ver una versión más detallada, diríjase al apéndice C.

1. **Preprocesamiento:**
    - Diarización de los audios.
    - Extracción de características con *wav2vec* o *wav2vec2.0*.

- Representación de la distribución temporal mediante estadísticos.

2. **Implementación de WGAN-GP:**

   - Definición de las arquitecturas del generador y discriminador (dependientes o no de una etiqueta).
   - Configuración de hiperparámetros y funciones de pérdida.
   - Implementación de la penalización de gradiente.

3. **Entrenamiento:**

   - Entrenamiento del modelo generativo adversarial con un enfoque estable y controlado.
   - Monitoreo constante de las pérdidas y métricas para asegurar la convergencia.

4. **Generación de Datos Sintéticos:**

   - Generación controlada y condicionada (de ser necesario) de nuevas muestras de audio.
   - Filtrado de muestras utilizando clasificadores para asegurar la calidad.

5. **Evaluación:**

   - Uso de múltiples clasificadores y métricas de evaluación.
   - Análisis del impacto de los datos sintéticos en el rendimiento de la clasificación.

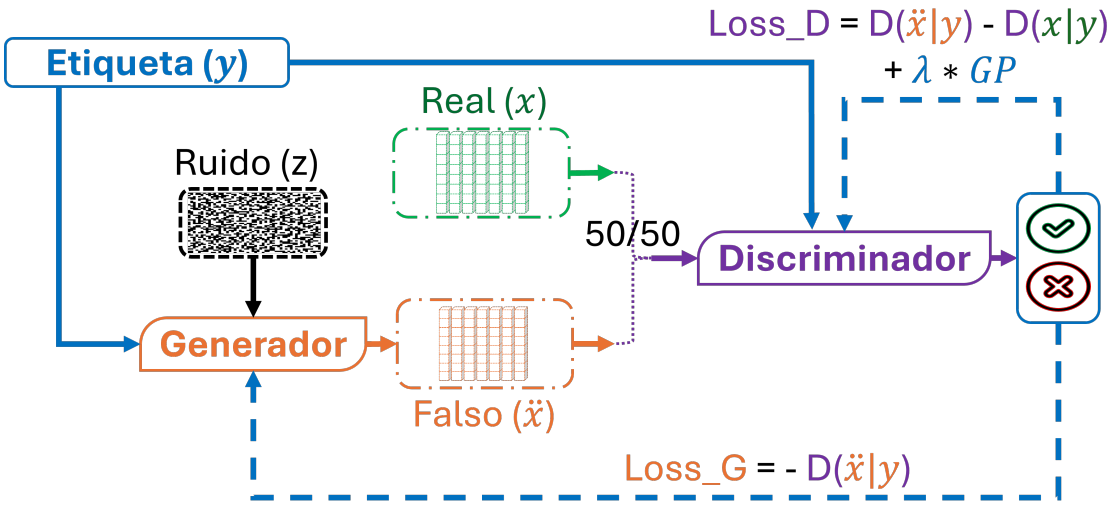


Figura 5.3: Modelo estructural del cWGAN-GP (CNN1D) utilizado para aumento de datos.

Evaluación de WGAN-GP (CNN1D) y cWGAN-GP (CNN1D)

En este caso se entrenaron dos modelos de GAN, el WGAN-GP(CNN1D) y cWGAN-GP(CNN1D), utilizando *wav2vec 2.0* para extraer características acústicas y representando su distribución temporal mediante nueve estadísticos, promedio, mediana, desviación estándar, valor máximo, valor mínimo, asimetría, curtosis, cuartiles 1 y 3 del dataset ADReSSo. Cabe anotar que en este caso, cuando se hizo el proceso de aumento, el clasificador que pudo identificar de mejor manera el aumento fue el SVM con kernel polinomial de grado 30. Las siguientes Figuras muestran el impacto en el F1 Score por clase al añadir los datos sintéticos generados por estos modelos al conjunto de entrenamiento del dataset SubADReSSo.

En los experimentos de aumento de datos con las WGAN-GP (CNN1D) y su versión condicional se observó que a pesar de que la clasificación inicial en 0, al añadir elementos sintéticos con la WGAN-GP (CNN1D) y cWGAN-GP (CNN1D), se logra llegar a un puntaje de 35.71, similar al que se obtenía al añadir elementos reales de CI (38.46 puntos). Por lo cual se podría concluir que no se logra llegar a un puntaje

superior para mejorar la clasificación, pero que los elementos que se están añadiendo efectivamente son similares a los reales, aunque no logran aportar más que la adición de elementos reales de CI a la clasificación de MCI. Otra anotación que podríamos mencionar, es que el modelo condicional cWGAN-GP (CNN1D) no logro generar adecuadamente la distinción entre ambas clases para generar datos más parecidos, por lo cual se podría pensar que los datos muy similares y la cantidad de datos que se tienen de cada clase no son suficientes para que los modelos propuestos de cWGAN-GP puedan generalizar dichas clases de manera satisfactoria. Para una descripción detallada de los experimentos realizados, se recomienda consultar la Tabla A.20 y A.21 en el apéndice A.

5.2.4 Comparación de Técnicas de Aumento basada en Embeddings Neuronales

Como resumen de los resultados obtenidos por las diferentes técnicas de aumento de datos basados en *embeddings* Neuronales que se aplicaron, se mostraron los datos que consideramos más importantes en las Tablas 5.3 y 5.4, para hacer una comparación más sencilla visualmente de las diferentes técnicas.

ElemAdd	Balance %	F1 Score CN	F1 Score MCI	Accuracy	F1 Score AVG	Augment
0	15.19	87.80	00.00	78.26	43.90	Original
75	110.13	68.85	38.71	58.70	53.78	CI
7	24.05	71.88	35.71	60.87	53.79	WGAN-CP (CNN1D)
8	25.31	72.69	35.78	61.74	54.23	cWGAN-CP (CNN1D)
68	101.27	72.49	45.20	63.70	58.85	*Smote*
239	317.72	76.64	47.65	67.83	62.15	*Smote*

Tabla 5.3: Resultados destacables al hacer aumento de datos sobre el Train SubADReSSo, generados a partir de *embeddings* de 9 estadísticos.

ElemAdd	Balance %	F1 Score CN	F1 Score MCI	Accuracy	F1 Score AVG	Augment
0	15.19	87.80	00.00	78.26	43.90	Original
67	100.00	66.67	37.50	56.52	52.08	CI
67	100.00	71.88	35.71	60.87	53.79	WGAN-CP (CNN1D)
67	100.00	71.88	35.71	60.87	53.79	cWGAN-CP (CNN1D)
67	100.00	71.89	43.66	62.83	57.77	*Smote*
146	200.00	75.06	45.05	65.87	60.05	*Smote*
225	300.00	76.52	47.03	67.60	61.77	*Smote*

Tabla 5.4: Resultados al balancear la clase MCI con la clase CN en Train SubADReSSo, generados a partir de *embeddings* de 9 estadísticos.

En las dos Tablas anteriores se encuentra en la primera columna (ElemAdd) la cantidad de elementos agregados de la clase CI del Train ADReSSo, al conjunto Train SubADReSSo, en la segunda columna se encuentra el porcentaje de balance que se tiene de la clase MCI del conjunto Train SubADReSSo con respecto a la clase CN, por ejemplo para la primera fila encontramos que se han agregado 0 instancias, pero hay un balance del 15.19%, esto es por la clase de interés MCI del conjunto SubADReSSo parte de 12 instancias mientras que la clase CN tiene 79 instancias. En las columnas 3, 4, 5 y 6 se encuentran las métricas obtenidas de F1 Score para la clase CN, MCI del Test SubADReSSo, además, del Accuracy y el Macro F1 Score, respectivamente. En la última columna podemos observar los diferentes tipos de aumentos aplicados directamente a los audios de la clase CI del Train ADReSSo, en la primera fila encontramos los datos, sin aplicar ningún aumento, en las filas 2, 3, 4, 5 y 6 encontramos las técnicas de adición de elementos Reales de la clase CI, aplicando WGAN-CP (CNN1D), aplicando cWGAN-CP (CNN1D) y *Smoote*.

Podemos deducir del desempeño similar que obtuvo el modelo WGAN-GP (CNN1D) y el modelo condicional cWGAN-GP (CNN1D), que este último no logro generar adecuadamente la distinción entre ambas clases para generar mejores datos que la WGAN-GP (CNN1D), por lo cual se podría pensar que los datos muy similares y la cantidad de datos que se tienen de cada clase no son suficientes para

que el modelo propuesto de cWGAN-GP pueda generalizar dichas clases de manera satisfactoria. Ahora, comparando entre métodos, se observa que la adición de elementos reales (CI) obtiene un mejor desempeño, aunque no dista mucho de la generación y adición de datos sintéticos por parte de la WGAN-CP (CNN1D). Dado lo anterior podemos decir que aunque las GANs aplicadas no superen la adición de elementos reales, la diferencia entre datos reales y sintéticos es mínima, por lo cual, podríamos intuir que de cierta manera los elementos que se están generando con las GANs, son similares a lo que podría aportar un elemento real. Por lo cual, la técnica con GANs se posiciona como prometedora y se tendrían que hacer más experimentos con una mayor cantidad de datos de la clase MCI en su entrenamiento, para descartarla o comprobar su factibilidad para el uso de aumento de datos en audio, con el fin de detectar MCI. Por otro lado, la técnica de *Smote*, a pesar de ser una técnica convencional, fue la que obtuvo mejores resultados al aplicarse a los *embeddings* neuronales.

5.3 Discusión

En esta sección se realizará una comparativa entre los resultados de aumento aplicados a representaciones clásicas y los aplicados a *embeddings* neuronales, recopilando los resultados en las Tablas 5.5 y 5.6 para que sea más sencillo visualizar los datos.

ElemAdd	Balance %	F1 Score CN	F1 Score MCI	Accuracy	F1 Score AVG	Augment
0	15.19	84.21	25.00	73.91	54.61	Original (Clas)
75	110.13	79.37	**55.17**	71.74	**67.27**	CI (Clas)
67	100.00	82.63	37.38	72.83	60.01	*Time Stretch* (Clas)
27	49.37	85.79	37.45	**76.96**	61.62	*Gain* (Clas)
38	63.29	78.74	38.44	68.48	58.59	*Pitch* (Clas)
44	70.89	79.09	40.69	69.13	59.89	*Inversion* (Clas)
39	64.56	83.12	42.94	74.13	63.03	*Shift* (Clas)
87	125.32	**85.81**	33.97	76.74	59.89	*Smote* (Clas)
141	193.67	84.55	39.34	75.43	61.95	*Smote* (Clas)
0	15.19	87.80	00.00	78.26	43.90	Original (Embed)
75	110.13	68.85	38.71	58.70	53.78	CI (Embed)
7	24.05	71.88	35.71	60.87	53.79	WGAN-CP (CNN1D) (Embed)
8	25.31	72.69	35.78	61.74	54.23	cWGAN-CP (CNN1D) (Embed)
68	101.27	72.49	45.20	63.70	58.85	*Smote* (Embed)
239	317.72	76.64	47.65	67.83	62.15	*Smote* (Embed)

Tabla 5.5: Comparación de Resultados destacables al hacer aumento de datos con las diferentes técnicas sobre el Train del Dataset SubADReSSo.

ElemAdd	Balance %	F1 Score CN	F1 Score MCI	Accuracy	F1 Score AVG	Augment
0	15.19	84.21	25.00	73.91	54.61	Original(Clas)
67	100.00	78.02	**50.09**	69.57	**64.05**	CI (Clas)
67	100.00	82.63	37.38	72.83	60.01	*Time Stretch* (Clas)
67	100.00	77.08	34.59	66.09	55.84	*Gain* (Clas)
67	100.00	73.81	33.18	62.39	53.50	*Pitch* (Clas)
67	100.00	76.01	37.72	65.43	56.87	*Inversion* (Clas)
67	100.00	77.72	37.28	67.17	57.50	*Shift* (Clas)
67	100.00	**84.73**	32.26	75.22	58.49	*Smote* (Clas)
146	200.00	84.55	39.34	**75.43**	61.95	*Smote* (Clas)
0	15.19	87.80	00.00	78.26	43.90	Original (Embed)
67	100.00	66.67	37.50	56.52	52.08	CI (Embed)
67	100.00	71.88	35.71	60.87	53.79	WGAN-CP(CNN1D) (Embed)
67	100.00	71.88	35.71	60.87	53.79	cWGAN-CP (CNN1D) (Embed)
67	100.00	71.89	43.66	62.83	57.77	*Smote* (Embed)
146	200.00	75.06	45.05	65.87	60.05	*Smote* (Embed)
225	300.00	76.52	47.03	67.60	61.77	*Smote* (Embed)

Tabla 5.6: Comparación de Resultados al balancear la clase MCI con la clase CN en Train SubADReSSo, con las diferentes técnicas de aumento de datos.

En las Tablas 5.5 y 5.6, se encuentra en la primera columna (ElemAdd) la cantidad de elementos agregados de la clase CI del Train ADReSSo, al conjunto Train SubADReSSo, en la segunda columna se encuentra el porcentaje de balance que se tiene de la clase MCI del conjunto Train SubADReSSo con respecto a la clase CN, por ejemplo para la primera fila encontramos que se han agregado 0 instancias, pero hay un balance del 15.19%, esto es por la clase de interés MCI del conjunto SubADReSSo parte de 12 instancias mientras que la clase CN tiene 79 instancias. En las columnas 3, 4, 5 y 6 se encuentran las métricas obtenidas de F1 Score para la clase CN, MCI del Test SubADReSSo, además, del Accuracy y el Macro F1 Score, respectivamente. En la última columna podemos observar los diferentes tipos de aumentos aplicados directamente a los audios de la clase CI del Train ADReSSo, pero, esta vez se agregó (Clas) y (Embed) para poder distinguir sobre que representación se aplicó cada aumento; (Clas) se utilizará para las representaciones clásicas, para los cuales se utilizó los 20 primeros coeficientes de MFCC y (Embed) se utilizará para hacer referencia a las representaciones de *embeddings* neuronales, que fueron extraídas con wav2vec2.0, todos los resultados para las técnicas evaluadas en las Tablas de esta subsección utilizaron un clasificador SVM con Kernel Polinomial.

Al contrastar los resultados de clasificación de MCI basados en el F1-Score, considerando especialmente los escenarios con un dataset balanceado, se observa que la estrategia más efectiva fue la adición de elementos reales de la clase CI a la representación clásica. La técnica *SMOTE*, por otro lado, mostró un mejor rendimiento al ser aplicada a los *embeddings* neuronales, constituyendo la segunda mejor estrategia para la clasificación de MCI.

Los modelos WGAN-CP (CNN1D) y cWGAN-CP (CNN1D), si bien superaron a SMOTE con representaciones clásicas y a las clasificaciones de referencia en ambas representaciones, mostraron un rendimiento levemente inferior a las técnicas de aumento [1] que modifican a la señal de audio directamente; este resultado, aunque

[1] *TimeStretch*, *Shift* y *PolarityInversion*

alentador por la funcionalidad demostrada de estas GANs, sugiere la necesidad de utilizar conjuntos de datos más extensos, debido a la naturaleza de los modelos de deep learning, así como la exploración de diferentes configuraciones de parámetros para las GANs.

Según los resultados anteriores indican que la eficacia de las técnicas de aumento de datos depende en gran medida de la representación utilizada y del método de clasificación. Por lo tanto, la selección de la mejor estrategia de aumento debe considerar la optimización conjunta de estos tres elementos: la técnica de aumento, la representación y el clasificador.

CAPÍTULO 6

CONCLUSIONES Y TRABAJO FUTURO

Los principales hallazgos de este trabajo son:

- Se determinó que los coeficientes MFCC combinados con un clasificador SVM con kernel polinomial, proporcionan el mejor rendimiento para la clasificación de MCI utilizando características acústicas clásicas.
- La segmentación de audio no mejoró el rendimiento de la clasificación en el conjunto de datos SubADReSSo, lo que indica que el uso del audio completo es más efectivo en este contexto. Es complejo encontrar una forma para representar los audios de una manera estandarizada, puesto que todos tienen diferente temporalidad, esta era una de las razones por la cual se había segmentado por segundos, en un principio, para conservar la temporalidad.
- La inclusión de datos de pacientes con deterioro cognitivo avanzado (CI) en el conjunto de entrenamiento mejora significativamente la clasificación de MCI. Este resultado sugiere que la información presente en las características acústicas de individuos con CI es relevante para la detección temprana de MCI, y que la incorporación de datos de diferentes etapas del deterioro cognitivo puede contribuir al desarrollo de modelos más robustos.
- Las técnicas que modifican la señal como aumento de datos en audio, como

PitchShift, *TimeStretch*, *Shift*, *GainTransition* y *PolarityInversion*, aplicadas a los datos de la clase CI, también mejoran el rendimiento de la clasificación de MCI. Aunque su impacto es menor en comparación con la inclusión de datos reales de la clase CI, estas técnicas proporcionan una alternativa viable para aumentar la cantidad y diversidad de datos de entrenamiento.

- La técnica SMOTE, aplicada tanto a características acústicas clásicas como a *embeddings*, a pesar de su sencillez, se presenta como una estrategia efectiva para generar datos sintéticos que mejoran la clasificación de MCI. Este hallazgo destaca la versatilidad de SMOTE como método de aumento de datos en este contexto.

- A pesar de utilizar uno de los modelos más estables de GAN, la GAN condicional ha tenido dificultades para generar ambas clases, probablemente sea debido a que se tienen muy pocos datos, para que la GAN pueda generalizar las clases y poder obtener los beneficios de una GAN condicional.

  Las Redes Generativas Adversarias, específicamente las WGAN-GP, muestran potencial para la generación de datos sintéticos para la detección de MCI, aunque su rendimiento aún no supera a la adición de datos reales de CI, las técnicas de aumento de adición de elementos reales, las que modifican la señal o SMOTE. Este trabajo representa una primera exploración del uso de GANs en este campo, abriendo nuevas vías de investigación para el desarrollo de modelos generativos más sofisticados y estrategias de entrenamiento optimizadas.

- Los resultados demuestran que no existe una representación acústica universalmente superior entre las clásicas (MFCC) y los *embeddings* neuronales (*wav2vec 2.0*) para la detección de MCI. La eficacia de cada técnica de aumento de datos depende en gran medida de la representación utilizada. En el caso de las representaciones clásicas, la adición de datos reales de la clase CI obtuvo el mejor rendimiento, seguida por las técnicas que modifican direc-

tamente la señal de audio (*PitchShift, TimeStretch, Shift, GainTransition* y *PolarityInversion*) y, por último, SMOTE. Por otro lado, con los *embeddings* neuronales, SMOTE mostró ser la técnica más efectiva, superando la adición de datos de CI y la generación de datos sintéticos con la WGAN-GP (CNN1D). Esta diferencia en el rendimiento resalta la importancia de considerar la interacción entre la representación de los datos y las técnicas de aumento al abordar problemas de clasificación con datos limitados.

- En cuanto a la comparación entre SMOTE y WGAN-GP (CNN1D) para el aumento de datos sobre las representaciones, se observó una interesante relación. Si bien SMOTE aplicado a los *embeddings* neuronales obtuvo un F1-Score para MCI superior al obtenido con la adición de datos sintéticos generados por la WGAN-GP (CNN1D), esta última, a su vez, superó el rendimiento de SMOTE aplicado a las representaciones clásicas. Este resultado sugiere que, bajo condiciones similares (es decir, aplicadas a los *embeddings*), SMOTE se desempeña mejor que la WGAN-GP (CNN1D) para la tarea específica de clasificación de MCI. Sin embargo, la WGAN-GP (CNN1D) aún presenta un rendimiento competitivo y merece mayor exploración, considerando la posibilidad de mejoras con conjuntos de datos más extensos que permitan un mejor entrenamiento del modelo generativo.

El aumento de datos es una estrategia efectiva para mejorar la detección de MCI a partir del habla, especialmente en escenarios con datos limitados. La inclusión de datos de pacientes con CI y la técnica SMOTE demostraron ser particularmente prometedoras. Si bien las GANs no superaron estas técnicas en este estudio, su aplicación en este contexto es novedosa y merece mayor exploración en futuras investigaciones como las que proponemos a continuación:

- Evaluar las técnicas de aumento en otros conjuntos de datos de habla espontánea para analizar la generalización de los resultados.

- Investigar la aplicación de técnicas de extracción de características más avanzadas, como modelos de *Transformer* preentrenados en audio, y su combinación con las técnicas de aumento de datos exploradas en este trabajo.

- Explorar la extracción de características de la capa de contexto, impulsada por *transformers* que ofrece *wav2vec 2.0* luego de un reentrenamiento de la red.

- Investigar la implementación de otros modelos de GANs, modelos basados en *transformers* o la modificación de la estructura propuesta, para la generación de datos sintéticos más realistas y representativos.

- Investigar la implementación de un modelo de clasificación más robusto que implique redes neuronales, en especial, capas convolucionales unidimensionales que permitan el análisis paralelo de los datos según su clase.

Referencias

[lib] Feature extraction — librosa 0.10.2 documentation.

[2] (2020). Alzheimer's dementia recognition through spontaneous speech (only): The adresso challenge. In *Proceedings of INTERSPEECH 2020*. Last accessed 2023/06/20.

[3] (2023). Multilingual alzheimer's dementia recognition through spontaneous speech. In *Proceedings of ICASSP 2023 SPGC*. Last accessed 2023/06/20.

[4] Agbavor, F. and Liang, H. (2022). Artificial intelligence-enabled end-to-end detection and assessment of alzheimer's disease using voice. *Brain Sciences 2023, Vol. 13, Page 28*, 13:28.

[5] Altinok, D. (2024). Explainable multimodal fusion for dementia detection from text and speech. pages 236–251.

[6] Alzheimer's Association (s.f.). ¿Qué es el Alzheimer? | Español | Alzheimer's Association. Consultado el 30 de agosto de 2024, en https://www.alz.org/alzheimer-demencia/que-es-la-enfermedad-de-alzheimer.

[7] Arjovsky, M., Chintala, S., and Bottou, L. (2017). Wasserstein GAN. *https://arxiv.org/abs/1701.07875*.

[8] Baevski, A., Zhou, H., Mohamed, A., and Auli, M. (2020). wav2vec 2.0: A framework for self-supervised learning of speech representations.

[9] Balagopalan, A. and Novikova, J. (2020). Augmenting bert carefully with underrepresented linguistic features. *ML4H Extended Abstract Arxiv Index*, pages 1–7.

[10] Balagopalan, A. and Novikova, J. (2021). Comparing acoustic-based approaches for alzheimer's disease detection. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 6:4176–4180.

[11] Balamurali, B. T. and Chen, J. M. (2024). Performance assessment of chatgpt versus bard in detecting alzheimer's dementia. *Diagnostics 2024, Vol. 14, Page 817*, 14:817.

[12] Beltrami, D., Gagliardi, G., Favretti, R. R., Ghidoni, E., Tamburini, F., and Calzà, L. (2018). Speech analysis by natural language processing techniques: A possible tool for very early detection of cognitive decline? *Frontiers in Aging Neuroscience*, 10:414837.

[13] Berube, S., Nonnemacher, J., Demsky, C., Glenn, S., Saxena, S., Wright, A., Tippett, D. C., and Hillis, A. E. (2019). Stealing cookies in the twenty-first century: Measures of spoken narrative in healthy versus speakers with aphasia. *American Journal of Speech-Language Pathology*, 28:321.

[14] Betker, J. (2022). TorToiSe text-to-speech.

[15] Bisogni, C., Loia, V., Nappi, M., and Pero, C. (2024). Acoustic features analysis for explainable machine learning-based audio spoofing detection. *Computer Vision and Image Understanding*, 249:104145.

[16] Borade, B. D. and Deshmukh, R. R. (2022). Emotion recognition from non-native marathi speech using mfcc and lpc techniques. *I) Journal*, 11.

[17] Calzà, L., Gagliardi, G., Favretti, R. R., and Tamburini, F. (2021). Linguistic features and automatic classifiers for identifying mild cognitive impairment and dementia. *Computer Speech Language*, 65:101113.

[18] Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). Smote: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16:321–357.

[19] Chen, J., Ye, J., Tang, F., and Zhou, J. (2021). Automatic detection of alzheimer's disease using spontaneous speech only. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 6:4181–4185.

[20] Chien, Y. W., Hong, S. Y., Cheah, W. T., Yao, L. H., Chang, Y. L., and Fu, L. C. (2019). An automatic assessment system for alzheimer's disease based on speech using feature sequence generator and recurrent neural network. *Scientific Reports 2019 9:1*, 9:1–10.

[21] Chittaragi, N. B. and Koolagudi, S. G. (2021). Dialect identification using chroma-spectral shape features with ensemble technique. *Computer Speech Language*, 70:101230.

[22] Daneshfar, F., Kabudian, S. J., and Neekabadi, A. (2020). Speech emotion recognition using hybrid spectral-prosodic features of speech signal/glottal waveform, metaheuristic-based dimensionality reduction, and gaussian elliptical basis function network classifier. *Applied Acoustics*, 166:107360.

[23] Dovetto, F. M. and Marra, F. (2024). Silence and aging: A correlation between pauses and dementia. In *Conference on Corpora for Language and Aging Research APRIL 10-12, 2024*, page 16.

[24] Esmaeilpour, M., Cardinal, P., and Koerich, A. L. (2020). Unsupervised feature

learning for environmental sound classification using weighted cycle-consistent generative adversarial network. *Applied Soft Computing*, 86.

[25] Eyben, F., Scherer, K. R., Schuller, B. W., Sundberg, J., Andre, E., Busso, C., Devillers, L. Y., Epps, J., Laukka, P., Narayanan, S. S., and Truong, K. P. (2016). The geneva minimalistic acoustic parameter set (gemaps) for voice research and affective computing. *IEEE Transactions on Affective Computing*, 7:190–202.

[26] Folstein, M. F., Folstein, S. E., and McHugh, P. R. (1975). "mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12:189–198.

[27] Gauder, L., Pepino, L., Ferrer, L., and Riera, P. (2021). Alzheimer disease recognition using speech-based embeddings from pre-trained models.

[28] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial networks. *Science Robotics*, 3:2672–2680.

[29] Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., and Courville, A. (2017). Improved training of wasserstein gans.

[30] He, R., Al-Tamimi, J., Sánchez-Benavides, G., Montaña-Valverde, G., Gispert, J. D., Grau-Rivera, O., Suárez-Calvet, M., Minguillon, C., Fauria, K., Navarro, A., and Hinzen, W. (2024). Atypical cortical hierarchy in a-positive older adults and its reflection in spontaneous speech. *Brain Research*, 1830:148806.

[31] Ilias, L. and Askounis, D. (2023). Context-aware attention layers coupled with optimal transport domain adaptation and multimodal fusion methods for recognizing dementia from spontaneous speech. *Knowledge-Based Systems*, 277:110834.

[32] Jordal, I., Tamazian, A., Dhyani, T., askskro, Chourdakis, E. T., Karpov, N., Landschoot, C., Angonin, C., Sarioglu, O., BakerBunker, kvilouras, Çoban, E. B.,

Gritskevich, E., Mirus, F., Lee, J.-Y., Choi, K., MarvinLvn, SolomidHero, and Alumäe, T. (2024). iver56/audiomentations: v0.36.0.

[33] Konig, A., Satt, A., Sorin, A., Hoory, R., Derreumaux, A., David, R., and Robert, P. H. (2017). Use of speech analyses within a mobile application for the assessment of cognitive impairment in elderly people. *Current Alzheimer Research*, 15:120–129.

[34] Li, Y. A., Zare, A., and Mesgarani, N. (2021). Starganv2-vc: A diverse, unsupervised, non-parallel framework for natural-sounding voice conversion. In *INTERSPEECH*.

[35] Librosa (s.f.). librosa.feature.mfcc — librosa 0.10.2.post1 documentation. Consultado el 30 de agosto de 2024.

[36] Lin, S. Y., Chang, H. L., Wai, T., Fu, L. C., and Chang, Y. L. (2022a). Contrast-enhanced automatic cognitive impairment detection system with pause-encoder. In *Conference Proceedings IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, volume 2022-October, pages 2796–2801.

[37] Lin, S. Y., Lin, P. C., Lin, Y. C., Lee, Y. J., Wang, C. Y., Peng, S. W., and Wang, P. N. (2022b). The clinical course of early and late mild cognitive impairment. *Frontiers in Neurology*, 13:685636.

[38] Liu, Z., Guo, Z., Ling, Z., and Li, Y. (2021). Detecting alzheimer's disease from speech using neural networks with bottleneck features and data augmentation. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) - Proceedings*, volume 2021-June, pages 7323–7327.

[39] Luz, S., Haider, F., de la Fuente, S., Fromm, D., and MacWhinney, B. (2021). Detecting cognitive decline using speech only: The adresso challenge. *medRxiv*, page 2021.03.24.21254263.

[40] Madhu, A. and Kumaraswamy, S. (2019). Data augmentation using generative adversarial network for environmental sound classification. *European Signal Processing Conference*, 2019-September.

[41] Martínez-Nicolás, I., Llorente, T. E., Martínez-Sánchez, F., and Meilán, J. J. G. (2021). Ten years of research on automatic voice and speech analysis of people with alzheimer's disease and mild cognitive impairment: A systematic review article. *Frontiers in Psychology*, 12:620251.

[42] Mertes, S., Baird, A., Schiller, D., Schuller, B. W., and Andre, E. (2020). An evolutionary-based generative approach for audio data augmentation. In *IEEE 22nd International Workshop on Multimedia Signal Processing, MMSP 2020*. Institute of Electrical and Electronics Engineers Inc.

[43] Mirheidari, B., Blackburn, D., O'malley, R., Venneri, A., Walker, T., Reuber, M., and Christensen, H. (2020). Improving cognitive impairment classification by generative neural network-based feature augmentation.

[44] Mirza, M. and Osindero, S. (2014). Conditional generative adversarial nets.

[45] Olachea-Hernández, C.-A., Villaseñor-Pineda, L., Hernández-Farías, D.-I., Montes-y Gómez, M., and González-Hernández, F.-I. (2025). Detecting alzheimer's disease through the use of language impairment features. pages 142–153.

[46] Organización Mundial de la Salud (s.f.). Demencia. Consultado el 30 de agosto de 2024 en https://www.who.int/es/news-room/fact-sheets/detail/dementia.

[47] Pan, Y., Mirheidari, B., Harris, J. M., Thompson, J. C., Jones, M., Snowden, J. S., Blackburn, D., and Christensen, H. (2021). Using the outputs of different automatic speech recognition paradigms for acoustic- and bert-based alzheimer's dementia detection through spontaneous speech.

[48] Pappagari, R., Cho, J., Joshi, S., Moro-Velazquez, L., Zelasko, P. , Villalba, J., and Dehak, N. (2021). Automatic detection and assessment of alzheimer disease using speech and language technologies in low-resource scenarios.

[49] Pérez-Toro, P. A., Bayerl, S. P., Arias-Vergara, T., Vasquez-Correa, J. C., Klumpp, P., Schuster, M., Nöth, E., Orozco-Arroyave, J. R., and Riedhammer, K. (2021). Influence of the interviewer on the automatic assessment of alzheimer's disease in the context of the adresso challenge.

[50] Qian, Y., Hu, H., and Tan, T. (2019). Data augmentation using generative adversarial networks for robust speech recognition. *Speech Communication*, 114:1–9.

[51] Ramesh, V., Vatanparvar, K., Nemati, E., Nathan, V., Rahman, M. M., and Kuang, J. (2020). Coughgan: Generating synthetic coughs that improve respiratory disease classification. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2020-July:5682–5688.

[52] Rodríguez, J. (2013). Las pausas en el discurso de individuos con demencia tipo alzheimer. estudio de casos. *Lengua y Habla*, pages 253–267.

[53] Rohanian, M., Hough, J., and Purver, M. (2021). Alzheimer's dementia recognition using acoustic, lexical, disfluency and speech pause features robust to noisy inputs.

[54] Rosselli, M., Ardila, A., and Rosselli, D. M. (2012). Deterioro cognitivo leve: Definición y clasificación. *Revista Neuropsicología, Neuropsiquiatría y Neurociencias*, 12:151–162.

[55] Schneider, S., Baevski, A., Collobert, R., and Auli, M. (2019). wav2vec: Unsupervised pre-training for speech recognition.

[56] Seibold, M., Hoch, A., Farshad, M., Navab, N., and Fürnstahl, P. (2022). Conditional generative data augmentation for clinical audio datasets. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 13437 LNCS, pages 345–354.

[57] Shilandari, A., Marvi, H., Khosravi, H., and Wang, W. (2022). Speech emotion recognition using data augmentation method by cycle-generative adversarial networks. *Signal Image Video Processing*, 16:1955–1962.

[58] Toth, L., Hoffmann, I., Gosztolya, G., Vincze, V., Szatloczki, G., Banreti, Z., Pakaski, M., and Kalman, J. (2017). A speech recognition-based solution for the automatic detection of mild cognitive impairment from spontaneous speech. *Current Alzheimer Research*, 15:130–138.

[59] Usher Institute (s.f.). Usher Institute — The University of Edinburgh. last accessed 2023/06/20.

[60] Wang, E. K., Yu, J., Chen, C. M., Kumari, S., and Rodrigues, J. J. (2022). Data augmentation for internet of things dialog system. *Mobile Networks and Applications*, 27:158–171.

[61] Wang, N., Cao, Y., Hao, S., Shao, Z., and Subbalakshmi, K. P. (2021). Modular multi-modal attention network for alzheimer's disease detection using patient audio and language data.

[62] Wang, S., Yang, Y., Wu, Z., Qian, Y., and Yu, K. (2020). Data augmentation using deep generative models for embedding based speaker recognition. *IEEE/ACM Transactions on Audio Speech and Language Processing*, 28:2598–2609.

[63] Weiner, J., H.-C. S. T. (2016). Speech-based detection of alzheimer's disease in conversational german. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH. 08-12-September-2016*, page 1938–1942.

[64] Weiner, J., Frankenberg, C., Telaar, D., Wendelstein, B., Schröder, J., and Schultz, T. (2016). Towards automatic transcription of ILSE — an interdisciplinary longitudinal study of adult development and aging. In Calzolari, N., Choukri, K., Declerck, T., Goggi, S., Grobelnik, M., Maegaard, B., Mariani, J., Mazo, H., Moreno, A., Odijk, J., and Piperidis, S., editors, *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 718–725, Portorož, Slovenia. European Language Resources Association (ELRA).

[65] Yang, Q., Li, X., Ding, X., Xu, F., and Ling, Z. (2022). Deep learning-based speech analysis for alzheimer's disease detection: a literature review. *Alzheimer's Research Therapy*, 14:1–16.

[66] Yella, N. and Rajan, B. (2021). Data augmentation using gan for sound based covid 19 diagnosis. In *Proceedings of the 11th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, IDAACS 2021*, volume 2, pages 606–609. Institute of Electrical and Electronics Engineers Inc.

[67] Zhang, Z., Han, J., Qian, K., Janott, C., Guo, Y., and Schuller, B. (2020). Snore-gans: Improving automatic snore sound classification with synthesized data. *IEEE Journal of Biomedical and Health Informatics*, 24:300–310.

[68] Zhao, Y., Togneri, R., and Sreeram, V. (2020). Replay anti-spoofing countermeasure based on data augmentation with post selection. *Computer Speech and Language*, 64.

[69] Zhu, Y., Obyat, A., Liang, X., Batsis, J. A., and Roth, R. M. (2021). Wavbert: Exploiting semantic and non-semantic speech using wav2vec and bert for dementia detection.

APÉNDICE A

RESULTADOS DETALLADOS

A.1 Referencia para Representaciones Clásicas

Esta sección presenta los resultados de la evaluación de las representaciones acústicas clásicas (MFCCs, Chroma, eGeMAPS) con distintos clasificadores para establecer una baseline. Se detallan métricas de desempeño para cada combinación, tanto para el conjunto de datos ADReSSo como para el conjunto SubADReSSo.

Tabla A.1: Resultados de clasificación, evaluando el Test ADReSSo en cada una de las características clásicas seleccionadas. Todos los clasificadores fueron entrenados con el Train ADReSSo.

Clasificador	Caract	TN	FP	FN	TP	F1 CN	F1 AD	Accur	F1_AVG
ADA	MFCC [20]	21	15	19	16	55.26	48.48	52.11	51.87
	MFCC 20Δ	20	16	13	22	57.97	60.27	59.15	59.12
	Chroma	19	17	21	14	50.00	42.42	46.48	46.21
	eGeMaps	21	15	17	18	56.76	52.94	54.93	54.85
RndForest	MFCC [20]	24	12	14	21	64.86	61.76	63.38	63.31
	MFCC 20Δ	23	13	15	20	62.16	58.82	60.56	60.49
	Chroma	17	19	20	15	46.58	43.48	45.07	45.03
	eGeMaps	26	10	11	24	71.23	69.57	70.42	70.40
kNN (k=5)	MFCC [20]	26	10	19	16	64.20	52.46	59.15	58.33
	MFCC 20Δ	20	16	15	20	56.34	56.34	56.34	56.34
	Chroma	21	15	11	24	61.76	64.86	63.38	63.31
	eGeMaps	30	6	21	14	68.97	50.91	61.97	59.94
kNN (k=10)	MFCC [20]	29	7	19	16	69.05	55.17	63.38	62.11

Continua en la siguiente página

Tabla A.1 – Continuación de la página anterior

Clasificador	File_csv	TN	FP	FN	TP	F1_CN	F1_AD	Accur	F1_AVG
	MFCC 20Δ	27	9	20	15	65.06	50.85	59.15	57.95
	Chroma	29	7	24	11	65.17	41.51	56.34	53.34
	eGeMaps	32	4	23	12	70.33	47.06	61.97	58.69
kNN (k=30)	MFCC [20]	28	8	20	15	66.67	51.72	60.56	59.20
	MFCC 20Δ	22	14	12	23	62.86	63.89	63.38	63.37
	Chroma	30	6	32	3	61.22	13.64	46.48	37.43
	eGeMaps	23	13	16	19	61.33	56.72	59.15	59.02
SVM Linear	MFCC [20]	25	11	15	20	65.79	60.61	63.38	63.20
	MFCC 20Δ	12	24	16	19	37.50	48.72	43.66	43.11
	Chroma	25	11	25	10	58.14	35.71	49.30	46.93
	eGeMaps	24	12	15	20	64.00	59.70	61.97	61.85
SVM Poly (d=20)	MFCC [20]	24	12	15	20	64.00	59.70	61.97	61.85
	MFCC 20Δ	17	19	11	24	53.13	61.54	57.75	57.33
	Chroma	19	17	19	16	51.35	47.06	49.30	49.21
	eGeMaps	24	12	17	18	62.34	55.38	59.15	58.86
SVM Poly (d=30)	MFCC [20]	22	14	13	22	61.97	61.97	61.97	61.97
	MFCC 20Δ	19	17	9	26	59.37	66.67	63.38	63.02
	Chroma	19	17	17	18	52.78	51.43	52.11	52.10
	eGeMaps	24	12	18	17	61.54	53.13	57.75	57.33

En la Tabla A.1, la información se encuentra distribuida de la siguiente manera;en la primera columna el clasificador utilizado, "ADA" para el clasificador *ADABoost*, "RndForest" es *Random forest*, "kNN" es el clasificador de k vecinos mas cercanos donde la k en estas pruebas es igual a 5, 10 y 30. También se encuentra el clasificador SVM con kernel lineal y Polinomial, donde este último esta evaluado con grado 20 y 30. En la segunda corresponde al conjunto de características extraídas. Las columnas 3, 4, 5 y 6, corresponden a las variables de la matriz de confusión, TP (*True Positive*) representa a los aciertos del clasificador en la clase de interés y TN (*True Negative*) corresponden a los aciertos del clasificador en la clase de control "CN" (*Cognitively Normal*); en la columna 7 y 8 encontramos los valores de F1 Score por clase; ya por último encontramos la columna del Accuracy y el F1 Average, respectivamente.

Tabla A.2: Resultados de clasificación, evaluando el Test subADReSSo en cada una de las características clásicas seleccionadas. Los diferentes clasificadores fueron entrenados con el Train subADReSSo.

Clasificador	Caract	TN	FP	FN	TP	F1 CN	F1 MCI	Accur	F1 AVG
ADA	MFCC [20]	33	3	10	0	83.54	0.00	71.74	41.77
	MFCC 20Δ	35	1	10	0	86.42	0.00	76.09	43.21
	Chroma	34	2	10	0	85.00	0.00	73.91	42.50
	eGeMaps	35	1	10	0	86.42	0.00	76.09	43.21
RndForest	MFCC [20]	36	0	10	0	87.80	0.00	78.26	43.90
	MFCC 20Δ	36	0	10	0	87.80	0.00	78.26	43.90
	Chroma	36	0	10	0	87.80	0.00	78.26	43.90
	eGeMaps	36	0	10	0	87.80	0.00	78.26	43.90
kNN (k=5)	MFCC [20]	36	0	10	0	87.80	0.00	78.26	43.90
	MFCC 20Δ	36	0	10	0	87.80	0.00	78.26	43.90
	Chroma	36	0	10	0	87.80	0.00	78.26	43.90
	eGeMaps	36	0	10	0	87.80	0.00	78.26	43.90
kNN (k=10)	MFCC [20]	36	0	10	0	87.80	0.00	78.26	43.90
	MFCC 20Δ	36	0	10	0	87.80	0.00	78.26	43.90
	Chroma	36	0	10	0	87.80	0.00	78.26	43.90
	eGeMaps	36	0	10	0	87.80	0.00	78.26	43.90
kNN (k=30)	MFCC [20]	36	0	10	0	87.80	0.00	78.26	43.90
	MFCC 20Δ	36	0	10	0	87.80	0.00	78.26	43.90
	Chroma	36	0	10	0	87.80	0.00	78.26	43.90
	eGeMaps	36	0	10	0	87.80	0.00	78.26	43.90
SVM Linear	MFCC [20]	33	3	10	0	83.54	0.00	71.74	41.77
	MFCC 20Δ	34	2	10	0	85.00	0.00	73.91	42.50
	Chroma	29	7	7	3	80.56	30.00	69.57	55.28
	eGeMaps	33	3	10	0	83.54	0.00	71.74	41.77
SVM Poly(d=20)	MFCC [20]	32	4	8	2	84.21	25.00	73.91	54.61
	MFCC 20Δ	32	4	10	0	82.05	0.00	69.57	41.03
	Chroma	32	4	7	3	85.33	35.29	76.09	60.31
	eGeMaps	35	1	10	0	86.42	0.00	76.09	43.21
SVM Poly (d=30)	MFCC [20]	32	4	8	2	84.21	25.00	73.91	54.61
	MFCC 20Δ	33	3	10	0	83.54	0.00	71.74	41.77
	Chroma	30	6	7	3	82.19	31.58	71.74	56.89
	eGeMaps	35	1	10	0	86.42	0.00	76.09	43.21

En la Tabla A.2, la información se encuentra distribuida de la siguiente manera;en la primera columna el clasificador utilizado, "ADA" para el clasificador *AD-*

ABoost, "RndForest" es *Random forest*, "kNN" es el clasificador de k vecinos mas cercanos donde la k en estas pruebas es igual a 5, 10 y 30. También se encuentra el clasificador SVM con kernel lineal y Polinomial, donde este último esta evaluado con grado 20 y 30. En la segunda corresponde al conjunto de características extraídas. Las columnas 3, 4, 5 y 6, corresponden a las variables de la matriz de confusión, TP representa a los aciertos del clasificador en la clase de interés y TN corresponden a los aciertos del clasificador en la clase de control "CN"; en la columna 7 y 8 encontramos los valores de F1 Score por clase; ya por último encontramos la columna del Accuracy y el F1 Average, respectivamente.

Tabla A.3: Resultados de clasificación, evaluando Test ADReSSo, con audios segmentados a 3 segundos, utilizando los 20 primeros coeficientes de MFCC como caracterización y SVM con kernel polinomial de grado 30 como clasificador.

					Clase CN			Clase AD				F1 Sc
Agg	TN	FP	FN	TP	Prec	Recl	F1 Score	Prec	Recl	F1 Score	Accur	AVG
0.1	0	36	0	35	0.00	0.00	0.00	49.30	100.00	66.04	49.30	33.02
0.2	0	36	1	34	0.00	0.00	0.00	48.57	97.14	64.76	47.89	32.38
0.3	3	33	3	32	50.00	8.33	14.29	49.23	91.43	64.00	49.30	39.14
0.4	7	29	4	31	63.64	19.44	29.79	51.67	88.57	65.26	53.52	47.53
0.5	20	16	10	25	66.67	55.56	60.61	60.98	71.43	65.79	63.38	63.20
0.6	28	8	21	14	57.14	77.78	65.88	63.64	40.00	49.12	59.15	57.50
0.7	35	1	31	4	53.03	97.22	68.63	80.00	11.43	20.00	54.93	44.31
0.8	36	0	35	0	50.70	100.00	67.29	0.00	0.00	0.00	50.70	33.64
0.9	36	0	35	0	50.70	100.00	67.29	0.00	0.00	0.00	50.70	33.64
1	36	0	35	0	50.70	100.00	67.29	0.00	0.00	0.00	50.70	33.64

Tabla A.4: Resultados de clasificación, evaluando Test ADReSSo, con audios segmentados a 5 segundos, utilizando los 20 primeros coeficientes de MFCC como caracterización y SVM con kernel polinomial de grado 30 como clasificador.

					Clase CN			Clase AD				F1 Sc
Agg	TN	FP	FN	TP	Prec	Recl	F1 Score	Prec	Recl	F1 Score	Accur	AVG
0.1	2	34	1	34	66.67	5.56	10.26	50.00	97.14	66.02	50.70	38.14
0.2	7	29	3	32	70.00	19.44	30.43	52.46	91.43	66.67	54.93	48.55
0.3	14	22	6	29	70.00	38.89	50.00	56.86	82.86	67.44	60.56	58.72
0.4	17	19	15	20	53.13	47.22	50.00	51.28	57.14	54.05	52.11	52.03
0.5	26	10	21	14	55.32	72.22	62.65	58.33	40.00	47.46	56.34	55.05
0.6	30	6	23	12	56.60	83.33	67.42	66.67	34.29	45.28	59.15	56.35
0.7	34	2	30	5	53.13	94.44	68.00	71.43	14.29	23.81	54.93	45.90
0.8	35	1	32	3	52.24	97.22	67.96	75.00	8.57	15.38	53.52	41.67
0.9	36	1	34	2	51.47	98.61	67.63	37.50	4.29	7.69	52.11	37.66
1	36	0	35	0	50.70	100.00	67.29	0.00	0.00	0.00	50.70	33.64

Tabla A.5: Resultados de clasificación, evaluando Test ADReSSo, con audios segmentados a 10 segundos, utilizando los 20 primeros coeficientes de MFCC como caracterización y SVM con kernel polinomial de grado 30 como clasificador.

					Clase CN			Clase AD				F1 Sc
Agg	TN	FP	FN	TP	Prec	Recl	F1 Score	Prec	Recl	F1 Score	Accur	AVG
0.1	4	32	2	33	66.67	11.11	19.05	50.77	94.29	66.00	52.11	42.52
0.2	6	30	2	33	75.00	16.67	27.27	52.38	94.29	67.35	54.93	47.31
0.3	14	22	3	32	82.35	38.89	52.83	59.26	91.43	71.91	64.79	62.37
0.4	17	19	9	26	65.38	47.22	54.84	57.78	74.29	65.00	60.56	59.92
0.5	19	17	14	21	57.58	52.78	55.07	55.26	60.00	57.53	56.34	56.30
0.6	23	13	22	13	51.11	63.89	56.79	50.00	37.14	42.62	50.70	49.71
0.7	30	6	25	10	54.55	83.33	65.93	62.50	28.57	39.22	56.34	52.57
0.8	31	5	30	5	50.82	86.11	63.92	50.00	14.29	22.22	50.70	43.07
0.9	33	4	31	4	51.17	90.28	65.29	55.00	11.43	18.61	51.41	41.95
1	34	2	32	3	51.52	94.44	66.67	60.00	8.57	15.00	52.11	40.83

En la Tabla A.3, A.4 y A.5, la información se encuentra distribuida de la siguiente manera; en la primera columna la forma decimal del porcentaje de agregación para la clasificación. Las columnas 2, 3, 4 y 5, corresponden a las variables de la matriz de confusión, TP representa a los aciertos del clasificador en la clase de interés "MCI" y TN corresponden a los aciertos del clasificador en la clase de control "CN"; luego encontramos las métricas de la matriz de clasificación por clase (Precision, Recall y F1 Score), en la columnas 6, 7 y 8, las correspondientes a la clase CN y en las columnas 9, 10 y 11 la clase MCI; ya por último encontramos la columna del Accuracy y el F1 Average, respectivamente.

Tabla A.6: Resultados de clasificación, evaluando Test SubADReSSo, con audios segmentados a 3 segundos, utilizando los 20 primeros coeficientes de MFCC como caracterización y SVM con kernel polinomial de grado 30 como clasificador.

					Clase CN			Clase MCI				F1 Sc
Agg	TN	FP	FN	TP	Prec	Recl	F1 Score	Prec	Recl	F1 Score	Accur	AVG
0.1	10	26	3	7	76.92	27.78	40.82	21.21	70.00	32.56	36.96	36.69
0.2	21	15	8	2	72.41	58.33	64.62	11.76	20.00	14.81	50.00	39.72
0.3	34	2	8	2	80.95	94.44	87.18	50.00	20.00	28.57	78.26	57.88
0.4	36	0	9	1	80.00	100.00	88.89	100.00	10.00	18.18	80.43	53.54
0.5	36	0	10	0	78.26	100.00	87.80	0.00	0.00	0.00	78.26	43.90
0.6	36	0	10	0	78.26	100.00	87.80	0.00	0.00	0.00	78.26	43.90
0.7	36	0	10	0	78.26	100.00	87.80	0.00	0.00	0.00	78.26	43.90
0.8	36	0	10	0	78.26	100.00	87.80	0.00	0.00	0.00	78.26	43.90
0.9	36	0	10	0	78.26	100.00	87.80	0.00	0.00	0.00	78.26	43.90
1	36	0	10	0	78.26	100.00	87.80	0.00	0.00	0.00	78.26	43.90

Tabla A.7: Resultados de clasificación, evaluando Test SubADReSSo, con audios segmentados a 5 segundos, utilizando los 20 primeros coeficientes de MFCC como caracterización y SVM con kernel polinomial de grado 30 como clasificador.

					Clase CN			Clase MCI				F1 Sc
Agg	TN	FP	FN	TP	Prec	Recl	F1 Score	Prec	Recl	F1 Score	Accur	AVG
0.1	25	11	6	4	80.65	69.44	74.63	26.67	40.00	32.00	63.04	53.31
0.2	32	4	8	2	80.00	88.89	84.21	33.33	20.00	25.00	73.91	54.61
0.3	36	0	9	1	80.00	100.00	88.89	100.00	10.00	18.18	80.43	53.54
0.4	36	0	9	1	80.00	100.00	88.89	100.00	10.00	18.18	80.43	53.54
0.5	36	0	9	1	80.00	100.00	88.89	100.00	10.00	18.18	80.43	53.54
0.6	36	0	9	1	80.00	100.00	88.89	100.00	10.00	18.18	80.43	53.54
0.7	36	0	10	0	78.26	100.00	87.80	0.00	0.00	0.00	78.26	43.90
0.8	36	0	10	0	78.26	100.00	87.80	0.00	0.00	0.00	78.26	43.90
0.9	36	0	10	0	78.26	100.00	87.80	0.00	0.00	0.00	78.26	43.90
1	36	0	10	0	78.26	100.00	87.80	0.00	0.00	0.00	78.26	43.90

Tabla A.8: Resultados de clasificación, evaluando Test SubADReSSo, con audios segmentados a 10 segundos, utilizando los 20 primeros coeficientes de MFCC como caracterización y SVM con kernel polinomial de grado 30 como clasificador.

					Clase CN			Clase MCI				F1 Sc
Agg	TN	FP	FN	TP	Prec	Recl	F1 Score	Prec	Recl	F1 Score	Accur	AVG
0.1	21	15	5	5	80.77	58.33	67.74	25.00	50.00	33.33	56.52	50.54
0.2	26	10	7	3	78.79	72.22	75.36	23.08	30.00	26.09	63.04	50.72
0.3	29	7	8	2	78.38	80.56	79.45	22.22	20.00	21.05	67.39	50.25
0.4	33	3	8	2	80.49	91.67	85.71	40.00	20.00	26.67	76.09	56.19
0.5	34	2	9	1	79.07	94.44	86.08	33.33	10.00	15.38	76.09	50.73
0.6	34	2	10	0	77.27	94.44	85.00	0.00	0.00	0.00	73.91	42.50
0.7	36	0	10	0	78.26	100.00	87.80	0.00	0.00	0.00	78.26	43.90
0.8	36	0	10	0	78.26	100.00	87.80	0.00	0.00	0.00	78.26	43.90
0.9	36	0	10	0	78.26	100.00	87.80	0.00	0.00	0.00	78.26	43.90
1	36	0	10	0	78.26	100.00	87.80	0.00	0.00	0.00	78.26	43.90

En la Tabla A.3, A.4 y A.5, la información se encuentra distribuida de la siguiente manera; en la primera columna la forma decimal del porcentaje de agregación para la clasificación. Las columnas 2, 3, 4 y 5, corresponden a las variables de la matriz de confusión, TP representa a los aciertos del clasificador en la clase de interés "MCI" y TN corresponden a los aciertos del clasificador en la clase de control "CN"; luego encontramos las métricas de la matriz de clasificación por clase (Precision, Recall y F1 Score), en la columnas 6, 7 y 8, las correspondientes a la clase CN y en las columnas 9, 10 y 11 la clase MCI; ya por último encontramos la columna del Accuracy y el F1 Average, respectivamente.

A.2 Aumentos de Datos para Representaciones Clásicas

En esta sección se detallan los resultados de los experimentos de aumento de datos aplicados a las representaciones clásicas, incluyendo la adición de datos de la clase CI, las técnicas que modifican la señal de audio y SMOTE. Se presentan los resultados para cada técnica de aumento, variando la cantidad de datos añadidos.

Tabla A.9: Evaluación de la clasificación Test SubADReSSo, añadiendo datos reales de la clase CI a la clase MCI, usando MFCC [0-20] como extractor de características y SVM Polinomial de grado 30 como clasificador.

						Clase CN			Clase MCI				F1 Sc
Add	Add%	TN	FP	FN	TP	Prec	Recl	F1 Sc	Prec	Recl	F1 Sc	Accur	AVG
0	15.19	32	4	8	2	80.00	88.89	84.21	33.33	20.00	25.00	73.91	54.61
1	16.46	31.1	4.9	8	2	79.47	86.39	82.74	31.33	20.00	24.09	71.96	53.41
2	17.72	31	5	8.1	1.9	79.22	86.11	82.48	29.78	19.00	22.89	71.52	52.69
3	18.99	31	5	8.1	1.9	79.21	86.11	82.47	30.97	19.00	22.97	71.52	52.72
4	20.25	31.1	4.9	8.2	1.8	79.08	86.39	82.52	29.54	18.00	21.79	71.52	52.16
5	21.52	31.6	4.4	8.2	1.8	79.38	87.78	83.34	31.49	18.00	22.30	72.61	52.82
6	22.78	31.6	4.4	8.1	1.9	79.59	87.78	83.42	33.48	19.00	23.19	72.83	53.31
7	24.05	30.9	5.1	8.1	1.9	79.21	85.83	82.29	30.32	19.00	22.26	71.30	52.28
8	25.32	30.7	5.3	7.9	2.1	79.51	85.28	82.22	31.64	21.00	24.20	71.30	53.21
9	26.58	30.7	5.3	7.8	2.2	79.73	85.28	82.31	35.24	22.00	25.08	71.52	53.69
10	27.85	30.9	5.1	7.8	2.2	79.84	85.83	82.62	35.96	22.00	25.30	71.96	53.96
11	29.11	30.5	5.5	7.9	2.1	79.40	84.72	81.88	33.22	21.00	23.96	70.87	52.92
12	30.38	30.7	5.3	7.8	2.2	79.71	85.28	82.32	34.87	22.00	25.26	71.52	53.79
13	31.65	30.8	5.2	7.6	2.4	80.22	85.56	82.72	36.57	24.00	27.14	72.17	54.93
14	32.91	30.8	5.2	7.5	2.5	80.42	85.56	82.83	37.52	25.00	28.14	72.39	55.48
15	34.18	30.7	5.3	7.3	2.7	80.84	85.28	82.89	38.54	27.00	29.54	72.61	56.22
16	35.44	31	5	7.3	2.7	80.99	86.11	83.35	40.37	27.00	30.01	73.26	56.68
17	36.71	31.2	4.8	7	3	81.78	86.67	83.98	44.37	30.00	32.94	74.35	58.46
18	37.97	31.2	4.8	6.8	3.2	82.16	86.67	84.23	44.75	32.00	35.33	74.78	59.78
19	39.24	30.7	5.3	6.8	3.2	81.99	85.28	83.43	42.64	32.00	33.87	73.70	58.65
20	40.51	30.2	5.8	7	3	81.23	83.89	82.41	39.20	30.00	31.57	72.17	56.99
21	41.77	30.1	5.9	7.1	2.9	80.95	83.61	82.10	38.35	29.00	30.47	71.74	56.29
22	43.04	30	6	7.1	2.9	80.95	83.33	82.03	32.36	29.00	29.95	71.52	55.99
23	44.30	29.9	6.1	6.9	3.1	81.42	83.06	82.09	33.01	31.00	31.14	71.74	56.62
24	45.57	29.7	6.3	6.9	3.1	81.30	82.50	81.72	33.48	31.00	30.90	71.30	56.31
25	46.84	29.5	6.5	6.8	3.2	81.44	81.94	81.53	32.84	32.00	31.38	71.09	56.45

Continua en la siguiente página

Tabla A.9 – Continuación de la página anterior

						Clase CN			Clase MCI				F1 Sc
Add	Add%	TN	FP	FN	TP	Prec	Recl	F1 Sc	Prec	Recl	F1 Sc	Accur	AVG
26	48.10	29.3	6.7	6.9	3.1	81.08	81.39	81.09	31.41	31.00	30.35	70.43	55.72
27	49.37	28.9	7.1	6.8	3.2	81.05	80.28	80.54	31.89	32.00	31.04	69.78	55.79
28	50.63	28.6	7.4	6.2	3.8	82.27	79.44	80.69	35.21	38.00	35.58	70.43	58.13
29	51.90	28.5	7.5	5.9	4.1	83.09	79.17	80.86	36.36	41.00	37.30	70.87	59.08
30	53.16	28.5	7.5	5.9	4.1	83.07	79.17	80.83	37.58	41.00	37.44	70.87	59.14
31	54.43	28.3	7.7	6	4	82.71	78.61	80.35	36.67	40.00	36.43	70.22	58.39
32	55.70	28	8	5.7	4.3	83.31	77.78	80.14	37.62	43.00	38.11	70.22	59.13
33	56.96	27	9	5.5	4.5	83.29	75.00	78.72	33.36	45.00	37.75	68.48	58.23
34	58.23	28.1	7.9	5.3	4.7	84.39	78.06	80.78	38.49	47.00	41.20	71.30	60.99
35	59.49	27.8	8.2	5	5	84.94	77.22	80.54	39.71	50.00	42.98	71.30	61.76
36	60.76	27.5	8.5	4.4	5.6	86.41	76.39	80.69	41.77	56.00	46.48	71.96	63.59
37	62.03	27.7	8.3	4.5	5.5	86.28	76.94	80.95	41.44	55.00	45.90	72.17	63.42
38	63.29	27.8	8.2	4.4	5.6	86.67	77.22	81.27	41.93	56.00	46.55	72.61	63.91
39	64.56	27.7	8.3	4.5	5.5	86.29	76.94	80.96	41.41	55.00	45.87	72.17	63.42
40	65.82	27.6	8.4	4.7	5.3	85.77	76.67	80.51	40.23	53.00	44.22	71.52	62.37
41	67.09	27.7	8.3	4.7	5.3	85.81	76.94	80.72	40.40	53.00	44.37	71.74	62.54
42	68.35	27.4	8.6	4.6	5.4	85.85	76.11	80.37	39.45	54.00	44.60	71.30	62.48
43	69.62	27.4	8.6	4.6	5.4	85.84	76.11	80.38	39.40	54.00	44.64	71.30	62.51
44	70.89	27.1	8.9	4.3	5.7	86.65	75.28	80.23	39.46	57.00	45.66	71.30	62.94
45	72.15	26.4	9.6	3.9	6.1	87.50	73.33	79.52	38.66	61.00	46.72	70.65	63.12
46	73.42	26.4	9.6	3.9	6.1	87.38	73.33	79.55	38.79	61.00	46.98	70.65	63.26
47	74.68	26	10	3.8	6.2	87.40	72.22	78.98	38.17	62.00	47.04	70.00	63.01
48	75.95	26	10	3.9	6.1	87.09	72.22	78.85	37.88	61.00	46.49	69.78	62.67
49	77.22	26.1	9.9	3.9	6.1	87.14	72.50	79.00	38.31	61.00	46.70	70.00	62.85
50	78.48	26	10	3.8	6.2	87.41	72.22	78.97	38.27	62.00	47.01	70.00	62.99
51	79.75	25.8	10.2	3.8	6.2	87.32	71.67	78.63	37.66	62.00	46.64	69.57	62.63
52	81.01	25.7	10.3	3.7	6.3	87.58	71.39	78.54	37.93	63.00	47.09	69.57	62.82
53	82.28	25.8	10.2	3.8	6.2	87.34	71.67	78.60	37.78	62.00	46.67	69.57	62.63
54	83.54	25.7	10.3	3.5	6.5	88.22	71.39	78.77	38.71	65.00	48.17	70.00	63.47
55	84.81	25.4	10.6	3.5	6.5	88.11	70.56	78.23	37.86	65.00	47.59	69.35	62.91
56	86.08	25.6	10.4	3.3	6.7	88.84	71.11	78.84	39.01	67.00	49.02	70.22	63.93
57	87.34	25.6	10.4	3.3	6.7	88.84	71.11	78.84	39.01	67.00	49.02	70.22	63.93
58	88.61	25.5	10.5	3.2	6.8	89.03	70.83	78.79	39.22	68.00	49.56	70.22	64.18
59	89.87	25.4	10.6	3.2	6.8	88.96	70.56	78.60	39.07	68.00	49.44	70.00	64.02
60	91.14	25.4	10.6	3.2	6.8	88.99	70.56	78.59	39.07	68.00	49.40	70.00	64.00
61	92.41	25.2	10.8	3.2	6.8	88.90	70.00	78.23	38.56	68.00	49.05	69.57	63.64
62	93.67	25	11	3.2	6.8	88.80	69.44	77.85	38.14	68.00	48.72	69.13	63.29
63	94.94	25.1	10.9	3.2	6.8	88.83	69.72	78.05	38.35	68.00	48.90	69.35	63.47
64	96.20	24.9	11.1	3.2	6.8	88.76	69.17	77.65	37.96	68.00	48.56	68.91	63.11

Continua en la siguiente página

Tabla A.9 – Continuación de la página anterior

						Clase CN			Clase MCI				F1 Sc
Add	Add%	TN	FP	FN	TP	Prec	Recl	F1 Sc	Prec	Recl	F1 Sc	Accur	AVG
65	97.47	24.8	11.2	3.1	6.9	89.07	68.89	77.58	38.06	69.00	48.88	68.91	63.23
66	98.73	25	11	3.1	6.9	89.17	69.44	77.94	38.58	69.00	49.24	69.35	63.59
67	100.00	24.9	11.1	2.9	7.1	89.78	69.17	78.02	38.92	71.00	50.09	69.57	64.05
68	101.27	24.8	11.2	2.7	7.3	90.37	68.89	78.07	39.38	73.00	50.99	69.78	64.53
69	102.53	24.6	11.4	2.2	7.8	91.84	68.33	78.33	40.63	78.00	53.39	70.43	65.86
70	103.80	24.4	11.6	2.2	7.8	91.78	67.78	77.95	40.18	78.00	53.01	70.00	65.48
71	105.06	24.4	11.6	2.2	7.8	91.78	67.78	77.95	40.18	78.00	53.01	70.00	65.48
72	106.33	24.3	11.7	2.2	7.8	91.75	67.50	77.75	39.97	78.00	52.82	69.78	65.29
73	107.59	24.7	11.3	2.2	7.8	91.87	68.61	78.54	40.79	78.00	53.55	70.65	66.04
74	108.86	24.8	11.2	2.2	7.8	91.90	68.89	78.73	41.00	78.00	53.73	70.87	66.23
75	110.13	25	11	2	8	92.59	69.44	79.37	42.11	80.00	55.17	71.74	67.27

En la Tabla A.9, la información se encuentra distribuida de la siguiente manera; en la primera columna los elementos agregados al conjunto de Train, en la segunda columna, el *Balance Rate* en porcentaje, que surge de dividir el número actual de elementos en el conjunto de Train sobre sobre el número total de elementos objetivos, esto multiplicado por 100. Las columnas 3, 4, 5 y 6, corresponden a las variables de la matriz de confusión, TP representa a los aciertos del clasificador en la clase de interés y TN corresponden a los aciertos del clasificador en la clase de control "CN"; luego encontramos las métricas de la matriz de clasificación por clase (Precision, Recall y F1 Score), en la columnas 7, 8 y 9, las correspondientes a la clase CN y en las columnas 10, 11 y 12 la clase de interés; ya por último encontramos la columna del Accuracy y el F1 Average, respectivamente.

Tabla A.10: Evaluación de la clasificación Test SubADReSSo, añadiendo datos sintéticos con SMOTE a la clase MCI, usando MFCC [0-20] como extractor de características y SVM Polinomial de grado 30 como clasificador.

						Clase CN			Clase MCI				F1 Sc
Add	Add%	TN	FP	FN	TP	Prec	Recl	F1 Sc	Prec	Recl	F1 Sc	Accur	AVG
0	15.19	32	4	8	2	80.00	88.89	84.21	33.33	20.00	25.00	73.91	54.61
1	16.46	32.1	3.9	8.5	1.5	79.07	89.17	83.80	27.69	15.00	19.24	73.04	51.52
2	17.72	32.2	3.8	8.7	1.3	78.73	89.44	83.73	25.67	13.00	17.06	72.83	50.40
3	18.99	31.8	4.2	8.6	1.4	78.72	88.33	83.23	24.86	14.00	17.69	72.17	50.46
4	20.25	31.9	4.1	8.6	1.4	78.77	88.61	83.39	25.40	14.00	17.84	72.39	50.62
5	21.52	32.5	3.5	8.5	1.5	79.27	90.28	84.39	31.71	15.00	19.77	73.91	52.08
6	22.78	32.8	3.2	8.6	1.4	79.23	91.11	84.74	32.00	14.00	19.06	74.35	51.90
7	24.05	32.7	3.3	8.5	1.5	79.37	90.83	84.70	32.17	15.00	20.10	74.35	52.40
8	25.32	32.9	3.1	8.7	1.3	79.08	91.39	84.78	31.00	13.00	17.99	74.35	51.38
9	26.58	33.3	2.7	8.6	1.4	79.47	92.50	85.48	35.83	14.00	19.80	75.43	52.64
10	27.85	33.2	2.8	8.6	1.4	79.42	92.22	85.33	34.83	14.00	19.61	75.22	52.47
11	29.11	33.3	2.7	8.6	1.4	79.47	92.50	85.47	34.83	14.00	19.61	75.43	52.54
12	30.38	33.1	2.9	8.8	1.2	79.00	91.94	84.96	29.00	12.00	16.65	74.57	50.81
13	31.65	32.9	3.1	8.5	1.5	79.48	91.39	84.99	33.79	15.00	20.20	74.78	52.60
14	32.91	32.7	3.3	8.7	1.3	79.01	90.83	84.47	26.88	13.00	17.05	73.91	50.76
15	34.18	32.7	3.3	8.7	1.3	79.01	90.83	84.47	26.88	13.00	17.05	73.91	50.76
16	35.44	32.9	3.1	8.7	1.3	79.11	91.39	84.77	28.21	13.00	17.32	74.35	51.05
17	36.71	32.9	3.1	8.5	1.5	79.48	91.39	84.99	32.21	15.00	19.99	74.78	52.49
18	37.97	33	3	8.5	1.5	79.53	91.67	85.14	33.21	15.00	20.18	75.00	52.66
19	39.24	32.6	3.4	8.5	1.5	79.33	90.56	84.53	30.40	15.00	19.51	74.13	52.02
20	40.51	32.7	3.3	8.4	1.6	79.61	90.83	84.81	31.36	16.00	20.50	74.57	52.65
21	41.77	32.7	3.3	8.4	1.6	79.61	90.83	84.81	31.36	16.00	20.50	74.57	52.65
22	43.04	33	3	8.4	1.6	79.76	91.67	85.25	32.50	16.00	20.81	75.22	53.03
23	44.30	33.4	2.6	8.4	1.6	79.96	92.78	85.85	35.83	16.00	21.23	76.09	53.54
24	45.57	33.3	2.7	8.4	1.6	79.91	92.50	85.70	35.83	16.00	21.23	75.87	53.47
25	46.84	33.3	2.7	8.3	1.7	80.09	92.50	85.81	39.17	17.00	22.77	76.09	54.29
26	48.10	33.3	2.7	8.3	1.7	80.09	92.50	85.81	39.17	17.00	22.77	76.09	54.29
27	49.37	33.3	2.7	8.3	1.7	80.09	92.50	85.81	39.17	17.00	22.77	76.09	54.29
28	50.63	33.3	2.7	8.2	1.8	80.28	92.50	85.92	40.83	18.00	24.09	76.30	55.00
29	51.90	33.3	2.7	8.2	1.8	80.28	92.50	85.92	40.83	18.00	24.09	76.30	55.00
30	53.16	33.2	2.8	8.2	1.8	80.23	92.22	85.77	40.17	18.00	23.92	76.09	54.85
31	54.43	33.1	2.9	8.2	1.8	80.18	91.94	85.62	39.17	18.00	23.73	75.87	54.68
32	55.70	33.1	2.9	8.2	1.8	80.18	91.94	85.62	39.17	18.00	23.73	75.87	54.68
33	56.96	33.1	2.9	8.2	1.8	80.18	91.94	85.62	39.17	18.00	23.73	75.87	54.68
34	58.23	33.1	2.9	8.2	1.8	80.18	91.94	85.62	39.17	18.00	23.73	75.87	54.68

Continua en la siguiente página

Tabla A.10 – Continuación de la página anterior

						Clase CN			Clase MCI				F1 Sc
Add	Add%	TN	FP	FN	TP	Prec	Recl	F1 Sc	Prec	Recl	F1 Sc	Accur	AVG
35	59.49	33.2	2.8	8.2	1.8	80.24	92.22	85.77	39.67	18.00	23.83	76.09	54.80
36	60.76	33.1	2.9	8.1	1.9	80.38	91.94	85.74	39.67	19.00	25.02	76.09	55.38
37	62.03	33.1	2.9	8.1	1.9	80.38	91.94	85.74	39.67	19.00	25.02	76.09	55.38
38	63.29	33.5	2.5	8.1	1.9	80.57	93.06	86.34	43.00	19.00	25.73	76.96	56.03
39	64.56	33.5	2.5	8.1	1.9	80.57	93.06	86.34	43.00	19.00	25.73	76.96	56.03
40	65.82	33.5	2.5	8.2	1.8	80.39	93.06	86.23	40.50	18.00	24.30	76.74	55.27
41	67.09	33.6	2.4	8.3	1.7	80.24	93.33	86.27	39.83	17.00	23.17	76.74	54.72
42	68.35	33.6	2.4	8.3	1.7	80.24	93.33	86.27	39.83	17.00	23.17	76.74	54.72
43	69.62	33.6	2.4	8.3	1.7	80.24	93.33	86.27	39.83	17.00	23.17	76.74	54.72
44	70.89	33.6	2.4	8.3	1.7	80.24	93.33	86.27	39.83	17.00	23.17	76.74	54.72
45	72.15	33.7	2.3	8.2	1.8	80.48	93.61	86.52	42.33	18.00	24.60	77.17	55.56
46	73.42	33.7	2.3	8.1	1.9	80.66	93.61	86.63	44.83	19.00	26.03	77.39	56.33
47	74.68	33.6	2.4	8.1	1.9	80.62	93.33	86.48	44.17	19.00	25.86	77.17	56.17
48	75.95	33.5	2.5	8.1	1.9	80.57	93.06	86.33	43.17	19.00	25.67	76.96	56.00
49	77.22	33.3	2.7	8.1	1.9	80.47	92.50	86.04	41.50	19.00	25.32	76.52	55.68
50	78.48	33.2	2.8	8	2	80.62	92.22	86.00	42.17	20.00	26.44	76.52	56.22
51	79.75	33.1	2.9	7.9	2.1	80.77	91.94	85.97	41.33	21.00	27.32	76.52	56.64
52	81.01	33.2	2.8	7.9	2.1	80.82	92.22	86.12	42.33	21.00	27.51	76.74	56.81
53	82.28	33.1	2.9	7.8	2.2	80.97	91.94	86.08	42.33	22.00	28.40	76.74	57.24
54	83.54	33.1	2.9	7.8	2.2	80.97	91.94	86.08	42.33	22.00	28.40	76.74	57.24
55	84.81	33.2	2.8	7.8	2.2	81.02	92.22	86.23	43.33	22.00	28.59	76.96	57.41
56	86.08	33.2	2.8	7.8	2.2	81.02	92.22	86.23	43.33	22.00	28.59	76.96	57.41
57	87.34	33	3	7.7	2.3	81.15	91.67	86.04	42.78	23.00	29.05	76.74	57.55
58	88.61	33	3	7.7	2.3	81.15	91.67	86.04	42.78	23.00	29.05	76.74	57.55
59	89.87	33	3	7.7	2.3	81.15	91.67	86.04	42.78	23.00	29.05	76.74	57.55
60	91.14	33	3	7.7	2.3	81.15	91.67	86.04	42.78	23.00	29.05	76.74	57.55
61	92.41	33.1	2.9	7.8	2.2	81.01	91.94	86.07	42.78	22.00	27.86	76.74	56.97
62	93.67	32.9	3.1	7.6	2.4	81.34	91.39	86.00	42.22	24.00	29.45	76.74	57.72
63	94.94	32.9	3.1	7.6	2.4	81.34	91.39	86.00	42.22	24.00	29.45	76.74	57.72
64	96.20	32.9	3.1	7.6	2.4	81.34	91.39	86.00	42.22	24.00	29.45	76.74	57.72
65	97.47	32.8	3.2	7.5	2.5	81.48	91.11	85.96	42.89	25.00	30.57	76.74	58.27
66	98.73	32.8	3.2	7.5	2.5	81.48	91.11	85.96	42.89	25.00	30.57	76.74	58.27
67	100.00	32.8	3.2	7.5	2.5	81.48	91.11	85.96	42.89	25.00	30.57	76.74	58.27
68	101.27	32.8	3.2	7.4	2.6	81.69	91.11	86.08	43.60	26.00	31.53	76.96	58.80
69	102.53	32.8	3.2	7.4	2.6	81.69	91.11	86.08	43.60	26.00	31.53	76.96	58.80
70	103.80	32.8	3.2	7.4	2.6	81.69	91.11	86.08	43.60	26.00	31.53	76.96	58.80
71	105.06	32.8	3.2	7.4	2.6	81.69	91.11	86.08	43.60	26.00	31.53	76.96	58.80
72	106.33	32.8	3.2	7.4	2.6	81.69	91.11	86.08	43.60	26.00	31.53	76.96	58.80
73	107.59	32.8	3.2	7.4	2.6	81.69	91.11	86.08	43.60	26.00	31.53	76.96	58.80

Continua en la siguiente página

Tabla A.10 – Continuación de la página anterior

						Clase CN			Clase MCI				F1 Sc
Add	Add%	TN	FP	FN	TP	Prec	Recl	F1 Sc	Prec	Recl	F1 Sc	Accur	AVG
74	108.86	32.8	3.2	7.4	2.6	81.69	91.11	86.08	43.60	26.00	31.53	76.96	58.80
75	110.13	32.8	3.2	7.4	2.6	81.69	91.11	86.08	43.60	26.00	31.53	76.96	58.80
76	111.39	32.8	3.2	7.4	2.6	81.69	91.11	86.08	43.60	26.00	31.53	76.96	58.80
77	112.66	32.8	3.2	7.3	2.7	81.88	91.11	86.19	44.94	27.00	32.82	77.17	59.51
78	113.92	32.8	3.2	7.3	2.7	81.88	91.11	86.19	44.94	27.00	32.82	77.17	59.51
79	115.19	32.8	3.2	7.3	2.7	81.88	91.11	86.19	44.94	27.00	32.82	77.17	59.51
80	116.46	32.8	3.2	7.3	2.7	81.88	91.11	86.19	44.94	27.00	32.82	77.17	59.51
81	117.72	32.8	3.2	7.2	2.8	82.08	91.11	86.31	45.94	28.00	33.91	77.39	60.11
82	118.99	32.8	3.2	7.2	2.8	82.08	91.11	86.31	45.94	28.00	33.91	77.39	60.11
83	120.25	32.8	3.2	7.2	2.8	82.08	91.11	86.31	45.94	28.00	33.91	77.39	60.11
84	121.52	32.5	3.5	7.1	2.9	82.17	90.28	85.95	44.94	29.00	34.16	76.96	60.06
85	122.78	32.5	3.5	7.1	2.9	82.17	90.28	85.95	44.94	29.00	34.16	76.96	60.06
86	124.05	32.4	3.6	7.1	2.9	82.12	90.00	85.81	43.94	29.00	33.97	76.74	59.89
87	125.32	32.4	3.6	7.1	2.9	82.12	90.00	85.81	43.94	29.00	33.97	76.74	59.89
88	126.58	32.4	3.6	7.1	2.9	82.12	90.00	85.81	43.83	29.00	33.94	76.74	59.88
89	127.85	32.3	3.7	7.1	2.9	82.08	89.72	85.65	43.38	29.00	33.73	76.52	59.69
90	129.11	32.1	3.9	6.9	3.1	82.41	89.17	85.57	43.83	31.00	35.28	76.52	60.43
91	130.38	32	4	6.9	3.1	82.36	88.89	85.43	42.83	31.00	35.09	76.30	60.26
92	131.65	32	4	6.9	3.1	82.36	88.89	85.43	42.83	31.00	35.09	76.30	60.26
93	132.91	31.9	4.1	6.9	3.1	82.31	88.61	85.28	42.16	31.00	34.92	76.09	60.10
94	134.18	32	4	6.9	3.1	82.36	88.89	85.42	43.16	31.00	35.11	76.30	60.27
95	135.44	32	4	7	3	82.17	88.89	85.31	41.49	30.00	33.79	76.09	59.55
96	136.71	31.9	4.1	7	3	82.13	88.61	85.15	41.05	30.00	33.58	75.87	59.37
97	137.97	31.9	4.1	7	3	82.13	88.61	85.15	41.05	30.00	33.58	75.87	59.37
98	139.24	31.6	4.4	7.1	2.9	81.77	87.78	84.58	38.48	29.00	32.09	75.00	58.34
99	140.51	31.6	4.4	7.1	2.9	81.77	87.78	84.58	38.48	29.00	32.09	75.00	58.34
100	141.77	31.7	4.3	7.1	2.9	81.82	88.06	84.73	39.14	29.00	32.26	75.22	58.50
101	143.04	31.7	4.3	7.1	2.9	81.82	88.06	84.73	39.14	29.00	32.26	75.22	58.50
102	144.30	31.7	4.3	7.1	2.9	81.82	88.06	84.73	39.14	29.00	32.26	75.22	58.50
103	145.57	31.7	4.3	7.1	2.9	81.82	88.06	84.73	39.14	29.00	32.26	75.22	58.50
104	146.84	31.7	4.3	7.1	2.9	81.82	88.06	84.73	39.14	29.00	32.26	75.22	58.50
105	148.10	31.6	4.4	7.1	2.9	81.77	87.78	84.59	38.31	29.00	32.15	75.00	58.37
106	149.37	31.6	4.4	7.1	2.9	81.77	87.78	84.59	38.31	29.00	32.15	75.00	58.37
107	150.63	31.8	4.2	7.1	2.9	81.88	88.33	84.90	38.88	29.00	32.30	75.43	58.60
108	151.90	31.8	4.2	7.1	2.9	81.88	88.33	84.90	38.88	29.00	32.30	75.43	58.60
109	153.16	31.8	4.2	7.1	2.9	81.88	88.33	84.90	38.88	29.00	32.30	75.43	58.60
110	154.43	31.8	4.2	7.1	2.9	81.88	88.33	84.90	38.88	29.00	32.30	75.43	58.60
111	155.70	31.8	4.2	7.1	2.9	81.88	88.33	84.90	38.88	29.00	32.30	75.43	58.60
112	156.96	31.8	4.2	7	3	82.08	88.33	85.01	40.21	30.00	33.47	75.65	59.24

Continua en la siguiente página

Tabla A.10 – Continuación de la página anterior

						Clase CN			Clase MCI				F1 Sc
Add	Add%	TN	FP	FN	TP	Prec	Recl	F1 Sc	Prec	Recl	F1 Sc	Accur	AVG
113	158.23	31.8	4.2	7	3	82.08	88.33	85.01	40.21	30.00	33.47	75.65	59.24
114	159.49	31.8	4.2	7	3	82.08	88.33	85.01	40.21	30.00	33.47	75.65	59.24
115	160.76	31.8	4.2	7	3	82.08	88.33	85.01	40.21	30.00	33.47	75.65	59.24
116	162.03	31.8	4.2	7	3	82.08	88.33	85.01	40.21	30.00	33.47	75.65	59.24
117	163.29	31.7	4.3	6.8	3.2	82.42	88.06	85.08	41.88	32.00	35.68	75.87	60.38
118	164.56	31.7	4.3	6.8	3.2	82.42	88.06	85.08	41.88	32.00	35.68	75.87	60.38
119	165.82	31.7	4.3	6.8	3.2	82.42	88.06	85.08	41.88	32.00	35.68	75.87	60.38
120	167.09	31.7	4.3	6.8	3.2	82.42	88.06	85.08	41.88	32.00	35.68	75.87	60.38
121	168.35	31.4	4.6	6.6	3.4	82.70	87.22	84.83	42.18	34.00	36.99	75.65	60.91
122	169.62	31.4	4.6	6.6	3.4	82.70	87.22	84.83	42.18	34.00	36.99	75.65	60.91
123	170.89	31.4	4.6	6.6	3.4	82.70	87.22	84.83	42.18	34.00	36.99	75.65	60.91
124	172.15	31.4	4.6	6.6	3.4	82.70	87.22	84.83	42.18	34.00	36.99	75.65	60.91
125	173.42	31.4	4.6	6.6	3.4	82.70	87.22	84.83	42.18	34.00	36.99	75.65	60.91
126	174.68	31.4	4.6	6.6	3.4	82.70	87.22	84.83	42.18	34.00	36.99	75.65	60.91
127	175.95	31.4	4.6	6.6	3.4	82.70	87.22	84.83	42.18	34.00	36.99	75.65	60.91
128	177.22	31.4	4.6	6.6	3.4	82.70	87.22	84.83	42.18	34.00	36.99	75.65	60.91
129	178.48	31.4	4.6	6.6	3.4	82.70	87.22	84.83	42.18	34.00	36.99	75.65	60.91
130	179.75	31.2	4.8	6.4	3.6	83.03	86.67	84.75	42.63	36.00	38.54	75.65	61.65
131	181.01	31.2	4.8	6.4	3.6	83.03	86.67	84.75	42.63	36.00	38.54	75.65	61.65
132	182.28	31	5	6.4	3.6	82.94	86.11	84.44	41.47	36.00	38.10	75.22	61.27
133	183.54	31.1	4.9	6.4	3.6	82.99	86.39	84.59	42.30	36.00	38.21	75.43	61.40
134	184.81	31.1	4.9	6.4	3.6	82.99	86.39	84.59	42.30	36.00	38.21	75.43	61.40
135	186.08	31.1	4.9	6.4	3.6	82.99	86.39	84.59	42.30	36.00	38.21	75.43	61.40
136	187.34	31.1	4.9	6.4	3.6	82.99	86.39	84.59	42.30	36.00	38.21	75.43	61.40
137	188.61	31.1	4.9	6.4	3.6	82.99	86.39	84.59	42.30	36.00	38.21	75.43	61.40
138	189.87	31.1	4.9	6.4	3.6	82.99	86.39	84.59	42.30	36.00	38.21	75.43	61.40
139	191.14	31.1	4.9	6.4	3.6	82.99	86.39	84.59	42.30	36.00	38.21	75.43	61.40
140	192.41	31.1	4.9	6.4	3.6	82.99	86.39	84.59	42.30	36.00	38.21	75.43	61.40
141	193.67	31	5	6.3	3.7	83.14	86.11	84.55	42.97	37.00	39.34	75.43	61.95
142	194.94	31	5	6.3	3.7	83.14	86.11	84.55	42.97	37.00	39.34	75.43	61.95
143	196.20	31	5	6.3	3.7	83.14	86.11	84.55	42.97	37.00	39.34	75.43	61.95
144	197.47	31	5	6.3	3.7	83.14	86.11	84.55	42.97	37.00	39.34	75.43	61.95
145	198.73	31	5	6.3	3.7	83.14	86.11	84.55	42.97	37.00	39.34	75.43	61.95
146	200.00	31	5	6.3	3.7	83.14	86.11	84.55	42.97	37.00	39.34	75.43	61.95
147	201.27	31	5	6.3	3.7	83.14	86.11	84.55	42.97	37.00	39.34	75.43	61.95
148	202.53	31	5	6.3	3.7	83.14	86.11	84.55	42.97	37.00	39.34	75.43	61.95
149	203.80	31	5	6.3	3.7	83.14	86.11	84.55	42.97	37.00	39.34	75.43	61.95
150	205.06	31	5	6.3	3.7	83.14	86.11	84.55	42.97	37.00	39.34	75.43	61.95
151	206.33	31	5	6.3	3.7	83.14	86.11	84.55	42.97	37.00	39.34	75.43	61.95

Continua en la siguiente página

Tabla A.10 – Continuación de la página anterior

						Clase CN			Clase MCI				F1 Sc
Add	Add%	TN	FP	FN	TP	Prec	Recl	F1 Sc	Prec	Recl	F1 Sc	Accur	AVG
152	207.59	31	5	6.3	3.7	83.14	86.11	84.55	42.97	37.00	39.34	75.43	61.95
153	208.86	31	5	6.3	3.7	83.14	86.11	84.55	42.97	37.00	39.34	75.43	61.95
154	210.13	31	5	6.3	3.7	83.14	86.11	84.55	42.97	37.00	39.34	75.43	61.95
155	211.39	31	5	6.3	3.7	83.14	86.11	84.55	42.97	37.00	39.34	75.43	61.95
156	212.66	31	5	6.3	3.7	83.14	86.11	84.55	42.97	37.00	39.34	75.43	61.95
157	213.92	31	5	6.3	3.7	83.14	86.11	84.55	42.97	37.00	39.34	75.43	61.95
158	215.19	31	5	6.3	3.7	83.14	86.11	84.55	42.97	37.00	39.34	75.43	61.95
159	216.46	31	5	6.3	3.7	83.14	86.11	84.55	42.97	37.00	39.34	75.43	61.95

En la Tabla A.10, la información se encuentra distribuida de la siguiente manera; en la primera columna los elementos agregados al conjunto de Train, en la segunda columna, el *Balance Rate* en porcentaje, que surge de dividir el número actual de elementos en el conjunto de Train sobre sobre el número total de elementos objetivos, esto multiplicado por 100. Las columnas 3, 4, 5 y 6, corresponden a las variables de la matriz de confusión, TP representa a los aciertos del clasificador en la clase de interés y TN corresponden a los aciertos del clasificador en la clase de control "CN"; luego encontramos las métricas de la matriz de clasificación por clase (Precision, Recall y F1 Score), en la columnas 7, 8 y 9, las correspondientes a la clase CN y en las columnas 10, 11 y 12 la clase de interés; ya por último encontramos la columna del Accuracy y el F1 Average, respectivamente.

Tabla A.11: Evaluación de la clasificación Test SubADReSSo, utilizando la técnica de TimeStretch, mientras se añaden elementos a la clase MCI.

						Clase CN			Clase MCI				F1 Sc
Add	Add%	TN	FP	FN	TP	Prec	Recl	F1 Sc	Prec	Recl	F1 Sc	Accur	AVG
0	15.19	32	4	8	2	80.00	88.89	84.21	33.33	20.00	25.00	73.91	54.61
1	16.46	31.2	4.8	7.9	2.1	79.79	86.67	83.08	30.65	21.00	24.84	72.39	53.96
2	17.72	31	5	7.9	2.1	79.69	86.11	82.77	29.70	21.00	24.54	71.96	53.66
3	18.99	31.1	4.9	7.9	2.1	79.74	86.39	82.92	30.18	21.00	24.69	72.17	53.81
4	20.25	31.1	4.9	7.9	2.1	79.74	86.39	82.93	30.06	21.00	24.67	72.17	53.80
5	21.52	31.4	4.6	7.9	2.1	79.89	87.22	83.38	32.20	21.00	25.18	72.83	54.28
6	22.78	31.5	4.5	7.9	2.1	79.94	87.50	83.52	33.81	21.00	25.35	73.04	54.44
7	24.05	31.2	4.8	8	2	79.56	86.67	82.93	31.57	20.00	24.00	72.17	53.46
8	25.32	31.1	4.9	7.9	2.1	79.71	86.39	82.87	32.32	21.00	24.92	72.17	53.90
9	26.58	31.1	4.9	7.9	2.1	79.69	86.39	82.84	33.35	21.00	25.00	72.17	53.92
10	27.85	31.3	4.7	8	2	79.60	86.94	83.06	33.01	20.00	24.25	72.39	53.65
11	29.11	31.2	4.8	8	2	79.52	86.67	82.87	33.17	20.00	24.25	72.17	53.56
12	30.38	31.4	4.6	7.9	2.1	79.84	87.22	83.30	34.75	21.00	25.45	72.83	54.37
13	31.65	31.5	4.5	7.8	2.2	80.12	87.50	83.61	35.16	22.00	26.54	73.26	55.07
14	32.91	31	5	7.6	2.4	80.29	86.11	83.06	33.63	24.00	27.72	72.61	55.39
15	34.18	30.8	5.2	7.6	2.4	80.18	85.56	82.73	32.98	24.00	27.37	72.17	55.05
16	35.44	31.1	4.9	7.7	2.3	80.13	86.39	83.10	33.31	23.00	26.87	72.61	54.98
17	36.71	31.1	4.9	7.7	2.3	80.12	86.39	83.10	33.37	23.00	26.90	72.61	55.00
18	37.97	31	5	7.7	2.3	80.08	86.11	82.96	32.44	23.00	26.70	72.39	54.83
19	39.24	30.9	5.1	7.7	2.3	80.02	85.83	82.79	32.22	23.00	26.59	72.17	54.69
20	40.51	30.8	5.2	7.7	2.3	79.97	85.56	82.63	31.87	23.00	26.46	71.96	54.55
21	41.77	31	5	7.7	2.3	80.08	86.11	82.96	32.48	23.00	26.73	72.39	54.85
22	43.04	31	5	7.7	2.3	80.07	86.11	82.97	32.48	23.00	26.81	72.39	54.89
23	44.30	30.9	5.1	7.7	2.3	80.02	85.83	82.81	32.00	23.00	26.67	72.17	54.74
24	45.57	30.8	5.2	7.8	2.2	79.77	85.56	82.55	30.33	22.00	25.42	71.74	53.98
25	46.84	30.6	5.4	7.7	2.3	79.88	85.00	82.34	30.41	23.00	26.08	71.52	54.21
26	48.10	30.7	5.3	7.5	2.5	80.35	85.28	82.72	32.45	25.00	28.09	72.17	55.41
27	49.37	30.8	5.2	7.5	2.5	80.36	85.56	82.84	34.29	25.00	28.68	72.39	55.76
28	50.63	30.6	5.4	7.4	2.6	80.48	85.00	82.63	34.19	26.00	29.21	72.17	55.92
29	51.90	30.9	5.1	7.2	2.8	81.10	85.83	83.38	36.04	28.00	31.29	73.26	57.33
30	53.16	31.1	4.9	7.2	2.8	81.19	86.39	83.69	37.17	28.00	31.68	73.70	57.68
31	54.43	31	5	7.2	2.8	81.15	86.11	83.53	36.46	28.00	31.46	73.48	57.50
32	55.70	30.7	5.3	7.2	2.8	81.01	85.28	83.06	35.14	28.00	30.90	72.83	56.98
33	56.96	30.7	5.3	7.2	2.8	81.01	85.28	83.06	35.14	28.00	30.90	72.83	56.98
34	58.23	30.8	5.2	7.2	2.8	81.06	85.56	83.22	35.42	28.00	31.02	73.04	57.12
35	59.49	30.6	5.4	7.2	2.8	80.96	85.00	82.90	34.61	28.00	30.70	72.61	56.80
36	60.76	30.8	5.2	7.2	2.8	81.06	85.56	83.21	35.69	28.00	30.97	73.04	57.09

Continua en la siguiente página

Tabla A.11 – Continuación de la página anterior

						Clase CN			Clase MCI				F1 Sc
Add	Add%	TN	FP	FN	TP	Prec	Recl	F1 Sc	Prec	Recl	F1 Sc	Accur	AVG
37	62.03	30.8	5.2	7.2	2.8	81.06	85.56	83.21	35.69	28.00	30.97	73.04	57.09
38	63.29	30.8	5.2	7.2	2.8	81.06	85.56	83.21	35.69	28.00	30.97	73.04	57.09
39	64.56	30.7	5.3	7.1	2.9	81.21	85.28	83.16	36.28	29.00	31.88	73.04	57.52
40	65.82	30.8	5.2	6.9	3.1	81.68	85.56	83.56	38.21	31.00	34.07	73.70	58.81
41	67.09	30.9	5.1	6.9	3.1	81.73	85.83	83.72	38.49	31.00	34.19	73.91	58.95
42	68.35	30.8	5.2	6.9	3.1	81.68	85.56	83.56	37.98	31.00	34.00	73.70	58.78
43	69.62	30.8	5.2	6.9	3.1	81.68	85.56	83.56	37.98	31.00	34.00	73.70	58.78
44	70.89	31	5	6.7	3.3	82.20	86.11	84.09	40.77	33.00	36.29	74.57	60.19
45	72.15	31	5	6.7	3.3	82.20	86.11	84.09	40.77	33.00	36.29	74.57	60.19
46	73.42	30.9	5.1	6.7	3.3	82.15	85.83	83.93	40.24	33.00	36.09	74.35	60.01
47	74.68	30.9	5.1	6.8	3.2	81.93	85.83	83.82	39.57	32.00	35.25	74.13	59.53
48	75.95	30.6	5.4	6.8	3.2	81.78	85.00	83.34	38.41	32.00	34.74	73.48	59.04
49	77.22	30.5	5.5	6.8	3.2	81.73	84.72	83.18	37.99	32.00	34.57	73.26	58.87
50	78.48	30.4	5.6	6.8	3.2	81.69	84.44	83.02	37.27	32.00	34.31	73.04	58.67
51	79.75	30.4	5.6	6.8	3.2	81.69	84.44	83.02	37.27	32.00	34.31	73.04	58.67
52	81.01	30.5	5.5	6.7	3.3	81.96	84.72	83.30	38.38	33.00	35.38	73.48	59.34
53	82.28	30.1	5.9	6.7	3.3	81.76	83.61	82.65	36.84	33.00	34.65	72.61	58.65
54	83.54	30	6	6.7	3.3	81.70	83.33	82.48	36.61	33.00	34.52	72.39	58.50
55	84.81	29.9	6.1	6.7	3.3	81.65	83.06	82.32	36.05	33.00	34.29	72.17	58.31
56	86.08	29.9	6.1	6.7	3.3	81.65	83.06	82.32	36.05	33.00	34.29	72.17	58.31
57	87.34	29.8	6.2	6.7	3.3	81.61	82.78	82.17	35.34	33.00	34.03	71.96	58.10
58	88.61	29.8	6.2	6.6	3.4	81.84	82.78	82.29	36.01	34.00	34.87	72.17	58.58
59	89.87	29.9	6.1	6.5	3.5	82.11	83.06	82.56	37.12	35.00	35.92	72.61	59.24
60	91.14	29.9	6.1	6.5	3.5	82.10	83.06	82.55	37.34	35.00	36.00	72.61	59.28
61	92.41	29.7	6.3	6.5	3.5	81.99	82.50	82.22	36.73	35.00	35.70	72.17	58.96
62	93.67	29.7	6.3	6.5	3.5	81.99	82.50	82.22	36.73	35.00	35.70	72.17	58.96
63	94.94	29.5	6.5	6.5	3.5	81.91	81.94	81.91	35.62	35.00	35.23	71.74	58.57
64	96.20	29.7	6.3	6.5	3.5	82.03	82.50	82.25	36.06	35.00	35.46	72.17	58.86
65	97.47	29.7	6.3	6.4	3.6	82.25	82.50	82.36	36.72	36.00	36.31	72.39	59.34
66	98.73	29.8	6.2	6.3	3.7	82.52	82.78	82.63	37.94	37.00	37.38	72.83	60.01
67	100.00	29.8	6.2	6.3	3.7	82.52	82.78	82.63	37.94	37.00	37.38	72.83	60.01
68	101.27	29.9	6.1	6.5	3.5	82.12	83.06	82.56	37.03	35.00	35.87	72.61	59.22
69	102.53	29.8	6.2	6.7	3.3	81.62	82.78	82.17	35.31	33.00	33.96	71.96	58.07
70	103.80	29.7	6.3	6.8	3.2	81.35	82.50	81.90	34.20	32.00	32.91	71.52	57.40
71	105.06	29.6	6.4	6.8	3.2	81.30	82.22	81.74	33.79	32.00	32.73	71.30	57.23
72	106.33	29.5	6.5	6.8	3.2	81.26	81.94	81.58	33.25	32.00	32.53	71.09	57.06
73	107.59	29.3	6.7	6.8	3.2	81.15	81.39	81.26	32.50	32.00	32.20	70.65	56.73
74	108.86	29.3	6.7	6.8	3.2	81.15	81.39	81.26	32.50	32.00	32.20	70.65	56.73
75	110.13	29.3	6.7	6.8	3.2	81.15	81.39	81.26	32.50	32.00	32.20	70.65	56.73

Continua en la siguiente página

Tabla A.11 – Continuación de la página anterior

						Clase CN			Clase MCI				F1 Sc
Add	Add%	TN	FP	FN	TP	Prec	Recl	F1 Sc	Prec	Recl	F1 Sc	Accur	AVG
76	111.39	29.3	6.7	6.8	3.2	81.15	81.39	81.26	32.50	32.00	32.20	70.65	56.73
77	112.66	29.3	6.7	6.8	3.2	81.15	81.39	81.26	32.50	32.00	32.20	70.65	56.73
78	113.92	29.2	6.8	6.9	3.1	80.88	81.11	80.98	31.47	31.00	31.17	70.22	56.08
79	115.19	29.2	6.8	6.9	3.1	80.88	81.11	80.98	31.47	31.00	31.17	70.22	56.08
80	116.46	29.1	6.9	6.9	3.1	80.83	80.83	80.82	31.14	31.00	31.01	70.00	55.91
81	117.72	29.1	6.9	6.9	3.1	80.83	80.83	80.82	31.14	31.00	31.01	70.00	55.91
82	118.99	29	7	6.9	3.1	80.78	80.56	80.65	30.80	31.00	30.85	69.78	55.75
83	120.25	29	7	6.9	3.1	80.78	80.56	80.65	30.80	31.00	30.85	69.78	55.75
84	121.52	28.9	7.1	6.9	3.1	80.72	80.28	80.49	30.47	31.00	30.69	69.57	55.59
85	122.78	28.9	7.1	6.9	3.1	80.73	80.28	80.50	30.36	31.00	30.67	69.57	55.58
86	124.05	29	7	7	3	80.56	80.56	80.56	30.00	30.00	30.00	69.57	55.28
87	125.32	29	7	7	3	80.56	80.56	80.56	30.00	30.00	30.00	69.57	55.28

En la Tabla A.11, la información se encuentra distribuida de la siguiente manera; en la primera columna los elementos agregados al conjunto de Train, en la segunda columna, el *Balance Rate* en porcentaje, que surge de dividir el número actual de elementos en el conjunto de Train sobre sobre el número total de elementos objetivos, esto multiplicado por 100. Las columnas 3, 4, 5 y 6, corresponden a las variables de la matriz de confusión, TP representa a los aciertos del clasificador en la clase de interés y TN corresponden a los aciertos del clasificador en la clase de control "CN"; luego encontramos las métricas de la matriz de clasificación por clase (Precision, Recall y F1 Score), en la columnas 7, 8 y 9, las correspondientes a la clase CN y en las columnas 10, 11 y 12 la clase de interés; ya por último encontramos la columna del Accuracy y el F1 Average, respectivamente.

Tabla A.12: Evaluación de la clasificación Test SubADReSSo, utilizando la técnica de Shift, mientras se añaden elementos a la clase MCI.

						Clase CN			Clase MCI				F1 Sc
Add	Add%	TN	FP	FN	TP	Prec	Recl	F1 Sc	Prec	Recl	F1 Sc	Accur	AVG
0	15.19	32	4	8	2	80.00	88.89	84.21	33.33	20.00	25.00	73.91	54.61
1	16.46	31.8	4.2	8	2	79.89	88.33	83.90	32.57	20.00	24.73	73.48	54.31
2	17.72	31.5	4.5	8	2	79.74	87.50	83.43	31.26	20.00	24.30	72.83	53.87
3	18.99	31.7	4.3	8	2	79.84	88.06	83.74	32.40	20.00	24.61	73.26	54.17
4	20.25	31.6	4.4	8	2	79.79	87.78	83.58	31.93	20.00	24.47	73.04	54.02
5	21.52	31.2	4.8	8.1	1.9	79.36	86.67	82.83	29.52	19.00	23.00	71.96	52.92
6	22.78	31.2	4.8	8.1	1.9	79.37	86.67	82.83	29.28	19.00	22.83	71.96	52.83
7	24.05	31.8	4.2	8	2	79.86	88.33	83.85	34.16	20.00	24.99	73.48	54.42
8	25.32	31.1	4.9	7.8	2.2	79.92	86.39	83.01	32.02	22.00	25.92	72.39	54.46
9	26.58	31.2	4.8	7.7	2.3	80.19	86.67	83.28	33.27	23.00	27.03	72.83	55.16
10	27.85	31.3	4.7	7.6	2.4	80.45	86.94	83.55	34.52	24.00	28.14	73.26	55.85
11	29.11	31.3	4.7	7.6	2.4	80.45	86.94	83.55	34.52	24.00	28.14	73.26	55.85
12	30.38	32.5	3.5	7.4	2.6	81.43	90.28	85.57	47.41	26.00	32.43	76.30	59.00
13	31.65	31.9	4.1	7.5	2.5	80.87	88.61	84.47	44.99	25.00	30.81	74.78	57.64
14	32.91	32.3	3.7	7.5	2.5	81.06	89.72	85.08	49.61	25.00	31.59	75.65	58.34
15	34.18	32.1	3.9	7.1	2.9	81.84	89.17	85.24	51.15	29.00	34.82	76.09	60.03
16	35.44	33	3	7.1	2.9	82.36	91.67	86.67	56.07	29.00	35.73	78.04	61.20
17	36.71	32.9	3.1	6.9	3.1	82.72	91.39	86.75	56.07	31.00	37.66	78.26	62.20
18	37.97	32.9	3.1	6.8	3.2	82.94	91.39	86.87	56.79	32.00	38.57	78.48	62.72
19	39.24	32.9	3.1	6.8	3.2	82.94	91.39	86.87	56.79	32.00	38.57	78.48	62.72
20	40.51	32	4	6.6	3.4	82.99	88.89	85.74	48.62	34.00	38.33	76.96	62.04
21	41.77	31.4	4.6	6.4	3.6	83.12	87.22	85.03	46.03	36.00	39.18	76.09	62.10
22	43.04	31.4	4.6	6.4	3.6	83.12	87.22	85.03	46.03	36.00	39.18	76.09	62.10
23	44.30	31.2	4.8	6.4	3.6	83.03	86.67	84.70	45.53	36.00	38.73	75.65	61.72
24	45.57	31	5	6.3	3.7	83.14	86.11	84.45	45.67	37.00	39.27	75.43	61.86
25	46.84	30.6	5.4	6.2	3.8	83.13	85.00	83.95	43.25	38.00	39.75	74.78	61.85
26	48.10	31	5	6.3	3.7	83.10	86.11	84.48	44.20	37.00	39.65	75.43	62.06
27	49.37	30.9	5.1	6.3	3.7	83.05	85.83	84.33	43.65	37.00	39.41	75.22	61.87
28	50.63	31	5	6.3	3.7	83.12	86.11	84.50	44.09	37.00	39.49	75.43	62.00
29	51.90	30.2	5.8	6.3	3.7	82.75	83.89	83.17	41.13	37.00	37.96	73.70	60.57
30	53.16	30.2	5.8	6.1	3.9	83.16	83.89	83.40	43.13	39.00	39.99	74.13	61.69
31	54.43	30.3	5.7	6	4	83.44	84.17	83.68	44.02	40.00	40.96	74.57	62.32
32	55.70	30.3	5.7	5.9	4.1	83.65	84.17	83.73	47.94	41.00	42.17	74.78	62.95
33	56.96	30	6	5.8	4.2	83.82	83.33	83.40	44.57	42.00	41.83	74.35	62.62
34	58.23	30.1	5.9	5.8	4.2	83.87	83.61	83.56	45.01	42.00	42.04	74.57	62.80
35	59.49	30.1	5.9	5.8	4.2	83.87	83.61	83.54	45.22	42.00	42.06	74.57	62.80
36	60.76	30	6	5.8	4.2	83.82	83.33	83.38	44.78	42.00	41.85	74.35	62.62

Continua en la siguiente página

Tabla A.12 – Continuación de la página anterior

						Clase CN			Clase MCI				F1 Sc
Add	Add%	TN	FP	FN	TP	Prec	Recl	F1 Sc	Prec	Recl	F1 Sc	Accur	AVG
37	62.03	30	6	5.7	4.3	84.04	83.33	83.52	45.14	43.00	42.73	74.57	63.13
38	63.29	30	6	5.7	4.3	84.04	83.33	83.52	45.14	43.00	42.73	74.57	63.13
39	64.56	29.7	6.3	5.6	4.4	84.12	82.50	83.12	44.56	44.00	42.94	74.13	63.03
40	65.82	29	7	5.7	4.3	83.58	80.56	81.90	39.41	43.00	40.49	72.39	61.20
41	67.09	29	7	5.7	4.3	83.58	80.56	81.90	39.41	43.00	40.49	72.39	61.20
42	68.35	28.6	7.4	5.7	4.3	83.38	79.44	81.23	38.08	43.00	39.75	71.52	60.49
43	69.62	28.7	7.3	5.7	4.3	83.42	79.72	81.38	38.63	43.00	39.98	71.74	60.68
44	70.89	28.8	7.2	5.6	4.4	83.70	80.00	81.66	39.74	44.00	41.01	72.17	61.33
45	72.15	28.6	7.4	5.6	4.4	83.61	79.44	81.34	38.74	44.00	40.57	71.74	60.95
46	73.42	28.6	7.4	5.6	4.4	83.61	79.44	81.34	38.74	44.00	40.57	71.74	60.95
47	74.68	28	8	5.6	4.4	83.34	77.78	80.33	36.56	44.00	39.34	70.43	59.83
48	75.95	28.2	7.8	5.7	4.3	83.16	78.33	80.56	36.73	43.00	39.08	70.65	59.82
49	77.22	28.1	7.9	5.7	4.3	83.13	78.06	80.40	36.32	43.00	38.84	70.43	59.62
50	78.48	27.6	8.4	5.6	4.4	83.13	76.67	79.70	34.80	44.00	38.61	69.57	59.16
51	79.75	27.2	8.8	5.5	4.5	83.17	75.56	79.14	34.22	45.00	38.72	68.91	58.93
52	81.01	27.3	8.7	5.5	4.5	83.22	75.83	79.32	34.46	45.00	38.89	69.13	59.10
53	82.28	27.2	8.8	5.5	4.5	83.17	75.56	79.15	34.10	45.00	38.69	68.91	58.92
54	83.54	27.2	8.8	5.5	4.5	83.17	75.56	79.15	34.10	45.00	38.69	68.91	58.92
55	84.81	27.2	8.8	5.5	4.5	83.17	75.56	79.15	34.10	45.00	38.69	68.91	58.92
56	86.08	26.8	9.2	5.5	4.5	82.97	74.44	78.46	32.96	45.00	37.99	68.04	58.22
57	87.34	26.9	9.1	5.4	4.6	83.30	74.72	78.74	33.68	46.00	38.77	68.48	58.75
58	88.61	27	9	5.5	4.5	83.10	75.00	78.81	33.40	45.00	38.23	68.48	58.52
59	89.87	27	9	5.5	4.5	83.10	75.00	78.81	33.40	45.00	38.23	68.48	58.52
60	91.14	26.9	9.1	5.5	4.5	83.05	74.72	78.63	33.16	45.00	38.06	68.26	58.35
61	92.41	27	9	5.5	4.5	83.09	75.00	78.80	33.44	45.00	38.26	68.48	58.53
62	93.67	27.1	8.9	5.5	4.5	83.14	75.28	78.98	33.68	45.00	38.43	68.70	58.71
63	94.94	26.9	9.1	5.6	4.4	82.78	74.72	78.53	32.63	44.00	37.41	68.04	57.97
64	96.20	26.7	9.3	5.4	4.6	83.19	74.17	78.37	33.31	46.00	38.48	68.04	58.43
65	97.47	26.1	9.9	5.4	4.6	82.88	72.50	77.30	31.80	46.00	37.49	66.74	57.40
66	98.73	26.1	9.9	5.4	4.6	82.88	72.50	77.30	31.80	46.00	37.49	66.74	57.40
67	100.00	26.4	9.6	5.5	4.5	82.78	73.33	77.72	32.01	45.00	37.28	67.17	57.50
68	101.27	26.7	9.3	5.6	4.4	82.66	74.17	78.15	32.31	44.00	37.17	67.61	57.66
69	102.53	27.2	8.8	5.7	4.3	82.66	75.56	78.92	33.07	43.00	37.29	68.48	58.10
70	103.80	26.8	9.2	5.8	4.2	82.18	74.44	78.10	31.59	42.00	36.00	67.39	57.05
71	105.06	26.8	9.2	5.9	4.1	81.93	74.44	77.99	31.10	41.00	35.31	67.17	56.65
72	106.33	26.8	9.2	5.9	4.1	81.93	74.44	77.99	31.10	41.00	35.31	67.17	56.65
73	107.59	26.8	9.2	5.9	4.1	81.93	74.44	77.99	31.10	41.00	35.31	67.17	56.65
74	108.86	26.6	9.4	6	4	81.58	73.89	77.53	30.01	40.00	34.24	66.52	55.89
75	110.13	26.5	9.5	6	4	81.52	73.61	77.35	29.77	40.00	34.10	66.30	55.72

Continua en la siguiente página

Tabla A.12 – Continuación de la página anterior

						Clase CN			Clase MCI				F1 Sc
Add	Add%	TN	FP	FN	TP	Prec	Recl	F1 Sc	Prec	Recl	F1 Sc	Accur	AVG
76	111.39	26.6	9.4	6	4	81.59	73.89	77.54	29.93	40.00	34.22	66.52	55.88
77	112.66	26.6	9.4	6	4	81.59	73.89	77.54	29.93	40.00	34.22	66.52	55.88
78	113.92	26.5	9.5	6	4	81.53	73.61	77.36	29.71	40.00	34.07	66.30	55.72
79	115.19	26.4	9.6	6	4	81.47	73.33	77.18	29.49	40.00	33.93	66.09	55.55
80	116.46	26.2	9.8	6	4	81.36	72.78	76.82	29.05	40.00	33.64	65.65	55.23
81	117.72	26.3	9.7	6	4	81.42	73.06	77.00	29.27	40.00	33.78	65.87	55.39
82	118.99	26.3	9.7	6	4	81.42	73.06	77.00	29.27	40.00	33.78	65.87	55.39
83	120.25	26.3	9.7	6	4	81.42	73.06	77.00	29.27	40.00	33.78	65.87	55.39
84	121.52	26.2	9.8	6	4	81.36	72.78	76.82	29.05	40.00	33.64	65.65	55.23
85	122.78	26.1	9.9	6	4	81.31	72.50	76.65	28.79	40.00	33.48	65.43	55.06
86	124.05	26	10	6	4	81.25	72.22	76.47	28.57	40.00	33.33	65.22	54.90
87	125.32	26	10	6	4	81.25	72.22	76.47	28.57	40.00	33.33	65.22	54.90

En la Tabla A.12, la información se encuentra distribuida de la siguiente manera; en la primera columna los elementos agregados al conjunto de Train, en la segunda columna, el *Balance Rate* en porcentaje, que surge de dividir el número actual de elementos en el conjunto de Train sobre sobre el número total de elementos objetivos, esto multiplicado por 100. Las columnas 3, 4, 5 y 6, corresponden a las variables de la matriz de confusión, TP representa a los aciertos del clasificador en la clase de interés y TN corresponden a los aciertos del clasificador en la clase de control "CN"; luego encontramos las métricas de la matriz de clasificación por clase (Precision, Recall y F1 Score), en la columnas 7, 8 y 9, las correspondientes a la clase CN y en las columnas 10, 11 y 12 la clase de interés; ya por último encontramos la columna del Accuracy y el F1 Average, respectivamente.

Tabla A.13: Evaluación de la clasificación Test SubADReSSo, utilizando la técnica de Pitch, mientras se añaden elementos a la clase MCI.

						Clase CN			Clase MCI				F1 Sc
Add	Add%	TN	FP	FN	TP	Prec	Recl	F1 Sc	Prec	Recl	F1 Sc	Accur	AVG
0	15.19	32	4	8	2	80.00	88.89	84.21	33.33	20.00	25.00	73.91	54.61
1	16.46	32	4	7.9	2.1	80.21	88.89	84.32	34.29	21.00	26.03	74.13	55.18
2	17.72	32.3	3.7	8.2	1.8	79.77	89.72	84.43	33.31	18.00	22.81	74.13	53.62
3	18.99	32.1	3.9	8.3	1.7	79.47	89.17	84.01	31.29	17.00	21.43	73.48	52.72
4	20.25	31.8	4.2	8.2	1.8	79.52	88.33	83.64	30.98	18.00	21.88	73.04	52.76
5	21.52	31.9	4.1	8	2	79.96	88.61	84.02	33.50	20.00	24.53	73.70	54.27
6	22.78	31.9	4.1	7.7	2.3	80.57	88.61	84.34	38.19	23.00	27.81	74.35	56.07
7	24.05	32	4	7.5	2.5	81.04	88.89	84.71	41.44	25.00	29.88	75.00	57.30
8	25.32	31.9	4.1	7.3	2.7	81.46	88.61	84.79	42.13	27.00	31.33	75.22	58.06
9	26.58	31.3	4.7	7	3	81.80	86.94	84.16	41.32	30.00	33.11	74.57	58.63
10	27.85	30.9	5.1	7	3	81.68	85.83	83.48	39.40	30.00	31.72	73.70	57.60
11	29.11	30.6	5.4	7	3	81.51	85.00	82.99	38.53	30.00	31.31	73.04	57.15
12	30.38	30.4	5.6	6.9	3.1	81.60	84.44	82.78	37.69	31.00	32.13	72.83	57.46
13	31.65	29.9	6.1	6.6	3.4	82.02	83.06	82.34	36.57	34.00	33.93	72.39	58.13
14	32.91	29.7	6.3	6.6	3.4	81.86	82.50	81.94	37.18	34.00	33.89	71.96	57.91
15	34.18	29.3	6.7	6.6	3.4	81.59	81.39	81.28	36.02	34.00	33.75	71.09	57.52
16	35.44	29.5	6.5	6.5	3.5	81.85	81.94	81.74	37.80	35.00	35.43	71.74	58.59
17	36.71	29.6	6.4	6.5	3.5	81.96	82.22	81.95	37.33	35.00	35.32	71.96	58.64
18	37.97	29.5	6.5	6.5	3.5	81.91	81.94	81.80	36.85	35.00	35.18	71.74	58.49
19	39.24	29.4	6.6	6.6	3.4	81.62	81.67	81.53	35.76	34.00	34.21	71.30	57.87
20	40.51	29.5	6.5	6.5	3.5	81.95	81.94	81.80	36.57	35.00	34.93	71.74	58.37
21	41.77	29	7	6.5	3.5	81.67	80.56	80.97	35.06	35.00	34.24	70.65	57.60
22	43.04	28.9	7.1	6.4	3.6	81.87	80.28	80.92	35.22	36.00	34.82	70.65	57.87
23	44.30	29.1	6.9	6.6	3.4	81.52	80.83	80.99	34.44	34.00	33.19	70.65	57.09
24	45.57	29	7	6.6	3.4	81.48	80.56	80.85	33.55	34.00	32.86	70.43	56.85
25	46.84	29	7	6.6	3.4	81.48	80.56	80.85	33.55	34.00	32.86	70.43	56.85
26	48.10	29.2	6.8	6.6	3.4	81.61	81.11	81.19	33.95	34.00	33.06	70.87	57.12
27	49.37	28.7	7.3	6.5	3.5	81.54	79.72	80.41	33.45	35.00	33.18	70.00	56.80
28	50.63	28.8	7.2	6.7	3.3	81.15	80.00	80.36	32.41	33.00	31.69	69.78	56.03
29	51.90	28.9	7.1	6.5	3.5	81.64	80.28	80.77	34.37	35.00	33.82	70.43	57.29
30	53.16	28.3	7.7	6.4	3.6	81.64	78.61	79.91	32.18	36.00	33.13	69.35	56.52
31	54.43	27.6	8.4	6.1	3.9	82.04	76.67	79.10	31.49	39.00	34.20	68.48	56.65
32	55.70	27.7	8.3	6.1	3.9	82.09	76.94	79.26	31.90	39.00	34.37	68.70	56.81
33	56.96	27	9	5.9	4.1	82.17	75.00	78.31	31.14	41.00	35.01	67.61	56.66
34	58.23	27	9	5.8	4.2	82.41	75.00	78.42	31.72	42.00	35.76	67.83	57.09
35	59.49	26.9	9.1	5.8	4.2	82.35	74.72	78.25	31.49	42.00	35.65	67.61	56.95
36	60.76	26.7	9.3	5.6	4.4	82.70	74.17	78.12	32.35	44.00	36.99	67.61	57.55

Continua en la siguiente página

Tabla A.13 – Continuación de la página anterior

						Clase CN			Clase MCI				F1 Sc
Add	Add%	TN	FP	FN	TP	Prec	Recl	F1 Sc	Prec	Recl	F1 Sc	Accur	AVG
37	62.03	26.7	9.3	5.6	4.4	82.70	74.17	78.12	32.35	44.00	36.99	67.61	57.55
38	63.29	27	9	5.5	4.5	83.06	75.00	78.74	34.11	45.00	38.44	68.48	58.59
39	64.56	26.7	9.3	5.6	4.4	82.62	74.17	78.08	33.04	44.00	37.35	67.61	57.71
40	65.82	26.4	9.6	5.8	4.2	81.97	73.33	77.35	30.73	42.00	35.30	66.52	56.33
41	67.09	25.8	10.2	5.8	4.2	81.63	71.67	76.31	29.27	42.00	34.47	65.22	55.39
42	68.35	25.9	10.1	5.8	4.2	81.70	71.94	76.50	29.43	42.00	34.59	65.43	55.54
43	69.62	25.6	10.4	5.7	4.3	81.80	71.11	76.06	29.26	43.00	34.78	65.00	55.42
44	70.89	25.6	10.4	5.7	4.3	81.80	71.11	76.06	29.26	43.00	34.78	65.00	55.42
45	72.15	25.8	10.2	5.8	4.2	81.61	71.67	76.29	29.37	42.00	34.52	65.22	55.41
46	73.42	25.7	10.3	5.7	4.3	81.83	71.39	76.21	29.65	43.00	35.02	65.22	55.61
47	74.68	25.7	10.3	5.7	4.3	81.83	71.39	76.21	29.65	43.00	35.02	65.22	55.61
48	75.95	25.8	10.2	5.7	4.3	81.88	71.67	76.38	29.90	43.00	35.17	65.43	55.78
49	77.22	25.7	10.3	5.5	4.5	82.39	71.39	76.45	30.45	45.00	36.23	65.65	56.34
50	78.48	25.5	10.5	5.7	4.3	81.72	70.83	75.84	29.20	43.00	34.69	64.78	55.27
51	79.75	25.4	10.6	5.8	4.2	81.39	70.56	75.55	28.55	42.00	33.92	64.35	54.73
52	81.01	25.5	10.5	5.9	4.1	81.18	70.83	75.63	28.27	41.00	33.42	64.35	54.53
53	82.28	25.3	10.7	5.9	4.1	81.06	70.28	75.27	27.86	41.00	33.14	63.91	54.20
54	83.54	25.3	10.7	5.9	4.1	81.06	70.28	75.27	27.86	41.00	33.14	63.91	54.20
55	84.81	25.2	10.8	5.9	4.1	81.00	70.00	75.08	27.70	41.00	33.02	63.70	54.05
56	86.08	25.3	10.7	6	4	80.80	70.28	75.15	27.39	40.00	32.46	63.70	53.81
57	87.34	25.1	10.9	5.9	4.1	80.95	69.72	74.90	27.44	41.00	32.83	63.48	53.86
58	88.61	25.1	10.9	5.9	4.1	80.95	69.72	74.90	27.44	41.00	32.83	63.48	53.86
59	89.87	24.9	11.1	5.9	4.1	80.83	69.17	74.51	27.11	41.00	32.58	63.04	53.54
60	91.14	24.8	11.2	5.9	4.1	80.77	68.89	74.33	26.92	41.00	32.44	62.83	53.38
61	92.41	24.8	11.2	5.9	4.1	80.77	68.89	74.33	26.92	41.00	32.44	62.83	53.38
62	93.67	24.8	11.2	5.8	4.2	81.05	68.89	74.45	27.33	42.00	33.07	63.04	53.76
63	94.94	24.8	11.2	5.8	4.2	81.04	68.89	74.45	27.35	42.00	33.09	63.04	53.77
64	96.20	24.8	11.2	5.8	4.2	81.04	68.89	74.46	27.33	42.00	33.08	63.04	53.77
65	97.47	24.5	11.5	5.8	4.2	80.85	68.06	73.89	26.79	42.00	32.70	62.39	53.29
66	98.73	24.5	11.5	5.8	4.2	80.85	68.06	73.89	26.79	42.00	32.70	62.39	53.29
67	100.00	24.4	11.6	5.7	4.3	81.07	67.78	73.81	27.05	43.00	33.18	62.39	53.50
68	101.27	24.2	11.8	5.7	4.3	80.94	67.22	73.43	26.72	43.00	32.92	61.96	53.17
69	102.53	24.2	11.8	5.6	4.4	81.20	67.22	73.53	27.22	44.00	33.59	62.17	53.56
70	103.80	24.2	11.8	5.6	4.4	81.20	67.22	73.53	27.22	44.00	33.59	62.17	53.56
71	105.06	24.2	11.8	5.6	4.4	81.20	67.22	73.53	27.22	44.00	33.59	62.17	53.56
72	106.33	24	12	5.6	4.4	81.09	66.67	73.15	26.83	44.00	33.30	61.74	53.22
73	107.59	24	12	5.6	4.4	81.09	66.67	73.15	26.83	44.00	33.30	61.74	53.22
74	108.86	23.9	12.1	5.6	4.4	81.02	66.39	72.95	26.68	44.00	33.18	61.52	53.07
75	110.13	23.9	12.1	5.7	4.3	80.74	66.39	72.84	26.26	43.00	32.57	61.30	52.71

Continua en la siguiente página

Tabla A.13 – Continuación de la página anterior

						Clase CN			Clase MCI				F1 Sc
Add	Add%	TN	FP	FN	TP	Prec	Recl	F1 Sc	Prec	Recl	F1 Sc	Accur	AVG
76	111.39	24	12	5.7	4.3	80.80	66.67	73.04	26.44	43.00	32.72	61.52	52.88
77	112.66	24	12	5.7	4.3	80.80	66.67	73.04	26.44	43.00	32.72	61.52	52.88
78	113.92	23.8	12.2	5.7	4.3	80.67	66.11	72.66	26.09	43.00	32.46	61.09	52.56
79	115.19	23.8	12.2	5.7	4.3	80.67	66.11	72.66	26.09	43.00	32.46	61.09	52.56
80	116.46	23.8	12.2	5.7	4.3	80.67	66.11	72.66	26.09	43.00	32.46	61.09	52.56
81	117.72	23.7	12.3	5.7	4.3	80.62	65.83	72.47	25.90	43.00	32.32	60.87	52.39
82	118.99	23.4	12.6	5.8	4.2	80.14	65.00	71.78	25.00	42.00	31.34	60.00	51.56
83	120.25	23.3	12.7	5.8	4.2	80.07	64.72	71.58	24.85	42.00	31.23	59.78	51.40
84	121.52	23.1	12.9	5.8	4.2	79.94	64.17	71.19	24.54	42.00	30.98	59.35	51.08
85	122.78	23.1	12.9	5.8	4.2	79.94	64.17	71.19	24.54	42.00	30.98	59.35	51.08
86	124.05	23.1	12.9	5.8	4.2	79.94	64.17	71.19	24.54	42.00	30.98	59.35	51.08
87	125.32	23	13	6	4	79.31	63.89	70.77	23.53	40.00	29.63	58.70	50.20

En la Tabla A.13, la información se encuentra distribuida de la siguiente manera; en la primera columna los elementos agregados al conjunto de Train, en la segunda columna, el *Balance Rate* en porcentaje, que surge de dividir el número actual de elementos en el conjunto de Train sobre sobre el número total de elementos objetivos, esto multiplicado por 100. Las columnas 3, 4, 5 y 6, corresponden a las variables de la matriz de confusión, TP representa a los aciertos del clasificador en la clase de interés y TN corresponden a los aciertos del clasificador en la clase de control "CN"; luego encontramos las métricas de la matriz de clasificación por clase (Precision, Recall y F1 Score), en la columnas 7, 8 y 9, las correspondientes a la clase CN y en las columnas 10, 11 y 12 la clase de interés; ya por último encontramos la columna del Accuracy y el F1 Average, respectivamente.

Tabla A.14: Evaluación de la clasificación Test SubADReSSo, utilizando la técnica de Inversion, mientras se añaden elementos a la clase MCI.

						Clase CN			Clase MCI				F1 Sc
Add	Add%	TN	FP	FN	TP	Prec	Recl	F1 Sc	Prec	Recl	F1 Sc	Accur	AVG
0	15.19	32	4	8	2	80.00	88.89	84.21	33.33	20.00	25.00	73.91	54.61
1	16.46	31.9	4.1	7.9	2.1	80.15	88.61	84.16	34.00	21.00	25.90	73.91	55.03
2	17.72	32.1	3.9	8.2	1.8	79.64	89.17	84.12	33.38	18.00	23.03	73.70	53.58
3	18.99	32.5	3.5	8.3	1.7	79.65	90.28	84.61	34.19	17.00	22.30	74.35	53.46
4	20.25	32.4	3.6	8.3	1.7	79.60	90.00	84.46	33.86	17.00	22.22	74.13	53.34
5	21.52	33.2	2.8	8.1	1.9	80.41	92.22	85.88	42.00	19.00	25.37	76.30	55.62
6	22.78	33.2	2.8	7.7	2.3	81.22	92.22	86.34	45.88	23.00	29.84	77.17	58.09
7	24.05	33	3	7.7	2.3	81.13	91.67	86.05	43.21	23.00	29.43	76.74	57.74
8	25.32	32.8	3.2	7.6	2.4	81.22	91.11	85.85	43.71	24.00	30.32	76.52	58.08
9	26.58	32.8	3.2	7.6	2.4	81.22	91.11	85.85	42.71	24.00	30.27	76.52	58.06
10	27.85	32.8	3.2	7.6	2.4	81.22	91.11	85.86	42.64	24.00	30.36	76.52	58.11
11	29.11	32.7	3.3	7.5	2.5	81.37	90.83	85.82	42.93	25.00	31.23	76.52	58.52
12	30.38	32.6	3.4	7.4	2.6	81.54	90.56	85.79	42.88	26.00	32.00	76.52	58.89
13	31.65	32.4	3.6	7.1	2.9	82.05	90.00	85.82	44.85	29.00	34.82	76.74	60.32
14	32.91	32.6	3.4	7	3	82.36	90.56	86.23	47.10	30.00	36.13	77.39	61.18
15	34.18	32.5	3.5	6.9	3.1	82.53	90.28	86.19	47.25	31.00	36.81	77.39	61.50
16	35.44	32.3	3.7	7	3	82.23	89.72	85.79	44.54	30.00	35.51	76.74	60.65
17	36.71	31.9	4.1	7	3	82.02	88.61	85.17	42.37	30.00	34.90	75.87	60.04
18	37.97	31.5	4.5	7	3	81.84	87.50	84.55	40.42	30.00	34.07	75.00	59.31
19	39.24	31.4	4.6	7	3	81.79	87.22	84.40	39.74	30.00	33.82	74.78	59.11
20	40.51	31.2	4.8	7.2	2.8	81.27	86.67	83.86	37.24	28.00	31.60	73.91	57.73
21	41.77	30.6	5.4	7	3	81.47	85.00	83.13	34.72	30.00	31.76	73.04	57.45
22	43.04	30.5	5.5	7	3	81.41	84.72	82.97	34.45	30.00	31.65	72.83	57.31
23	44.30	30.7	5.3	6.9	3.1	81.71	85.28	83.39	36.41	31.00	33.02	73.48	58.21
24	45.57	30.8	5.2	6.8	3.2	82.01	85.56	83.68	37.24	32.00	33.93	73.91	58.80
25	46.84	30.4	5.6	6.8	3.2	81.83	84.44	83.00	35.95	32.00	33.14	73.04	58.07
26	48.10	30	6	6.9	3.1	81.41	83.33	82.23	33.83	31.00	31.60	71.96	56.92
27	49.37	29.9	6.1	6.8	3.2	81.62	83.06	82.19	34.12	32.00	32.18	71.96	57.18
28	50.63	29.6	6.4	6.8	3.2	81.49	82.22	81.72	32.32	32.00	31.42	71.30	56.57
29	51.90	29.2	6.8	6.8	3.2	81.32	81.11	81.03	30.85	32.00	30.62	70.43	55.83
30	53.16	28.7	7.3	6.8	3.2	81.05	79.72	80.21	29.31	32.00	29.94	69.35	55.07
31	54.43	28.6	7.4	6.9	3.1	80.79	79.44	79.96	27.92	31.00	28.84	68.91	54.40
32	55.70	28.6	7.4	6.7	3.3	81.22	79.44	80.18	29.52	33.00	30.64	69.35	55.41
33	56.96	28.4	7.6	6.3	3.7	82.01	78.89	80.29	32.09	37.00	33.85	69.78	57.07
34	58.23	28.4	7.6	6.3	3.7	82.05	78.89	80.29	31.67	37.00	33.57	69.78	56.93
35	59.49	28.4	7.6	6.3	3.7	82.05	78.89	80.29	31.67	37.00	33.57	69.78	56.93
36	60.76	27.6	8.4	6.1	3.9	81.96	76.67	79.11	31.80	39.00	34.57	68.48	56.84

Continua en la siguiente página

Tabla A.14 – Continuación de la página anterior

						Clase CN			Clase MCI				F1 Sc
Add	Add%	TN	FP	FN	TP	Prec	Recl	F1 Sc	Prec	Recl	F1 Sc	Accur	AVG
37	62.03	27.6	8.4	6.1	3.9	81.96	76.67	79.11	31.80	39.00	34.57	68.48	56.84
38	63.29	27.6	8.4	6	4	82.18	76.67	79.25	32.41	40.00	35.53	68.70	57.39
39	64.56	27.1	8.9	5.8	4.2	82.48	75.28	78.61	31.96	42.00	35.94	68.04	57.27
40	65.82	27	9	5.8	4.2	82.43	75.00	78.44	31.65	42.00	35.76	67.83	57.10
41	67.09	26.9	9.1	5.7	4.3	82.67	74.72	78.38	31.85	43.00	36.21	67.83	57.29
42	68.35	26.9	9.1	5.7	4.3	82.67	74.72	78.38	31.85	43.00	36.21	67.83	57.29
43	69.62	26.8	9.2	5.2	4.8	83.82	74.44	78.80	34.23	48.00	39.82	68.70	59.31
44	70.89	26.9	9.1	5.1	4.9	84.12	74.72	79.09	35.00	49.00	40.69	69.13	59.89
45	72.15	26.9	9.1	5.1	4.9	84.12	74.72	79.09	35.00	49.00	40.69	69.13	59.89
46	73.42	26.6	9.4	5.1	4.9	83.91	73.89	78.50	34.67	49.00	40.42	68.48	59.46
47	74.68	26.6	9.4	5.1	4.9	83.91	73.89	78.50	34.67	49.00	40.42	68.48	59.46
48	75.95	26.7	9.3	5.1	4.9	83.96	74.17	78.67	34.97	49.00	40.59	68.70	59.63
49	77.22	26.8	9.2	5.3	4.7	83.52	74.44	78.61	34.19	47.00	39.27	68.48	58.94
50	78.48	26.6	9.4	5.1	4.9	83.92	73.89	78.49	34.72	49.00	40.38	68.48	59.44
51	79.75	26.8	9.2	5.3	4.7	83.47	74.44	78.64	34.24	47.00	39.43	68.48	59.04
52	81.01	26.4	9.6	5.4	4.6	82.96	73.33	77.78	33.04	46.00	38.25	67.39	58.01
53	82.28	26.3	9.7	5.3	4.7	83.17	73.06	77.70	33.31	47.00	38.76	67.39	58.23
54	83.54	26.2	9.8	5.2	4.8	83.38	72.78	77.63	33.59	48.00	39.27	67.39	58.45
55	84.81	26.1	9.9	5.2	4.8	83.33	72.50	77.45	33.31	48.00	39.08	67.17	58.27
56	86.08	26.2	9.8	5.2	4.8	83.39	72.78	77.65	33.47	48.00	39.21	67.39	58.43
57	87.34	26.2	9.8	5.2	4.8	83.37	72.78	77.62	33.67	48.00	39.30	67.39	58.46
58	88.61	26.1	9.9	5.2	4.8	83.32	72.50	77.43	33.46	48.00	39.14	67.17	58.29
59	89.87	26	10	5.2	4.8	83.26	72.22	77.24	33.28	48.00	39.00	66.96	58.12
60	91.14	25.7	10.3	5.2	4.8	83.11	71.39	76.70	32.50	48.00	38.48	66.30	57.59
61	92.41	25.7	10.3	5.2	4.8	83.13	71.39	76.72	32.41	48.00	38.43	66.30	57.57
62	93.67	25.5	10.5	5.2	4.8	83.04	70.83	76.37	31.81	48.00	38.05	65.87	57.21
63	94.94	25.4	10.6	5.2	4.8	82.99	70.56	76.21	31.45	48.00	37.86	65.65	57.03
64	96.20	25.4	10.6	5.2	4.8	82.99	70.56	76.21	31.45	48.00	37.86	65.65	57.03
65	97.47	25.3	10.7	5.2	4.8	82.93	70.28	76.01	31.28	48.00	37.72	65.43	56.87
66	98.73	25.3	10.7	5.2	4.8	82.93	70.28	76.01	31.28	48.00	37.72	65.43	56.87
67	100.00	25.3	10.7	5.2	4.8	82.93	70.28	76.01	31.28	48.00	37.72	65.43	56.87
68	101.27	25.4	10.6	5.2	4.8	82.99	70.56	76.19	31.60	48.00	37.89	65.65	57.04
69	102.53	25.1	10.9	5.2	4.8	82.81	69.72	75.61	31.07	48.00	37.48	65.00	56.54
70	103.80	24.9	11.1	5.2	4.8	82.70	69.17	75.23	30.67	48.00	37.18	64.57	56.21
71	105.06	24.2	11.8	5.1	4.9	82.59	67.22	74.10	29.40	49.00	36.73	63.26	55.41
72	106.33	24.1	11.9	5.1	4.9	82.52	66.94	73.90	29.26	49.00	36.61	63.04	55.26
73	107.59	24.1	11.9	5.1	4.9	82.52	66.94	73.90	29.26	49.00	36.61	63.04	55.26
74	108.86	24.1	11.9	5.1	4.9	82.52	66.94	73.90	29.26	49.00	36.61	63.04	55.26
75	110.13	24.2	11.8	5.1	4.9	82.58	67.22	74.10	29.42	49.00	36.75	63.26	55.42

Continua en la siguiente página

Tabla A.14 – Continuación de la página anterior

						Clase CN			Clase MCI				F1 Sc
Add	Add%	TN	FP	FN	TP	Prec	Recl	F1 Sc	Prec	Recl	F1 Sc	Accur	AVG
76	111.39	23.9	12.1	5.1	4.9	82.40	66.39	73.51	28.91	49.00	36.34	62.61	54.93
77	112.66	23.9	12.1	5.1	4.9	82.39	66.39	73.50	28.98	49.00	36.37	62.61	54.94
78	113.92	24	12	5.1	4.9	82.46	66.67	73.71	29.07	49.00	36.47	62.83	55.09
79	115.19	23.9	12.1	5.1	4.9	82.40	66.39	73.51	28.91	49.00	36.34	62.61	54.93
80	116.46	23.9	12.1	5	5	82.68	66.39	73.63	29.33	50.00	36.95	62.83	55.29
81	117.72	23.8	12.2	5	5	82.62	66.11	73.43	29.17	50.00	36.81	62.61	55.12
82	118.99	23.8	12.2	5	5	82.62	66.11	73.43	29.17	50.00	36.81	62.61	55.12
83	120.25	23.7	12.3	5	5	82.56	65.83	73.23	29.01	50.00	36.68	62.39	54.96
84	121.52	23.5	12.5	5	5	82.44	65.28	72.84	28.66	50.00	36.41	61.96	54.63
85	122.78	23.3	12.7	5	5	82.32	64.72	72.45	28.33	50.00	36.14	61.52	54.30
86	124.05	23	13	5	5	82.14	63.89	71.87	27.78	50.00	35.71	60.87	53.79
87	125.32	23	13	5	5	82.14	63.89	71.87	27.78	50.00	35.71	60.87	53.79

En la Tabla A.14, la información se encuentra distribuida de la siguiente manera; en la primera columna los elementos agregados al conjunto de Train, en la segunda columna, el *Balance Rate* en porcentaje, que surge de dividir el número actual de elementos en el conjunto de Train sobre sobre el número total de elementos objetivos, esto multiplicado por 100. Las columnas 3, 4, 5 y 6, corresponden a las variables de la matriz de confusión, TP representa a los aciertos del clasificador en la clase de interés y TN corresponden a los aciertos del clasificador en la clase de control "CN"; luego encontramos las métricas de la matriz de clasificación por clase (Precision, Recall y F1 Score), en la columnas 7, 8 y 9, las correspondientes a la clase CN y en las columnas 10, 11 y 12 la clase de interés; ya por último encontramos la columna del Accuracy y el F1 Average, respectivamente.

Tabla A.15: Evaluación de la clasificación Test SubADReSSo, utilizando la técnica de Gain, mientras se añaden elementos a la clase MCI.

						Clase CN			Clase MCI				F1 Sc
Add	Add%	TN	FP	FN	TP	Prec	Recl	F1 Sc	Prec	Recl	F1 Sc	Accur	AVG
0	15.19	32	4	8	2	80.00	88.89	84.21	33.33	20.00	25.00	73.91	54.61
1	16.46	31.7	4.3	8	2	79.85	88.06	83.75	31.90	20.00	24.56	73.26	54.15
2	17.72	31.6	4.4	7.9	2.1	80.00	87.78	83.70	32.50	21.00	25.46	73.26	54.58
3	18.99	31.4	4.6	7.8	2.2	80.10	87.22	83.50	32.44	22.00	26.14	73.04	54.82
4	20.25	31.5	4.5	7.8	2.2	80.14	87.50	83.63	35.02	22.00	26.46	73.26	55.04
5	21.52	31.4	4.6	7.9	2.1	79.88	87.22	83.34	33.85	21.00	25.05	72.83	54.19
6	22.78	31.5	4.5	8	2	79.72	87.50	83.36	34.21	20.00	24.25	72.83	53.81
7	24.05	31.5	4.5	8.1	1.9	79.48	87.50	83.21	34.18	19.00	23.45	72.61	53.33
8	25.32	31.9	4.1	8.1	1.9	79.75	88.61	83.87	34.25	19.00	23.42	73.48	53.65
9	26.58	32.1	3.9	8.3	1.7	79.42	89.17	83.98	33.75	17.00	22.08	73.48	53.03
10	27.85	32.4	3.6	8.3	1.7	79.59	90.00	84.41	36.07	17.00	22.09	74.13	53.25
11	29.11	32.2	3.8	8.3	1.7	79.49	89.44	84.12	34.07	17.00	21.88	73.70	53.00
12	30.38	32.2	3.8	8	2	80.11	89.44	84.46	36.43	20.00	24.90	74.35	54.68
13	31.65	32.4	3.6	8	2	80.22	90.00	84.78	37.08	20.00	25.15	74.78	54.96
14	32.91	32.4	3.6	8	2	80.22	90.00	84.78	37.08	20.00	25.15	74.78	54.96
15	34.18	32.1	3.9	7.9	2.1	80.24	89.17	84.40	38.74	21.00	26.20	74.35	55.30
16	35.44	32.1	3.9	7.9	2.1	80.23	89.17	84.41	38.77	21.00	26.27	74.35	55.34
17	36.71	32.4	3.6	8	2	80.18	90.00	84.75	41.93	20.00	25.77	74.78	55.26
18	37.97	32.5	3.5	7.4	2.6	81.47	90.28	85.58	48.17	26.00	32.08	76.30	58.83
19	39.24	32.4	3.6	7.5	2.5	81.24	90.00	85.33	42.69	25.00	30.40	75.87	57.87
20	40.51	32.2	3.8	7.4	2.6	81.35	89.44	85.14	42.63	26.00	31.09	75.65	58.11
21	41.77	32.1	3.9	7.6	2.4	80.91	89.17	84.76	40.02	24.00	28.53	75.00	56.64
22	43.04	32.1	3.9	7.2	2.8	81.74	89.17	85.22	43.50	28.00	32.74	75.87	58.98
23	44.30	32.1	3.9	7.1	2.9	81.95	89.17	85.33	44.45	29.00	33.76	76.09	59.55
24	45.57	32	4	7.1	2.9	81.87	88.89	85.17	44.48	29.00	33.85	75.87	59.51
25	46.84	32.1	3.9	6.9	3.1	82.37	89.17	85.55	46.26	31.00	35.72	76.52	60.64
26	48.10	32.1	3.9	6.9	3.1	82.37	89.17	85.55	46.26	31.00	35.72	76.52	60.64
27	49.37	32.1	3.9	6.7	3.3	82.81	89.17	85.79	47.53	33.00	37.45	76.96	61.62
28	50.63	31.7	4.3	7	3	81.96	88.06	84.84	41.47	30.00	34.12	75.43	59.48
29	51.90	31.3	4.7	7	3	81.72	86.94	84.19	41.00	30.00	33.91	74.57	59.05
30	53.16	31.4	4.6	6.9	3.1	81.97	87.22	84.45	43.22	31.00	35.15	75.00	59.80
31	54.43	30.9	5.1	6.8	3.2	81.99	85.83	83.76	41.35	32.00	34.79	74.13	59.28
32	55.70	30.8	5.2	6.6	3.4	82.40	85.56	83.81	42.40	34.00	36.21	74.35	60.01
33	56.96	30.6	5.4	6.8	3.2	81.88	85.00	83.31	37.45	32.00	33.64	73.48	58.47
34	58.23	30.4	5.6	6.8	3.2	81.78	84.44	82.98	36.76	32.00	33.32	73.04	58.15
35	59.49	30.2	5.8	7	3	81.27	83.89	82.44	34.10	30.00	30.96	72.17	56.70
36	60.76	30.2	5.8	6.9	3.1	81.46	83.89	82.55	35.76	31.00	32.28	72.39	57.41

Continua en la siguiente página

Tabla A.15 – Continuación de la página anterior

						Clase CN			Clase MCI				F1 Sc
Add	Add%	TN	FP	FN	TP	Prec	Recl	F1 Sc	Prec	Recl	F1 Sc	Accur	AVG
37	62.03	30.2	5.8	6.8	3.2	81.68	83.89	82.67	36.54	32.00	33.20	72.61	57.93
38	63.29	29.7	6.3	6.6	3.4	81.87	82.50	82.12	35.17	34.00	34.15	71.96	58.13
39	64.56	29.3	6.7	6.5	3.5	81.87	81.39	81.58	34.52	35.00	34.51	71.30	58.05
40	65.82	28.9	7.1	6.4	3.6	81.87	80.28	80.98	34.45	36.00	34.76	70.65	57.87
41	67.09	28.8	7.2	6.4	3.6	81.82	80.00	80.82	33.91	36.00	34.57	70.43	57.69
42	68.35	28.7	7.3	6.4	3.6	81.75	79.72	80.61	34.38	36.00	34.51	70.22	57.56
43	69.62	28.4	7.6	6.3	3.7	81.84	78.89	80.23	33.52	37.00	34.73	69.78	57.48
44	70.89	28.5	7.5	6.2	3.8	82.11	79.17	80.51	34.63	38.00	35.78	70.22	58.14
45	72.15	28.5	7.5	6.3	3.7	81.87	79.17	80.39	34.10	37.00	35.04	70.00	57.72
46	73.42	28.5	7.5	6.3	3.7	81.87	79.17	80.39	34.10	37.00	35.04	70.00	57.72
47	74.68	28.5	7.5	6.3	3.7	81.87	79.17	80.39	34.10	37.00	35.04	70.00	57.72
48	75.95	28.5	7.5	6.3	3.7	81.87	79.17	80.39	34.10	37.00	35.04	70.00	57.72
49	77.22	28.3	7.7	6.2	3.8	82.02	78.61	80.21	33.47	38.00	35.34	69.78	57.78
50	78.48	27.8	8.2	6.1	3.9	82.05	77.22	79.51	32.08	39.00	35.04	68.91	57.27
51	79.75	27.9	8.1	6.3	3.7	81.60	77.50	79.43	31.44	37.00	33.79	68.70	56.61
52	81.01	27.9	8.1	6.5	3.5	81.12	77.50	79.20	30.27	35.00	32.24	68.26	55.72
53	82.28	27.9	8.1	6.3	3.7	81.58	77.50	79.42	31.64	37.00	33.89	68.70	56.66
54	83.54	27.9	8.1	6.3	3.7	81.58	77.50	79.42	31.64	37.00	33.89	68.70	56.66
55	84.81	28	8	6.2	3.8	81.88	77.78	79.72	32.39	38.00	34.77	69.13	57.24
56	86.08	27.9	8.1	6.2	3.8	81.83	77.50	79.57	31.97	38.00	34.60	68.91	57.09
57	87.34	28.1	7.9	6.2	3.8	81.94	78.06	79.90	32.58	38.00	34.90	69.35	57.40
58	88.61	27.8	8.2	6.1	3.9	82.01	77.22	79.52	32.29	39.00	35.24	68.91	57.38
59	89.87	27.4	8.6	6	4	82.05	76.11	78.93	31.82	40.00	35.33	68.26	57.13
60	91.14	27.4	8.6	6	4	82.05	76.11	78.93	31.82	40.00	35.33	68.26	57.13
61	92.41	27	9	5.9	4.1	82.05	75.00	78.34	31.48	41.00	35.53	67.61	56.93
62	93.67	26.9	9.1	5.9	4.1	82.00	74.72	78.16	31.26	41.00	35.38	67.39	56.77
63	94.94	26.7	9.3	5.9	4.1	81.89	74.17	77.81	30.77	41.00	35.10	66.96	56.45
64	96.20	26.7	9.3	5.9	4.1	81.89	74.17	77.81	30.77	41.00	35.10	66.96	56.45
65	97.47	26.3	9.7	6	4	81.40	73.06	76.96	29.45	40.00	33.81	65.87	55.39
66	98.73	26.3	9.7	6	4	81.40	73.06	76.96	29.45	40.00	33.81	65.87	55.39
67	100.00	26.3	9.7	5.9	4.1	81.64	73.06	77.08	30.05	41.00	34.59	66.09	55.84
68	101.27	26.7	9.3	5.9	4.1	81.88	74.17	77.81	30.85	41.00	35.14	66.96	56.47
69	102.53	26.6	9.4	5.9	4.1	81.82	73.89	77.62	30.65	41.00	35.01	66.74	56.31
70	103.80	26.5	9.5	5.9	4.1	81.76	73.61	77.44	30.46	41.00	34.87	66.52	56.16
71	105.06	26.5	9.5	5.9	4.1	81.76	73.61	77.44	30.46	41.00	34.87	66.52	56.16
72	106.33	26.3	9.7	6	4	81.41	73.06	76.97	29.36	40.00	33.79	65.87	55.38
73	107.59	26	10	6	4	81.26	72.22	76.43	28.64	40.00	33.28	65.22	54.86
74	108.86	25.8	10.2	6	4	81.14	71.67	76.07	28.23	40.00	33.01	64.78	54.54
75	110.13	25.9	10.1	6	4	81.20	71.94	76.26	28.42	40.00	33.14	65.00	54.70

Continua en la siguiente página

Tabla A.15 – Continuación de la página anterior

						Clase CN			Clase MCI				F1 Sc
Add	Add%	TN	FP	FN	TP	Prec	Recl	F1 Sc	Prec	Recl	F1 Sc	Accur	AVG
76	111.39	25.8	10.2	6	4	81.14	71.67	76.07	28.23	40.00	33.01	64.78	54.54
77	112.66	25.6	10.4	5.9	4.1	81.30	71.11	75.81	28.28	41.00	33.35	64.57	54.58
78	113.92	25.6	10.4	6	4	81.07	71.11	75.70	27.67	40.00	32.57	64.35	54.13
79	115.19	25.5	10.5	6	4	81.01	70.83	75.52	27.48	40.00	32.44	64.13	53.98
80	116.46	25.5	10.5	6.2	3.8	80.45	70.83	75.29	26.62	38.00	31.19	63.70	53.24
81	117.72	25.2	10.8	6.1	3.9	80.52	70.00	74.86	26.56	39.00	31.53	63.26	53.20
82	118.99	25.3	10.7	6.1	3.9	80.58	70.28	75.05	26.75	39.00	31.67	63.48	53.36
83	120.25	25.2	10.8	6.1	3.9	80.52	70.00	74.86	26.56	39.00	31.53	63.26	53.20
84	121.52	25.3	10.7	6.1	3.9	80.58	70.28	75.05	26.73	39.00	31.66	63.48	53.35
85	122.78	25.3	10.7	6.1	3.9	80.58	70.28	75.05	26.73	39.00	31.66	63.48	53.35
86	124.05	25.3	10.7	6.1	3.9	80.58	70.28	75.05	26.73	39.00	31.66	63.48	53.35
87	125.32	25	11	6	4	80.65	69.44	74.63	26.67	40.00	32.00	63.04	53.31

En la Tabla A.15, la información se encuentra distribuida de la siguiente manera; en la primera columna los elementos agregados al conjunto de Train, en la segunda columna, el *Balance Rate* en porcentaje, que surge de dividir el número actual de elementos en el conjunto de Train sobre sobre el número total de elementos objetivos, esto multiplicado por 100. Las columnas 3, 4, 5 y 6, corresponden a las variables de la matriz de confusión, TP representa a los aciertos del clasificador en la clase de interés y TN corresponden a los aciertos del clasificador en la clase de control "CN"; luego encontramos las métricas de la matriz de clasificación por clase (Precision, Recall y F1 Score), en la columnas 7, 8 y 9, las correspondientes a la clase CN y en las columnas 10, 11 y 12 la clase de interés; ya por último encontramos la columna del Accuracy y el F1 Average, respectivamente.

A.3 Referencia para Embeddings Neuronales

Esta sección presenta los resultados de la evaluación de las representaciones basadas en embeddings (wav2vec y wav2vec 2.0) con distintos clasificadores. Se detallan métricas de desempeño para cada combinación de modelo de embedding y clasificador, tanto para el conjunto de datos ADReSSo como para el SubADReSSo.

Tabla A.16: Comparación de modelos *wav2vec* usados como extractores de características y clasificadores elegidos para experimentos de referencia sobre el Test ADReSSo.

Model	Clasificador	TN	FP	FN	TP	Clase CN			Clase AD			Accur	F1 Sc
						Prec	Recl	F1 Sc	Prec	Recl	F1 Sc		AVG
w2v	ADA	25	11	11	24	69.44	69.44	69.44	68.57	68.57	68.57	69.01	69.01
	Polynomial (d=30)	22	14	8	27	73.33	61.11	66.67	65.85	77.14	71.05	69.01	68.86
	Polynomial (d=20)	23	13	10	25	69.70	63.89	66.67	65.79	71.43	68.49	67.61	67.58
	SVM (RBF)	27	9	14	21	65.85	75.00	70.13	70.00	60.00	64.62	67.61	67.37
w2v2_960_lv60	Rnd Forest	25	11	13	22	65.79	69.44	67.57	66.67	62.86	64.71	66.20	66.14
	SVM (RBF)	25	11	15	20	62.50	69.44	65.79	64.52	57.14	60.61	63.38	63.20
	kNN (*k*=20)	27	9	17	18	61.36	75.00	67.50	66.67	51.43	58.06	63.38	62.78
	kNN (*k*=30)	22	14	13	22	62.86	61.11	61.97	61.11	62.86	61.97	61.97	61.97
	Polynomial (d=20)	23	13	14	21	62.16	63.89	63.01	61.76	60.00	60.87	61.97	61.94
	Polynomial (d=30)	21	15	13	22	61.76	58.33	60.00	59.46	62.86	61.11	60.56	60.56
w2v2_lv60	Rnd Forest	25	11	13	22	65.79	69.44	67.57	66.67	62.86	64.71	66.20	66.14
	SVM (RBF)	25	11	15	20	62.50	69.44	65.79	64.52	57.14	60.61	63.38	63.20
	kNN (*k*=20)	27	9	17	18	61.36	75.00	67.50	66.67	51.43	58.06	63.38	62.78
	kNN (*k*=30)	22	14	13	22	62.86	61.11	61.97	61.11	62.86	61.97	61.97	61.97
	Polynomial (d=20)	23	13	14	21	62.16	63.89	63.01	61.76	60.00	60.87	61.97	61.94
	Polynomial (d=30)	21	15	13	22	61.76	58.33	60.00	59.46	62.86	61.11	60.56	60.56
w2v2_lv60_self	Rnd Forest	25	11	13	22	65.79	69.44	67.57	66.67	62.86	64.71	66.20	66.14
	SVM (RBF)	25	11	15	20	62.50	69.44	65.79	64.52	57.14	60.61	63.38	63.20
	kNN (*k*=20)	27	9	17	18	61.36	75.00	67.50	66.67	51.43	58.06	63.38	62.78
	kNN (*k*=30)	22	14	13	22	62.86	61.11	61.97	61.11	62.86	61.97	61.97	61.97
	Polynomial (d=20)	23	13	14	21	62.16	63.89	63.01	61.76	60.00	60.87	61.97	61.94
	Polynomial (d=30)	21	15	13	22	61.76	58.33	60.00	59.46	62.86	61.11	60.56	60.56

En la Tabla A.16, la información se encuentra distribuida de la siguiente manera; en la primera columna los modelos *wav2vec* utilizado como extractores de características;en la segunda columna, el clasificador utilizado, "ADA" para el

clasificador ADABoost, "RndForest" es Random forest, "kNN" es el clasificador de k vecinos mas cercanos donde la k en estas pruebas es igual a 5, 10 y 30. También se encuentra el clasificador SVM con kernel lineal y Polinomial, donde este último esta evaluado con grado 20 y 30. Las columnas 3, 4, 5 y 6, corresponden a las variables de la matriz de confusión, TP (True Positive) representa a los aciertos del clasificador en la clase de interés y TN (True Negative) corresponden a los aciertos del clasificador en la clase de control "CN"; luego encontramos las métricas de la matriz de clasificación por clase (Precision, Recall y F1 Score), en la columnas 7, 8 y 9, las correspondientes a la clase CN y en las columnas 10, 11 y 12 la clase de interés; ya por último encontramos la columna del Accuracy y el F1 Average, respectivamente.

Tabla A.17: Comparación de modelos *wav2vec* usados como extractores de características y clasificadores elegidos para experimentos de referencia sobre el Test SubADReSSo.

Model	Clasificador	TN	FP	FN	TP	Clase CN			Clase MCI			Accur	F1 Sc
						Prec	Recl	F1 Sc	Prec	Recl	F1 Sc		AVG
w2v	Rnd Forest	36	0	10	0	78.26	100.00	87.80	0.00	0.00	0.00	78.26	43.90
	kNN (*k*=5)	36	0	10	0	78.26	100.00	87.80	0.00	0.00	0.00	78.26	43.90
	kNN (*k*=10)	36	0	10	0	78.26	100.00	87.80	0.00	0.00	0.00	78.26	43.90
	kNN (*k*=20)	36	0	10	0	78.26	100.00	87.80	0.00	0.00	0.00	78.26	43.90
	kNN (*k*=30)	36	0	10	0	78.26	100.00	87.80	0.00	0.00	0.00	78.26	43.90
	SVM (RBF)	32	4	10	0	76.19	88.89	82.05	0.00	0.00	0.00	69.57	41.03
	Polynomial (d=30)	32	4	10	0	76.19	88.89	82.05	0.00	0.00	0.00	69.57	41.03
	ADA	36	0	10	0	78.26	100.00	87.80	0.00	0.00	0.00	78.26	43.90
w2v2_960_lv60	SVM (RBF)	35	1	9	1	79.55	97.22	87.50	50.00	10.00	16.67	78.26	52.08
	Polynomial (d=30)	36	0	10	0	78.26	100.00	87.80	0.00	0.00	0.00	78.26	43.90
	ADA	34	2	9	1	79.07	94.44	86.08	33.33	10.00	15.38	76.09	50.73
w2v2_lv60	SVM (RBF)	35	1	9	1	79.55	97.22	87.50	50.00	10.00	16.67	78.26	52.08
	Polynomial (d=30)	36	0	10	0	78.26	100.00	87.80	0.00	0.00	0.00	78.26	43.90
	ADA	34	2	9	1	79.07	94.44	86.08	33.33	10.00	15.38	76.09	50.73
w2v2_lv60_self2	SVM (RBF)	35	1	9	1	79.55	97.22	87.50	50.00	10.00	16.67	78.26	52.08
	Polynomial (d=30)	36	0	10	0	78.26	100.00	87.80	0.00	0.00	0.00	78.26	43.90
	ADA	34	2	9	1	79.07	94.44	86.08	33.33	10.00	15.38	76.09	50.73

En la Tabla A.17, la información se encuentra distribuida de la siguiente manera; en la primera columna los modelos *wav2vec* utilizado como extractores de características;en la segunda columna, el clasificador utilizado, "ADA" para el clasificador ADABoost, "RndForest" es Random forest, "kNN" es el clasificador de k vecinos mas cercanos donde la k en estas pruebas es igual a 5, 10 y 30. También se encuentra el clasificador SVM con kernel lineal y Polinomial, donde este último esta evaluado con grado 20 y 30. Las columnas 3, 4, 5 y 6, corresponden a las variables de la matriz de confusión, TP (True Positive) representa a los aciertos del clasificador en la clase de interés y TN (True Negative) corresponden a los aciertos del clasificador en la clase de control "CN"; luego encontramos las métricas de la matriz de clasificación por clase (Precision, Recall y F1 Score), en la columnas 7, 8 y 9, las correspondientes a la clase CN y en las columnas 10, 11 y 12 la clase

de interés; ya por último encontramos la columna del Accuracy y el F1 Average, respectivamente.

A.4 Aumento de Datos para Embeddings Neuronales

Esta sección presenta los resultados detallados de los experimentos de aumento de datos aplicados a las representaciones basadas en embeddings. Se desglosan métricas para cada estrategia de aumento (adición de elementos reales, SMOTE y generación con WGAN-GP(CNN1D) y cWGAN-GP(CNN1D)), en cada paso de adición.

Tabla A.18: Resultados de clasificación, evaluando el Test SubADReSSo, usando diferentes clasificadores entrenados con el Train SubADReSSo aumentado con la clase CI del Train ADReSSo, utilizando modelos de *wav2vec* para extracción de características.

Model	Clasificador	TN	FP	FN	TP	Clase CN			Clase MCI			Accur	F1 Sc
						Prec	Recl	F1 Sc	Prec	Recl	F1 Sc		AVG
w2v	ADA	25	11	4	6	86.21	69.44	76.92	35.29	60.00	44.44	67.39	60.68
	SVM(RBF)	27	9	6	4	81.82	75.00	78.26	30.77	40.00	34.78	67.39	56.52
	kNN(*k*=5)	26	10	6	4	81.25	72.22	76.47	28.57	40.00	33.33	65.22	54.90
	Polynom(d=30)	22	14	5	5	81.48	61.11	69.84	26.32	50.00	34.48	58.70	52.16
w2v2_960_lv60	Rnd Forest	25	11	5	5	83.33	69.44	75.76	31.25	50.00	38.46	65.22	57.11
	SVM(RBF)	25	11	5	5	83.33	69.44	75.76	31.25	50.00	38.46	65.22	57.11
	kNN(*k*=20)	27	9	6	4	81.82	75.00	78.26	30.77	40.00	34.78	67.39	56.52
	kNN(*k*=30)	22	14	4	6	84.62	61.11	70.97	30.00	60.00	40.00	60.87	55.48
	kNN(*k*=5)	26	10	6	4	81.25	72.22	76.47	28.57	40.00	33.33	65.22	54.90
	Polynom(d=30)	21	15	4	6	84.00	58.33	68.85	28.57	60.00	38.71	58.70	53.78
w2v2_lv60	Rnd Forest	25	11	5	5	83.33	69.44	75.76	31.25	50.00	38.46	65.22	57.11
	SVM(RBF)	25	11	5	5	83.33	69.44	75.76	31.25	50.00	38.46	65.22	57.11
	kNN(*k*=20)	27	9	6	4	81.82	75.00	78.26	30.77	40.00	34.78	67.39	56.52
	kNN(*k*=30)	22	14	4	6	84.62	61.11	70.97	30.00	60.00	40.00	60.87	55.48
	kNN(*k*=5)	26	10	6	4	81.25	72.22	76.47	28.57	40.00	33.33	65.22	54.90
	Polynom(d=30)	21	15	4	6	84.00	58.33	68.85	28.57	60.00	38.71	58.70	53.78
w2v2_lv60_self	Rnd Forest	25	11	5	5	83.33	69.44	75.76	31.25	50.00	38.46	65.22	57.11
	SVM(RBF)	25	11	5	5	83.33	69.44	75.76	31.25	50.00	38.46	65.22	57.11
	kNN(*k*=20)	27	9	6	4	81.82	75.00	78.26	30.77	40.00	34.78	67.39	56.52
	kNN(*k*=30)	22	14	4	6	84.62	61.11	70.97	30.00	60.00	40.00	60.87	55.48
	kNN(*k*=5)	26	10	6	4	81.25	72.22	76.47	28.57	40.00	33.33	65.22	54.90
	Polynom(d=30)	21	15	4	6	84.00	58.33	68.85	28.57	60.00	38.71	58.70	53.78
	kNN(*k*=10)	28	8	7	3	80.00	77.78	78.87	27.27	30.00	28.57	67.39	53.72
	Polynom(d=5)	22	14	5	5	81.48	61.11	69.84	26.32	50.00	34.48	58.70	52.16

En la Tabla A.18, la información se encuentra distribuida de la siguiente manera; en la primera columna los modelos *wav2vec* utilizado como extractores de características;en la segunda columna, el clasificador utilizado, "ADA" para el clasificador ADABoost, "RndForest" es Random forest, "kNN" es el clasificador de k vecinos mas cercanos donde la k en estas pruebas es igual a 5, 10 y 30. También se encuentra el clasificador SVM con kernel lineal y Polinomial, donde este último esta evaluado con grado 20 y 30. Las columnas 3, 4, 5 y 6, corresponden a las variables de la matriz de confusión, TP (True Positive) representa a los aciertos del clasificador en la clase de interés y TN (True Negative) corresponden a los aciertos del clasificador en la clase de control "CN"; luego encontramos las métricas de la matriz de clasificación por clase (Precision, Recall y F1 Score), en la columnas 7, 8 y 9, las correspondientes a la clase CN y en las columnas 10, 11 y 12 la clase de interés; ya por último encontramos la columna del Accuracy y el F1 Average, respectivamente.

Tabla A.19: Evaluación de la clasificación del Test SubADReSSo, añadiendo datos sintéticos con SMOTE a la clase MCI.

						Clase CN			Clase MCI				F1 Sc
Add	Add%	TN	FP	FN	TP	Prec	Recl	F1 Sc	Prec	Recl	F1 Sc	Accur	AVG
0	15.19	32	4	10	0	76.19	88.89	82.05	0.00	0.00	0.00	69.57	41.03
1	16.46	30.3	5.7	9.2	0.8	76.84	84.17	80.26	7.64	8.00	7.81	67.61	44.04
2	17.72	29.7	6.3	8.2	1.8	78.59	82.50	80.41	16.99	18.00	17.45	68.48	48.93
3	18.99	28.7	7.3	7.8	2.2	78.80	79.72	79.20	20.24	22.00	21.04	67.17	50.12
4	20.25	28.6	7.4	7.8	2.2	78.73	79.44	79.03	20.24	22.00	21.04	66.96	50.03
5	21.52	28	8	7.6	2.4	78.86	77.78	78.27	20.61	24.00	22.12	66.09	50.20
6	22.78	27.6	8.4	7.5	2.5	78.68	76.67	77.53	22.24	25.00	23.14	65.43	50.34
7	24.05	27.7	8.3	7.4	2.6	78.97	76.94	77.87	23.62	26.00	24.41	65.87	51.14
8	25.32	27.1	8.9	7.4	2.6	78.55	75.28	76.80	22.94	26.00	24.04	64.57	50.42
9	26.58	27.1	8.9	7.5	2.5	78.36	75.28	76.70	21.50	25.00	22.84	64.35	49.77
10	27.85	27.1	8.9	7.3	2.7	78.82	75.28	76.92	22.87	27.00	24.45	64.78	50.68
11	29.11	26.9	9.1	7.1	2.9	79.15	74.72	76.79	24.18	29.00	26.10	64.78	51.44
12	30.38	26.9	9.1	7	3	79.39	74.72	76.92	24.65	30.00	26.88	65.00	51.90
13	31.65	27.2	8.8	6.8	3.2	80.10	75.56	77.69	26.07	32.00	28.50	66.09	53.09
14	32.91	26.8	9.2	6.7	3.3	80.03	74.44	77.06	26.35	33.00	29.09	65.43	53.07
15	34.18	26.6	9.4	6.4	3.6	80.73	73.89	77.04	27.35	36.00	30.77	65.65	53.90
16	35.44	26.3	9.7	6.3	3.7	80.84	73.06	76.62	27.12	37.00	30.97	65.22	53.79
17	36.71	26	10	6.2	3.8	80.86	72.22	76.16	27.32	38.00	31.46	64.78	53.81
18	37.97	25.9	10.1	6.2	3.8	80.78	71.94	75.97	27.25	38.00	31.41	64.57	53.69
19	39.24	25.7	10.3	6.1	3.9	80.87	71.39	75.72	27.63	39.00	32.10	64.35	53.91
20	40.51	25.6	10.4	5.9	4.1	81.37	71.11	75.78	28.21	41.00	33.17	64.57	54.48
21	41.77	25.5	10.5	5.9	4.1	81.34	70.83	75.60	27.92	41.00	32.96	64.35	54.28
22	43.04	25.6	10.4	5.6	4.4	82.17	71.11	76.10	29.69	44.00	35.16	65.22	55.63
23	44.30	25.3	10.7	5.5	4.5	82.24	70.28	75.69	29.59	45.00	35.51	64.78	55.60
24	45.57	25.3	10.7	5.4	4.6	82.49	70.28	75.80	30.09	46.00	36.19	65.00	55.99
25	46.84	25.3	10.7	5.3	4.7	82.78	70.28	75.89	30.55	47.00	36.81	65.22	56.35
26	48.10	25.3	10.7	5.2	4.8	83.02	70.28	76.00	31.15	48.00	37.57	65.43	56.78
27	49.37	24.9	11.1	5.3	4.7	82.49	69.17	75.12	30.00	47.00	36.41	64.35	55.77
28	50.63	24.9	11.1	5.4	4.6	82.24	69.17	75.01	29.48	46.00	35.71	64.13	55.36
29	51.90	24.8	11.2	5.4	4.6	82.17	68.89	74.81	29.33	46.00	35.60	63.91	55.21
30	53.16	24.5	11.5	5.3	4.7	82.28	68.06	74.39	29.06	47.00	35.77	63.48	55.08
31	54.43	24.5	11.5	5.3	4.7	82.27	68.06	74.41	29.02	47.00	35.77	63.48	55.09
32	55.70	23.8	12.2	5.2	4.8	82.04	66.11	73.05	28.51	48.00	35.57	62.17	54.31
33	56.96	23.6	12.4	5.3	4.7	81.66	65.56	72.57	27.64	47.00	34.62	61.52	53.59
34	58.23	23.7	12.3	5.2	4.8	82.00	65.83	72.87	28.25	48.00	35.37	61.96	54.12
35	59.49	23.7	12.3	5.2	4.8	82.00	65.83	72.87	28.25	48.00	35.37	61.96	54.12
36	60.76	23.9	12.1	4.8	5.2	83.51	66.39	73.76	29.84	52.00	37.67	63.26	55.71

Continua en la siguiente página

Tabla A.19 – Continuación de la página anterior

						Clase CN			Clase MCI				F1 Sc
Add	Add%	TN	FP	FN	TP	Prec	Recl	F1 Sc	Prec	Recl	F1 Sc	Accur	AVG
37	62.03	23.8	12.2	4.5	5.5	84.59	66.11	73.91	30.59	55.00	38.96	63.70	56.43
38	63.29	23.7	12.3	4.3	5.7	85.06	65.83	73.94	31.36	57.00	40.14	63.91	57.04
39	64.56	23.6	12.4	4.3	5.7	84.99	65.56	73.74	31.23	57.00	40.04	63.70	56.89
40	65.82	23.4	12.6	4.3	5.7	84.83	65.00	73.31	31.01	57.00	39.85	63.26	56.58
41	67.09	23.4	12.6	4.3	5.7	84.83	65.00	73.31	31.01	57.00	39.85	63.26	56.58
42	68.35	23.2	12.8	4.4	5.6	84.43	64.44	72.76	30.33	56.00	38.98	62.61	55.87
43	69.62	23.3	12.7	4.4	5.6	84.48	64.72	72.97	30.48	56.00	39.12	62.83	56.05
44	70.89	23.3	12.7	4.4	5.6	84.48	64.72	72.97	30.48	56.00	39.12	62.83	56.05
45	72.15	23.1	12.9	4.5	5.5	83.98	64.17	72.39	29.93	55.00	38.40	62.17	55.40
46	73.42	23.4	12.6	4.5	5.5	84.22	65.00	73.02	30.28	55.00	38.70	62.83	55.86
47	74.68	23.3	12.7	4.5	5.5	84.15	64.72	72.82	30.14	55.00	38.57	62.61	55.69
48	75.95	23.3	12.7	4.5	5.5	84.15	64.72	72.82	30.14	55.00	38.57	62.61	55.69
49	77.22	23.3	12.7	4.4	5.6	84.61	64.72	72.93	30.40	56.00	39.00	62.83	55.97
50	78.48	23.3	12.7	4.4	5.6	84.61	64.72	72.93	30.40	56.00	39.00	62.83	55.97
51	79.75	23.2	12.8	4.3	5.7	84.76	64.44	72.89	30.52	57.00	39.43	62.83	56.16
52	81.01	23.1	12.9	4.2	5.8	85.02	64.17	72.80	30.73	58.00	39.84	62.83	56.32
53	82.28	23.1	12.9	4.1	5.9	85.32	64.17	72.92	31.10	59.00	40.41	63.04	56.67
54	83.54	23.1	12.9	4.1	5.9	85.32	64.17	72.92	31.10	59.00	40.41	63.04	56.67
55	84.81	23	13	4.1	5.9	85.30	63.89	72.73	30.87	59.00	40.22	62.83	56.47
56	86.08	23	13	4.1	5.9	85.30	63.89	72.73	30.87	59.00	40.22	62.83	56.47
57	87.34	22.9	13.1	4.1	5.9	85.29	63.61	72.54	30.63	59.00	40.02	62.61	56.28
58	88.61	22.5	13.5	3.9	6.1	85.70	62.50	71.93	30.76	61.00	40.60	62.17	56.27
59	89.87	22.4	13.6	3.6	6.4	86.59	62.22	72.10	31.67	64.00	42.13	62.61	57.11
60	91.14	22.5	13.5	3.8	6.2	86.00	62.50	72.06	31.15	62.00	41.20	62.39	56.63
61	92.41	22.6	13.4	3.7	6.3	86.34	62.78	72.37	31.74	63.00	41.94	62.83	57.16
62	93.67	22.2	13.8	3.9	6.1	85.36	61.67	71.16	30.70	61.00	40.51	61.52	55.84
63	94.94	22.3	13.7	3.9	6.1	85.41	61.94	71.35	30.92	61.00	40.68	61.74	56.02
64	96.20	22.3	13.7	3.8	6.2	85.70	61.94	71.43	31.33	62.00	41.28	61.96	56.35
65	97.47	22.4	13.6	3.8	6.2	85.76	62.22	71.62	31.52	62.00	41.42	62.17	56.52
66	98.73	22.3	13.7	3.8	6.2	85.72	61.94	71.43	31.29	62.00	41.24	61.96	56.33
67	100.00	22.3	13.7	3.4	6.6	86.87	61.94	71.89	32.97	66.00	43.66	62.83	57.77
68	101.27	22.5	13.5	3.2	6.8	87.54	62.50	72.49	34.22	68.00	45.20	63.70	58.85
69	102.53	22.5	13.5	3.2	6.8	87.54	62.50	72.49	34.22	68.00	45.20	63.70	58.85
70	103.80	22.3	13.7	3.2	6.8	87.46	61.94	72.12	33.74	68.00	44.82	63.26	58.47
71	105.06	22.3	13.7	3.2	6.8	87.46	61.94	72.12	33.74	68.00	44.82	63.26	58.47
72	106.33	22.2	13.8	3.3	6.7	87.09	61.67	71.80	33.21	67.00	44.13	62.83	57.97
73	107.59	22.2	13.8	3.3	6.7	87.09	61.67	71.80	33.21	67.00	44.13	62.83	57.97
74	108.86	22.7	13.3	3.3	6.7	87.33	63.06	72.70	34.48	67.00	45.01	63.91	58.86
75	110.13	22.7	13.3	3.3	6.7	87.33	63.06	72.70	34.48	67.00	45.01	63.91	58.86

Continua en la siguiente página

Tabla A.19 – Continuación de la página anterior

						Clase CN			Clase MCI				F1 Sc
Add	Add%	TN	FP	FN	TP	Prec	Recl	F1 Sc	Prec	Recl	F1 Sc	Accur	AVG
76	111.39	22.7	13.3	3.3	6.7	87.33	63.06	72.70	34.48	67.00	45.01	63.91	58.86
77	112.66	22.7	13.3	3.3	6.7	87.33	63.06	72.70	34.48	67.00	45.01	63.91	58.86
78	113.92	22.9	13.1	3.5	6.5	86.84	63.61	72.84	34.12	65.00	44.16	63.91	58.50
79	115.19	22.9	13.1	3.5	6.5	86.85	63.61	72.82	34.21	65.00	44.17	63.91	58.50
80	116.46	23.1	12.9	3.5	6.5	86.93	64.17	73.20	34.65	65.00	44.54	64.35	58.87
81	117.72	23	13	3.4	6.6	87.16	63.89	73.10	34.91	66.00	45.02	64.35	59.06
82	118.99	23	13	3.5	6.5	86.88	63.89	72.99	34.50	65.00	44.40	64.13	58.69
83	120.25	23	13	3.4	6.6	87.16	63.89	73.10	34.91	66.00	45.02	64.35	59.06
84	121.52	23	13	3.4	6.6	87.16	63.89	73.10	34.91	66.00	45.02	64.35	59.06
85	122.78	22.8	13.2	3.4	6.6	87.05	63.33	72.67	34.49	66.00	44.71	63.91	58.69
86	124.05	22.8	13.2	3.4	6.6	87.05	63.33	72.67	34.49	66.00	44.71	63.91	58.69
87	125.32	22.8	13.2	3.4	6.6	87.05	63.33	72.67	34.49	66.00	44.71	63.91	58.69
88	126.58	22.8	13.2	3.4	6.6	87.05	63.33	72.67	34.49	66.00	44.71	63.91	58.69
89	127.85	22.8	13.2	3.4	6.6	87.05	63.33	72.67	34.49	66.00	44.71	63.91	58.69
90	129.11	22.6	13.4	3.5	6.5	86.64	62.78	72.16	33.76	65.00	43.85	63.26	58.00
91	130.38	22.5	13.5	3.5	6.5	86.58	62.50	71.95	33.60	65.00	43.71	63.04	57.83
92	131.65	22.4	13.6	3.5	6.5	86.52	62.22	71.73	33.47	65.00	43.58	62.83	57.65
93	132.91	22.4	13.6	3.5	6.5	86.52	62.22	71.73	33.47	65.00	43.58	62.83	57.65
94	134.18	22.4	13.6	3.5	6.5	86.52	62.22	71.73	33.47	65.00	43.58	62.83	57.65
95	135.44	22.3	13.7	3.4	6.6	86.81	61.94	71.67	33.52	66.00	43.91	62.83	57.79
96	136.71	22.5	13.5	3.5	6.5	86.59	62.50	71.95	33.55	65.00	43.68	63.04	57.82
97	137.97	22.4	13.6	3.5	6.5	86.51	62.22	71.67	33.46	65.00	43.58	62.83	57.63
98	139.24	22.3	13.7	3.5	6.5	86.47	61.94	71.50	33.19	65.00	43.40	62.61	57.45
99	140.51	22.3	13.7	3.5	6.5	86.47	61.94	71.50	33.19	65.00	43.40	62.61	57.45
100	141.77	22.3	13.7	3.5	6.5	86.47	61.94	71.50	33.19	65.00	43.40	62.61	57.45
101	143.04	22.4	13.6	3.5	6.5	86.51	62.22	71.67	33.46	65.00	43.58	62.83	57.63
102	144.30	22.4	13.6	3.5	6.5	86.51	62.22	71.67	33.46	65.00	43.58	62.83	57.63
103	145.57	22.4	13.6	3.5	6.5	86.50	62.22	71.61	33.65	65.00	43.65	62.83	57.63
104	146.84	22.3	13.7	3.4	6.6	86.89	61.94	71.47	33.82	66.00	43.98	62.83	57.72
105	148.10	22.3	13.7	3.4	6.6	86.89	61.94	71.47	33.82	66.00	43.98	62.83	57.72
106	149.37	22.3	13.7	3.4	6.6	86.89	61.94	71.47	33.82	66.00	43.98	62.83	57.72
107	150.63	22.5	13.5	3.3	6.7	87.36	62.50	72.02	34.41	67.00	44.73	63.48	58.38
108	151.90	22.5	13.5	3.3	6.7	87.36	62.50	72.02	34.41	67.00	44.73	63.48	58.38
109	153.16	22.5	13.5	3.3	6.7	87.36	62.50	72.02	34.41	67.00	44.73	63.48	58.38
110	154.43	22.5	13.5	3.3	6.7	87.36	62.50	72.02	34.41	67.00	44.73	63.48	58.38
111	155.70	22.5	13.5	3.3	6.7	87.37	62.50	72.09	34.36	67.00	44.71	63.48	58.40
112	156.96	22.5	13.5	3.4	6.6	87.00	62.50	71.97	34.06	66.00	44.22	63.26	58.09
113	158.23	22.5	13.5	3.4	6.6	87.00	62.50	71.97	34.06	66.00	44.22	63.26	58.09
114	159.49	22.5	13.5	3.4	6.6	87.00	62.50	71.97	34.06	66.00	44.22	63.26	58.09

Continua en la siguiente página

Tabla A.19 – Continuación de la página anterior

						Clase CN			Clase MCI				F1 Sc
Add	Add%	TN	FP	FN	TP	Prec	Recl	F1 Sc	Prec	Recl	F1 Sc	Accur	AVG
115	160.76	22.5	13.5	3.4	6.6	87.00	62.50	71.97	34.06	66.00	44.22	63.26	58.09
116	162.03	22.5	13.5	3.4	6.6	87.00	62.50	71.97	34.06	66.00	44.22	63.26	58.09
117	163.29	22.5	13.5	3.4	6.6	87.00	62.50	71.97	34.06	66.00	44.22	63.26	58.09
118	164.56	22.5	13.5	3.4	6.6	87.00	62.50	71.97	34.06	66.00	44.22	63.26	58.09
119	165.82	22.5	13.5	3.4	6.6	87.00	62.50	71.97	34.06	66.00	44.22	63.26	58.09
120	167.09	22.5	13.5	3.4	6.6	87.00	62.50	71.97	34.06	66.00	44.22	63.26	58.09
121	168.35	22.5	13.5	3.4	6.6	87.00	62.50	71.97	34.06	66.00	44.22	63.26	58.09
122	169.62	22.5	13.5	3.4	6.6	87.00	62.50	71.97	34.06	66.00	44.22	63.26	58.09
123	170.89	22.5	13.5	3.4	6.6	87.00	62.50	71.97	34.06	66.00	44.22	63.26	58.09
124	172.15	22.5	13.5	3.4	6.6	86.98	62.50	71.91	34.23	66.00	44.31	63.26	58.11
125	173.42	22.3	13.7	3.4	6.6	86.89	61.94	71.56	33.63	66.00	43.93	62.83	57.74
126	174.68	22.4	13.6	3.4	6.6	86.95	62.22	71.82	33.74	66.00	44.04	63.04	57.93
127	175.95	22.3	13.7	3.4	6.6	86.91	61.94	71.64	33.45	66.00	43.84	62.83	57.74
128	177.22	22.3	13.7	3.4	6.6	86.91	61.94	71.64	33.45	66.00	43.84	62.83	57.74
129	178.48	22.3	13.7	3.4	6.6	86.91	61.94	71.64	33.45	66.00	43.84	62.83	57.74
130	179.75	22.3	13.7	3.4	6.6	86.91	61.94	71.64	33.45	66.00	43.84	62.83	57.74
131	181.01	22.9	13.1	3.4	6.6	87.22	63.61	73.05	34.30	66.00	44.62	64.13	58.84
132	182.28	23	13	3.5	6.5	86.87	63.89	73.11	34.20	65.00	44.32	64.13	58.72
133	183.54	23	13	3.5	6.5	86.87	63.89	73.11	34.20	65.00	44.32	64.13	58.72
134	184.81	23	13	3.5	6.5	86.87	63.89	73.11	34.20	65.00	44.32	64.13	58.72
135	186.08	23	13	3.5	6.5	86.87	63.89	73.11	34.20	65.00	44.32	64.13	58.72
136	187.34	23	13	3.5	6.5	86.87	63.89	73.11	34.20	65.00	44.32	64.13	58.72
137	188.61	23	13	3.5	6.5	86.87	63.89	73.11	34.20	65.00	44.32	64.13	58.72
138	189.87	23	13	3.5	6.5	86.87	63.89	73.11	34.20	65.00	44.32	64.13	58.72
139	191.14	23.1	12.9	3.5	6.5	86.94	64.17	73.35	34.31	65.00	44.44	64.35	58.89
140	192.41	23.3	12.7	3.5	6.5	87.05	64.72	73.85	34.53	65.00	44.67	64.78	59.26
141	193.67	23.3	12.7	3.5	6.5	87.05	64.72	73.85	34.53	65.00	44.67	64.78	59.26
142	194.94	23.2	12.8	3.5	6.5	87.00	64.44	73.68	34.26	65.00	44.49	64.57	59.08
143	196.20	23.2	12.8	3.5	6.5	87.00	64.44	73.68	34.26	65.00	44.49	64.57	59.08
144	197.47	23.2	12.8	3.5	6.5	87.00	64.44	73.68	34.26	65.00	44.49	64.57	59.08
145	198.73	23.2	12.8	3.5	6.5	87.00	64.44	73.68	34.26	65.00	44.49	64.57	59.08
146	200.00	23.9	12.1	3.6	6.4	86.93	66.39	75.06	35.07	64.00	45.05	65.87	60.05
147	201.27	23.9	12.1	3.6	6.4	86.93	66.39	75.06	35.07	64.00	45.05	65.87	60.05
148	202.53	23.9	12.1	3.6	6.4	86.93	66.39	75.06	35.07	64.00	45.05	65.87	60.05
149	203.80	23.8	12.2	3.6	6.4	86.86	66.11	74.82	34.95	64.00	44.93	65.65	59.88
150	205.06	23.8	12.2	3.6	6.4	86.86	66.11	74.82	34.95	64.00	44.93	65.65	59.88
151	206.33	23.8	12.2	3.6	6.4	86.86	66.11	74.82	34.95	64.00	44.93	65.65	59.88
152	207.59	23.8	12.2	3.6	6.4	86.86	66.11	74.82	34.95	64.00	44.93	65.65	59.88
153	208.86	23.8	12.2	3.6	6.4	86.86	66.11	74.82	34.95	64.00	44.93	65.65	59.88

Continua en la siguiente página

Tabla A.19 – Continuación de la página anterior

						Clase CN			Clase MCI				F1 Sc
Add	Add%	TN	FP	FN	TP	Prec	Recl	F1 Sc	Prec	Recl	F1 Sc	Accur	AVG
154	210.13	23.8	12.2	3.6	6.4	86.86	66.11	74.82	34.95	64.00	44.93	65.65	59.88
155	211.39	23.8	12.2	3.5	6.5	87.21	66.11	74.94	35.28	65.00	45.43	65.87	60.19
156	212.66	23.8	12.2	3.5	6.5	87.21	66.11	74.94	35.28	65.00	45.43	65.87	60.19
157	213.92	23.8	12.2	3.5	6.5	87.21	66.11	74.94	35.28	65.00	45.43	65.87	60.19
158	215.19	23.8	12.2	3.5	6.5	87.21	66.11	74.94	35.28	65.00	45.43	65.87	60.19
159	216.46	23.7	12.3	3.5	6.5	87.16	65.83	74.76	35.04	65.00	45.27	65.65	60.01
160	217.72	23.7	12.3	3.5	6.5	87.16	65.83	74.76	35.04	65.00	45.27	65.65	60.01
161	218.99	23.7	12.3	3.5	6.5	87.16	65.83	74.76	35.04	65.00	45.27	65.65	60.01
162	220.25	23.7	12.3	3.5	6.5	87.16	65.83	74.76	35.04	65.00	45.27	65.65	60.01
163	221.52	23.7	12.3	3.5	6.5	87.16	65.83	74.76	35.04	65.00	45.27	65.65	60.01
164	222.78	23.7	12.3	3.5	6.5	87.16	65.83	74.76	35.04	65.00	45.27	65.65	60.01
165	224.05	23.7	12.3	3.5	6.5	87.16	65.83	74.76	35.04	65.00	45.27	65.65	60.01
166	225.32	23.6	12.4	3.5	6.5	87.12	65.56	74.57	34.82	65.00	45.10	65.43	59.83
167	226.58	23.6	12.4	3.5	6.5	87.12	65.56	74.57	34.82	65.00	45.10	65.43	59.83
168	227.85	23.6	12.4	3.5	6.5	87.12	65.56	74.57	34.82	65.00	45.10	65.43	59.83
169	229.11	23.6	12.4	3.5	6.5	87.12	65.56	74.57	34.82	65.00	45.10	65.43	59.83
170	230.38	23.6	12.4	3.5	6.5	87.12	65.56	74.57	34.82	65.00	45.10	65.43	59.83
171	231.65	23.7	12.3	3.5	6.5	87.17	65.83	74.79	34.97	65.00	45.24	65.65	60.01
172	232.91	23.8	12.2	3.5	6.5	87.20	66.11	74.97	35.20	65.00	45.42	65.87	60.20
173	234.18	23.8	12.2	3.5	6.5	87.20	66.11	74.97	35.20	65.00	45.42	65.87	60.20
174	235.44	23.8	12.2	3.5	6.5	87.20	66.11	74.97	35.20	65.00	45.42	65.87	60.20
175	236.71	23.8	12.2	3.5	6.5	87.20	66.11	74.97	35.20	65.00	45.42	65.87	60.20
176	237.97	23.8	12.2	3.5	6.5	87.20	66.11	74.97	35.20	65.00	45.42	65.87	60.20
177	239.24	23.8	12.2	3.5	6.5	87.20	66.11	74.97	35.20	65.00	45.42	65.87	60.20
178	240.51	23.8	12.2	3.5	6.5	87.20	66.11	74.97	35.20	65.00	45.42	65.87	60.20
179	241.77	23.8	12.2	3.5	6.5	87.20	66.11	74.97	35.20	65.00	45.42	65.87	60.20
180	243.04	23.8	12.2	3.5	6.5	87.20	66.11	74.97	35.20	65.00	45.42	65.87	60.20
181	244.30	23.8	12.2	3.5	6.5	87.20	66.11	74.97	35.20	65.00	45.42	65.87	60.20
182	245.57	24.1	11.9	3.6	6.4	87.02	66.94	75.45	35.39	64.00	45.35	66.30	60.40
183	246.84	24.1	11.9	3.6	6.4	87.02	66.94	75.45	35.39	64.00	45.35	66.30	60.40
184	248.10	24.1	11.9	3.6	6.4	87.02	66.94	75.45	35.39	64.00	45.35	66.30	60.40
185	249.37	24.4	11.6	3.6	6.4	87.20	67.78	76.14	35.78	64.00	45.73	66.96	60.93
186	250.63	24.4	11.6	3.6	6.4	87.20	67.78	76.14	35.78	64.00	45.73	66.96	60.93
187	251.90	24.5	11.5	3.6	6.4	87.25	68.06	76.32	36.00	64.00	45.90	67.17	61.11
188	253.16	24.5	11.5	3.6	6.4	87.25	68.06	76.32	36.00	64.00	45.90	67.17	61.11
189	254.43	24.5	11.5	3.6	6.4	87.25	68.06	76.32	36.00	64.00	45.90	67.17	61.11
190	255.70	24.7	11.3	3.7	6.3	87.02	68.61	76.61	36.00	63.00	45.67	67.39	61.14
191	256.96	24.7	11.3	3.7	6.3	87.02	68.61	76.61	36.00	63.00	45.67	67.39	61.14
192	258.23	24.7	11.3	3.7	6.3	87.02	68.61	76.61	36.00	63.00	45.67	67.39	61.14

Continua en la siguiente página

Tabla A.19 – Continuación de la página anterior

						Clase CN			Clase MCI				F1 Sc
Add	Add%	TN	FP	FN	TP	Prec	Recl	F1 Sc	Prec	Recl	F1 Sc	Accur	AVG
193	259.49	24.7	11.3	3.7	6.3	87.02	68.61	76.61	36.00	63.00	45.67	67.39	61.14
194	260.76	24.7	11.3	3.7	6.3	87.02	68.61	76.61	36.00	63.00	45.67	67.39	61.14
195	262.03	24.6	11.4	3.7	6.3	86.98	68.33	76.42	35.79	63.00	45.49	67.17	60.96
196	263.29	24.5	11.5	3.7	6.3	86.94	68.06	76.23	35.56	63.00	45.31	66.96	60.77
197	264.56	24.5	11.5	3.7	6.3	86.94	68.06	76.23	35.56	63.00	45.31	66.96	60.77
198	265.82	24.5	11.5	3.7	6.3	86.94	68.06	76.23	35.56	63.00	45.31	66.96	60.77
199	267.09	24.5	11.5	3.7	6.3	86.94	68.06	76.23	35.56	63.00	45.31	66.96	60.77
200	268.35	24.5	11.5	3.7	6.3	86.94	68.06	76.23	35.56	63.00	45.31	66.96	60.77
201	269.62	24.5	11.5	3.7	6.3	86.94	68.06	76.23	35.56	63.00	45.31	66.96	60.77
202	270.89	24.5	11.5	3.7	6.3	86.94	68.06	76.23	35.56	63.00	45.31	66.96	60.77
203	272.15	24.5	11.5	3.7	6.3	86.94	68.06	76.23	35.56	63.00	45.31	66.96	60.77
204	273.42	24.6	11.4	3.7	6.3	86.99	68.33	76.41	35.80	63.00	45.48	67.17	60.94
205	274.68	24.6	11.4	3.7	6.3	86.99	68.33	76.41	35.80	63.00	45.48	67.17	60.94
206	275.95	24.6	11.4	3.7	6.3	86.99	68.33	76.41	35.80	63.00	45.48	67.17	60.94
207	277.22	24.6	11.4	3.7	6.3	86.99	68.33	76.41	35.80	63.00	45.48	67.17	60.94
208	278.48	24.6	11.4	3.7	6.3	86.99	68.33	76.41	35.80	63.00	45.48	67.17	60.94
209	279.75	24.6	11.4	3.8	6.2	86.69	68.33	76.29	35.44	62.00	44.92	66.96	60.60
210	281.01	24.6	11.4	3.8	6.2	86.69	68.33	76.29	35.44	62.00	44.92	66.96	60.60
211	282.28	24.6	11.4	3.8	6.2	86.69	68.33	76.29	35.44	62.00	44.92	66.96	60.60
212	283.54	24.6	11.4	3.7	6.3	86.99	68.33	76.41	35.81	63.00	45.49	67.17	60.95
213	284.81	24.6	11.4	3.7	6.3	86.99	68.33	76.41	35.81	63.00	45.49	67.17	60.95
214	286.08	24.6	11.4	3.7	6.3	86.99	68.33	76.41	35.81	63.00	45.49	67.17	60.95
215	287.34	24.6	11.4	3.7	6.3	86.99	68.33	76.41	35.81	63.00	45.49	67.17	60.95
216	288.61	24.5	11.5	3.6	6.4	87.24	68.06	76.33	35.99	64.00	45.90	67.17	61.11
217	289.87	24.5	11.5	3.6	6.4	87.24	68.06	76.33	35.99	64.00	45.90	67.17	61.11
218	291.14	24.5	11.5	3.5	6.5	87.53	68.06	76.45	36.37	65.00	46.48	67.39	61.47
219	292.41	24.4	11.6	3.4	6.6	87.87	67.78	76.35	36.52	66.00	46.81	67.39	61.58
220	293.67	24.6	11.4	3.5	6.5	87.64	68.33	76.61	36.59	65.00	46.61	67.61	61.61
221	294.94	24.6	11.4	3.4	6.6	87.95	68.33	76.73	36.95	66.00	47.17	67.83	61.95
222	296.20	24.5	11.5	3.4	6.6	87.90	68.06	76.52	36.80	66.00	47.03	67.61	61.77
223	297.47	24.5	11.5	3.4	6.6	87.90	68.06	76.52	36.80	66.00	47.03	67.61	61.77
224	298.73	24.5	11.5	3.4	6.6	87.90	68.06	76.52	36.80	66.00	47.03	67.61	61.77
225	300.00	24.5	11.5	3.4	6.6	87.90	68.06	76.52	36.80	66.00	47.03	67.61	61.77
226	301.27	24.5	11.5	3.4	6.6	87.90	68.06	76.52	36.80	66.00	47.03	67.61	61.77
227	302.53	24.5	11.5	3.4	6.6	87.90	68.06	76.52	36.80	66.00	47.03	67.61	61.77
228	303.80	24.5	11.5	3.4	6.6	87.90	68.06	76.52	36.80	66.00	47.03	67.61	61.77
229	305.06	24.5	11.5	3.4	6.6	87.90	68.06	76.52	36.80	66.00	47.03	67.61	61.77
230	306.33	24.5	11.5	3.4	6.6	87.90	68.06	76.52	36.80	66.00	47.03	67.61	61.77
231	307.59	24.5	11.5	3.4	6.6	87.90	68.06	76.52	36.80	66.00	47.03	67.61	61.77

Continua en la siguiente página

Tabla A.19 – Continuación de la página anterior

						Clase CN			Clase MCI				F1 Sc
Add	Add%	TN	FP	FN	TP	Prec	Recl	F1 Sc	Prec	Recl	F1 Sc	Accur	AVG
232	308.86	24.5	11.5	3.4	6.6	87.90	68.06	76.52	36.80	66.00	47.03	67.61	61.77
233	310.13	24.6	11.4	3.4	6.6	87.94	68.33	76.70	37.02	66.00	47.20	67.83	61.95
234	311.39	24.6	11.4	3.4	6.6	87.94	68.33	76.70	37.02	66.00	47.20	67.83	61.95
235	312.66	24.6	11.4	3.4	6.6	87.94	68.33	76.70	37.02	66.00	47.20	67.83	61.95
236	313.92	24.6	11.4	3.4	6.6	87.94	68.33	76.70	37.02	66.00	47.20	67.83	61.95
237	315.19	24.6	11.4	3.4	6.6	87.94	68.33	76.70	37.02	66.00	47.20	67.83	61.95
238	316.46	24.5	11.5	3.4	6.6	87.89	68.06	76.52	36.78	66.00	47.03	67.61	61.78
239	317.72	24.5	11.5	3.3	6.7	88.17	68.06	76.64	37.20	67.00	47.65	67.83	62.15
240	318.99	24.5	11.5	3.3	6.7	88.17	68.06	76.64	37.20	67.00	47.65	67.83	62.15

En la Tabla A.19, la información se encuentra distribuida de la siguiente manera; en la primera columna los elementos agregados al conjunto de Train, en la segunda columna, el *Balance Rate* en porcentaje, que surge de dividir el número actual de elementos en el conjunto de Train sobre sobre el número total de elementos objetivos, esto multiplicado por 100. Las columnas 3, 4, 5 y 6, corresponden a las variables de la matriz de confusión, TP representa a los aciertos del clasificador en la clase de interés y TN corresponden a los aciertos del clasificador en la clase de control "CN"; luego encontramos las métricas de la matriz de clasificación por clase (Precision, Recall y F1 Score), en la columnas 7, 8 y 9, las correspondientes a la clase CN y en las columnas 10, 11 y 12 la clase de interés; ya por último encontramos la columna del Accuracy y el F1 Average, respectivamente.

Tabla A.20: Evaluación de la clasificación Test SubADReSSo, añadiendo datos sintéticos con WGAN-GP(CNN1D) a la clase MCI.

						Clase CN			Clase MCI				F1 Sc
Add	Add%	TN	FP	FN	TP	Prec	Recl	F1 Sc	Prec	Recl	F1 Sc	Accur	AVG
0	15.19	36	0	10	0	78.26	100.00	87.80	0.00	0.00	0.00	78.26	43.90
1	16.46	30.1	5.9	7.6	2.4	80.22	83.61	80.96	17.33	24.00	19.29	70.65	50.12
2	17.72	27.6	8.4	6.6	3.4	81.04	76.67	77.92	23.72	34.00	26.54	67.39	52.23
3	18.99	23.5	12.5	5.2	4.8	81.90	65.28	72.61	27.75	48.00	35.10	61.52	53.86
4	20.25	23.5	12.5	5.2	4.8	81.90	65.28	72.61	27.75	48.00	35.10	61.52	53.86
5	21.52	23.4	12.6	5.2	4.8	81.84	65.00	72.42	27.58	48.00	34.98	61.30	53.70
6	22.78	23.1	12.9	5.1	4.9	81.93	64.17	71.96	27.50	49.00	35.22	60.87	53.59
7	24.05	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
8	25.32	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
9	26.58	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
10	27.85	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
11	29.11	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
12	30.38	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
13	31.65	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
14	32.91	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
15	34.18	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
16	35.44	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
17	36.71	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
18	37.97	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
19	39.24	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
20	40.51	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
21	41.77	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
22	43.04	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
23	44.30	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
24	45.57	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
25	46.84	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
26	48.10	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
27	49.37	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
28	50.63	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
29	51.90	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
30	53.16	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
31	54.43	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
32	55.70	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
33	56.96	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
34	58.23	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
35	59.49	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
36	60.76	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79

Continua en la siguiente página

Tabla A.20 – Continuación de la página anterior

						Clase CN			Clase MCI				F1 Sc
Add	Add%	TN	FP	FN	TP	Prec	Recl	F1 Sc	Prec	Recl	F1 Sc	Accur	AVG
37	62.03	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
38	63.29	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
39	64.56	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
40	65.82	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
41	67.09	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
42	68.35	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
43	69.62	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
44	70.89	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
45	72.15	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
46	73.42	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
47	74.68	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
48	75.95	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
49	77.22	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
50	78.48	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
51	79.75	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
52	81.01	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
53	82.28	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
54	83.54	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
55	84.81	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
56	86.08	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
57	87.34	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
58	88.61	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
59	89.87	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
60	91.14	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
61	92.41	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
62	93.67	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
63	94.94	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
64	96.20	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
65	97.47	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
66	98.73	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
67	100.00	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
68	101.27	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
69	102.53	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
70	103.80	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
71	105.06	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
72	106.33	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
73	107.59	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
74	108.86	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
75	110.13	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79

Continua en la siguiente página

Tabla A.20 – Continuación de la página anterior

						Clase CN			Clase MCI				F1 Sc
Add	Add%	TN	FP	FN	TP	Prec	Recl	F1 Sc	Prec	Recl	F1 Sc	Accur	AVG
76	111.39	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
77	112.66	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
78	113.92	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
79	115.19	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
80	116.46	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
81	117.72	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
82	118.99	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
83	120.25	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
84	121.52	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
85	122.78	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
86	124.05	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
87	125.32	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79

En la Tabla A.20, la información se encuentra distribuida de la siguiente manera; en la primera columna los elementos agregados al conjunto de Train, en la segunda columna, el *Balance Rate* en porcentaje, que surge de dividir el número actual de elementos en el conjunto de Train sobre sobre el número total de elementos objetivos, esto multiplicado por 100. Las columnas 3, 4, 5 y 6, corresponden a las variables de la matriz de confusión, TP representa a los aciertos del clasificador en la clase de interés y TN corresponden a los aciertos del clasificador en la clase de control "CN"; luego encontramos las métricas de la matriz de clasificación por clase (Precision, Recall y F1 Score), en la columnas 7, 8 y 9, las correspondientes a la clase CN y en las columnas 10, 11 y 12 la clase de interés; ya por último encontramos la columna del Accuracy y el F1 Average, respectivamente.

Tabla A.21: Evaluación de la clasificación Test ADReSSo, añadiendo datos sintéticos con cWGAN-GP(CNN1D) a la clase MCI.

						Clase CN			Clase MCI				F1 Sc
Add	Add%	TN	FP	FN	TP	Prec	Recl	F1 Sc	Prec	Recl	F1 Sc	Accur	AVG
0	15.19	36	0	10	0	78.26	100.00	87.80	0.00	0.00	0.00	78.26	43.90
1	16.46	34	2	9.3	0.7	78.65	94.44	85.42	7.78	7.00	6.36	75.43	45.89
2	17.72	28.6	7.4	7.4	2.6	79.77	79.44	78.80	20.50	26.00	21.12	67.83	49.96
3	18.99	24.7	11.3	5.7	4.3	81.32	68.61	74.26	27.57	43.00	33.27	63.04	53.77
4	20.25	24.7	11.3	5.7	4.3	81.32	68.61	74.26	27.57	43.00	33.27	63.04	53.77
5	21.52	24.3	11.7	5.3	4.7	82.09	67.50	73.97	28.98	47.00	35.63	63.04	54.80
6	22.78	24.2	11.8	5.3	4.7	82.04	67.22	73.81	28.68	47.00	35.46	62.83	54.63
7	24.05	23.6	12.4	5.2	4.8	81.95	65.56	72.77	28.06	48.00	35.28	61.74	54.03
8	25.32	23.5	12.5	5.1	4.9	82.16	65.28	72.69	28.33	49.00	35.78	61.74	54.23
9	26.58	23.5	12.5	5.1	4.9	82.16	65.28	72.69	28.33	49.00	35.78	61.74	54.23
10	27.85	23.5	12.5	5.1	4.9	82.16	65.28	72.69	28.33	49.00	35.78	61.74	54.23
11	29.11	23.5	12.5	5.1	4.9	82.16	65.28	72.69	28.33	49.00	35.78	61.74	54.23
12	30.38	23.5	12.5	5.1	4.9	82.16	65.28	72.69	28.33	49.00	35.78	61.74	54.23
13	31.65	23.1	12.9	5.1	4.9	81.93	64.17	71.96	27.50	49.00	35.22	60.87	53.59
14	32.91	23.1	12.9	5.1	4.9	81.93	64.17	71.96	27.50	49.00	35.22	60.87	53.59
15	34.18	23.1	12.9	5.1	4.9	81.93	64.17	71.96	27.50	49.00	35.22	60.87	53.59
16	35.44	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
17	36.71	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
18	37.97	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
19	39.24	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
20	40.51	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
21	41.77	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
22	43.04	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
23	44.30	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
24	45.57	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
25	46.84	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
26	48.10	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
27	49.37	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
28	50.63	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
29	51.90	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
30	53.16	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
31	54.43	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
32	55.70	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
33	56.96	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
34	58.23	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
35	59.49	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
36	60.76	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79

Continua en la siguiente página

Tabla A.21 – Continuación de la página anterior

						Clase CN			Clase MCI				F1 Sc
Add	Add%	TN	FP	FN	TP	Prec	Recl	F1 Sc	Prec	Recl	F1 Sc	Accur	AVG
37	62.03	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
38	63.29	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
39	64.56	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
40	65.82	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
41	67.09	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
42	68.35	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
43	69.62	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
44	70.89	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
45	72.15	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
46	73.42	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
47	74.68	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
48	75.95	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
49	77.22	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
50	78.48	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
51	79.75	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
52	81.01	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
53	82.28	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
54	83.54	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
55	84.81	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
56	86.08	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
57	87.34	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
58	88.61	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
59	89.87	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
60	91.14	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
61	92.41	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
62	93.67	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
63	94.94	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
64	96.20	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
65	97.47	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
66	98.73	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
67	100.00	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
68	101.27	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
69	102.53	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
70	103.80	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
71	105.06	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
72	106.33	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
73	107.59	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
74	108.86	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
75	110.13	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79

Continua en la siguiente página

Tabla A.21 – Continuación de la página anterior

						Clase CN			Clase MCI				F1 Sc
Add	Add%	TN	FP	FN	TP	Prec	Recl	F1 Sc	Prec	Recl	F1 Sc	Accur	AVG
76	111.39	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
77	112.66	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
78	113.92	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
79	115.19	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
80	116.46	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
81	117.72	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
82	118.99	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
83	120.25	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
84	121.52	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
85	122.78	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
86	124.05	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79
87	125.32	23	13	5	5	82.14	63.89	71.88	27.78	50.00	35.71	60.87	53.79

En la Tabla A.21, la información se encuentra distribuida de la siguiente manera; en la primera columna los elementos agregados al conjunto de Train, en la segunda columna, el *Balance Rate* en porcentaje, que surge de dividir el número actual de elementos en el conjunto de Train sobre sobre el número total de elementos objetivos, esto multiplicado por 100. Las columnas 3, 4, 5 y 6, corresponden a las variables de la matriz de confusión, TP representa a los aciertos del clasificador en la clase de interés y TN corresponden a los aciertos del clasificador en la clase de control "CN"; luego encontramos las métricas de la matriz de clasificación por clase (Precision, Recall y F1 Score), en la columnas 7, 8 y 9, las correspondientes a la clase CN y en las columnas 10, 11 y 12 la clase de interés; ya por último encontramos la columna del Accuracy y el F1 Average, respectivamente.

APÉNDICE B

EXTRACCIÓN CLÁSICA DE CARACTERÍSTICAS ACÚSTICAS

A partir de los trabajos de participantes en el reto ADReSSo se retuvieron aquellas técnicas más prometedoras. Entre ellas, están: (i) el conjunto de características eGeMAPS propuesto para extraer rasgos emocionales; (ii) los primeros 20 coeficientes cepstrales en las frecuencias de Mel (MFCC, por sus siglas en Inglés), que se orientan a las características auditivas dentro del rango del habla humana; (iii) los deltas de los primeros 20 coeficientes MFCC, que permiten ver el cambio dentro de estos coeficientes y (iv) los coeficientes Chroma orientados a describir la curva melódica del audio. Los últimos tres conjuntos de características se trabajaron calculando los estadísticos funcionales[1] de las series obtenidas al calcular los coeficientes por segmentos temporales[2].

Como se mencionó, se escogieron estos conjuntos eGeMAPS y MFCC debido a su amplio uso en la bibliografía, y se incluyó en este análisis las características Chroma, con la finalidad de observar su comportamiento, puesto que estos co-

[1] Se calcularon el promedio, desviación estándar, mediana, curtosis y asimetría de cada coeficiente obtenido para la serie temporal de datos que representa cada audio.

[2] Para el cálculo del conjunto de eGeMaps y Chroma, se utilizó la librería de **OpenSmile** y para los coeficientes de MFCC y sus Deltas se usó la librería **Librosa** para python.

eficientes caracterizan los espectros de semitonos con deformación de octava, reconocimiento de acordes y tonalidades. Se incluyó esta caracterización dada la posible aparición de la disfonía (alteración del timbre de voz) en los pacientes enfermos.

MFCC: Coeficientes Cepstrales de Frecuencias Mel

Esta técnica permite capturar las propiedades más relevantes de una señal de voz al modelar cómo el oído humano percibe las frecuencias del habla.

En el código, se carga la señal de audio utilizando la librería `librosa` con el método de reescalado 'soxr_vhq', el archivo de audio es cargado mediante la función `librosa.load`, que devuelve dos valores: la señal de audio y la frecuencia de muestreo. La longitud de la ventana de análisis (`frame`) se establece en 25 ms, y el solapamiento (`step`) en 10 ms. Estos parámetros son los comúnmente usados para el análisis de voz (16). La extracción de coeficientes MFCC también se complemento con el cálculo de los *delta* de dichos coeficientes. Tal como se usa para capturar la dinámica temporal de las características del habla (16).

Posteriormente, se extraen la cantidad de coeficientes cepstrales de Mel que se requieren mediante la función `librosa.feature.mfcc`, especificando el número de coeficientes (`n_mfcc=20`) y los parámetros de la Transformada de Fourier, para obtener una matriz de 20 coeficientes MFCC, donde cada columna corresponde a un marco temporal de la señal.

Una vez obtenidos los coeficientes, estos son normalizados mediante una normalización de media y varianza (Cepstral Mean and Variance Normalization - CMVN) usando `speechpy.processing.cmvn(matriz_mfcc.transpose(), variance_normalization = True)` , que ajusta los coeficientes MFCC para que tengan media cero y varianza unitaria. Esto ayuda a reducir las diferencias entre señales de diferentes grabaciones, eliminando efectos de ruido o de grabación no deseados, donde `variance_normalization`

`= True`, indica que se aplicó tanto normalización de la media como de la varianza, no solo ajustando el promedio de los coeficientes, sino también su dispersión.

Finalmente, a los coeficientes extraídos se les calcula el promedio, desviación estándar, mediana, curtosis y asimetría de cada coeficiente para la serie temporal de datos que representa cada audio, tal como se explica en la 2.1.

Deltas de MFCC

El cálculo de las deltas, o derivadas temporales de los coeficientes MFCC, mide cómo cambian los MFCC a lo largo del tiempo. Estos cambios pueden ser cruciales para capturar la dinámica de la señal de voz, como las variaciones rápidas que ocurren durante el habla. Se utilizo la función `librosa.feature.delta`, la cual calcula estas deltas utilizando la matriz de coeficientes MFCC producida por la función del cálculo anterior `librosa.feature.mfcc`, lo que permite identificar las variaciones en la señal sobre un periodo de tiempo, lo cual aporta información sobre la evolución temporal de la señal. Posteriormente los resultados obtenidos de esta función pasan por la misma función de normalización y cálculo de estadísticos funcionales por la cual paso el cálculo de los coeficientes MFCC.

Chroma

Para la extracción de características de *Chroma*, se utilizó la librería *OpenSmile*, donde se tuvo que interpretar la codificación interna de esta librería para generar un archivo de configuración, basados en las instrucciones que se encuentran en la librería de *OpenSmile* para otro conjunto de características más grande. Este código generado configura un flujo de procesamiento para la extracción de características *Chroma* a partir de señales de audio utilizando módulos internos de esta librería. Primero, el componente `cFramer` divide la señal en marcos de 64 ms con un sola-

pamiento de 10 ms. Luego, el componente `cWindower` aplica una ventana gaussiana a cada marco para suavizar los bordes. Posteriormente, `cTransformFFT` realiza una Transformada rápida de Fourier (FFT), convirtiendo los marcos al dominio de la frecuencia. A continuación, `cFFTmagphase` extrae la magnitud del espectro, y el componente `cTonespec` genera un espectro escalado en semitonos, cubriendo seis octavas. Finalmente, `cChroma` agrupa las frecuencias en 12 clases de semitonos para calcular las características *Chroma*.

Al finalizar esta etapa del código, los datos pasan por la misma función de normalización y cálculo de estadísticos funcionales, que se usó para el cálculo de los coeficientes de MFCC.

eGeMaps

El *Geneva Minimalistic Parameter Set* (GeMaps) es un subconjunto de características acústicas cuidadosamente seleccionadas para capturar información esencial sobre el comportamiento de la voz, optimizado para la eficiencia y relevancia en la investigación de voz, afectividad y salud mental. Este conjunto contiene 18 descriptores de bajo nivel LLD por sus siglas en Inglés, *Low-Level Descriptors*, organizados en tres categorías principales: parámetros relacionados con la frecuencia, la energía/amplitud, y el contenido espectral. A continuación, se detallan técnicamente los descriptores de cada categoría. Se aplican diferentes estadísticos funcionales a cada uno de los elementos, para hacerlo mas sencillo de leer y conocer que datos estadísticos se extraen de cada característica, se escribirá al final de cada definición el acrónimo correspondiente.

- AM: *Arithmetic Mean* (Media Aritmética).
- CV: *Coefficient of Variation* (Coeficiente de Variación, que es la desviación

estándar normalizada por la media aritmética).

- P20: 20-th *Percentile* (percentil 20).
- P50: 50-th *Percentile* (percentil 50 o mediana).
- P80: 80-th *Percentile* (percentil 80).
- R20-80: *Range of 20-th to 80-th Percentile* (rango entre el percentil 20 y el percentil 80).
- MSlopeR: *Mean of Rising Slope* (media de la pendiente de las partes ascendentes de la señal).
- MSlopeF: *Mean of Falling Slope* (media de la pendiente de las partes descendentes de la señal).
- SSlopeR: *Standard Deviation of Rising Slope* (desviación estándar de la pendiente de las partes ascendentes de la señal).
- SSlopeF: *Standard Deviation of Falling Slope* (desviación estándar de la pendiente de las partes descendentes de la señal).
- AMU: *Arithmetic Mean over Unvoiced segments* (la media aritmética se calcula únicamente sobre los segmentos no vocalizados de la señal).
- AMV: *Arithmetic Mean in Voiced regions* (media aritmética en las regiones vocalizadas).
- CVV: *Coefficient of Variation in Voiced regions* (coeficiente de variación en las regiones vocalizadas).

1. Parámetros relacionados con la frecuencia: Estos parámetros se centran en la modulación de la frecuencia fundamental (F0) y la resonancia vocal:

- *Pitch* (F0): La frecuencia fundamental de la voz se mide en una escala logarítmica de semitonos, comenzando en 27.5 Hz. Es crucial para la detección de entonación y patrones prosódicos. AM, CV, P20, P50, P80, R20-80, MSlopeR, MSlopeF, SSlopeR, SSlopeF.
- *Jitter*: Describe la variación ciclo a ciclo en la duración de los periodos de F0. Se utiliza para detectar inestabilidades en la producción de la voz, típicamente asociadas con patologías vocales o estados afectivos. AM, CV.
- Formantes 1, 2 y 3 (F1, F2, F3): Frecuencia de los tres primeros formantes, que son picos de energía en el espectro acústico causados por resonancias en el tracto vocal. Proporcionan información sobre la configuración del tracto vocal y son útiles para analizar la calidad de la voz. AM, CV.
- Ancho de banda del Formante 1 (F1BW): Captura la dispersión de la energía en torno al primer formante, relacionada con la claridad y apertura de las vocales. AM, CV.

2. Parámetros relacionados con la energía/amplitud: Estos descriptores cuantifican la variación en la intensidad de la voz y las diferencias en la amplitud entre ciclos de F0:

    - Shimmer: Variación ciclo a ciclo en la amplitud de la onda acústica. Un shimmer elevado puede indicar desórdenes en la producción vocal o tensiones emocionales. AM, CV.
    - Loudness: Estimación perceptual de la intensidad de la voz basada en la curva de respuesta del oído humano. La medida se deriva del espectro auditivo y es útil para identificar emociones como la *ira* o la *excitación*. AM, CV, P20, P50, P80, R20-80, MSlopeR, MSlopeF, SSlopeR, SSlopeF.

- HNR (Harmonics-to-Noise Ratio): Relación entre la energía en los componentes armónicos de la señal y la energía de componentes aleatorios (ruido). Es un indicador de la calidad vocal y la presencia de ruido en la voz. AM, CV.

3. Parámetros espectrales (balance y relación armónica): Esta categoría evalúa el balance de energía en diferentes regiones del espectro y la relación entre los armónicos de la señal de voz:

    - Alpha Ratio: Relación entre la energía en las bandas de frecuencia baja (50–1000 Hz) y alta (1–5 kHz). Esta medida es clave para identificar variaciones en la calidad vocal. AM, CV, AMU.
    - Índice Hammarberg: Relación entre el pico más fuerte de energía en la región de 0-2 kHz y el pico más fuerte en la región de 2-5 kHz. Es útil para evaluar el brillo de la voz y la resonancia del tracto vocal. AM, CV.
    - Pendiente espectral 0-500 Hz y 500-1500 Hz: Pendiente obtenida de la regresión lineal aplicada al espectro de potencia en las bandas de 0-500 Hz y 500-1500 Hz. Describe cómo se distribuye la energía en las frecuencias bajas y medias, útil para capturar el balance entre los sonidos agudos y graves. AM, CV, AMU.
    - Energía relativa de los formantes 1, 2 y 3: Relación entre la energía en los formantes F1, F2 y F3 con la energía en el pico fundamental (F0). Es importante para describir la distribución de energía en las vocales. AM, CV.
    - Diferencia armónica H1-H2: Relación de amplitud entre el primer armónico (H1) y el segundo armónico (H2). Un indicador de la calidad vocal, típicamente asociado a la tensión en las cuerdas vocales. AM, CV.
    - Diferencia armónica H1-A3: Relación entre el primer armónico (H1) y el armónico más fuerte en la región del tercer formante (A3). Esta medida

es útil para caracterizar la resonancia del tracto vocal y la articulación. AM, CV.

Además, se incluyen 6 características temporales:

- Tasa de picos de sonoridad: Es el número de picos de sonoridad por segundo.
- Duración media y desviación estándar de las regiones sonoras: Se refiere a la duración de los segmentos con F0>0.
- Duración media y desviación estándar de las regiones no sonoras: Se refiere a la duración de los segmentos con F0=0, aproximando pausas.
- Tasa de regiones sonoras: Es el número de regiones sonoras (segmentos con F0>0) por segundo, similar a la tasa de sílabas.

El *"Extended Parameter Set"*, con el cual se conformaría el eGeMaps, es un conjunto extendido de parámetros acústicos diseñados para complementar el conjunto minimalista. Este conjunto extendido se propone debido a que el minimalista no incluye parámetros cepstrales ni muchos parámetros dinámicos, que son clave para modelar estados afectivos.

El conjunto extendido agrega 7 descriptores de bajo nivel (LLD) a los 18 ya presentes en el conjunto minimalista:

4. Parámetros espectrales (balance, forma, dinámica):
    - Coeficientes cepstrales en la frecuencia Mel (MFCC) 1-4. AM, CV, AMV.
    - Flujo espectral, que mide la diferencia entre los espectros de dos fotogramas consecutivos. AM, CV, AMU, AMV, CVV.
5. Parámetros relacionados con la frecuencia:
    - Ancho de banda de los formantes 2 y 3, para completar los parámetros de los formantes 1-3. AM, CV.

6. Adicionalmente, se incluye el nivel sonoro equivalente. Esto da como resultado 26 parámetros adicionales. En total, al combinarse con el conjunto minimalista, el conjunto extendido de parámetros acústicos de Geneva (eGeMAPS) comprende 88 parámetros en su totalidad.

APÉNDICE C

WGAN-GP

Arquitectura General de los Modelos WGAN-GP Implementados

A continuación, se realiza la descripción técnica de cada una de las partes de los modelos implementados para generación de datos de interés basado en el modelo WGAN-GP. La WGAN-GP (CNN1D) y cWGAN-GP (CNN1D) cuentan con una estructura de redes convolucionales y convoluciones transpuestas unidimensionales, lo que permitiría un análisis más detallado e interrelacionado de los datos por parte de las redes generadora y discriminadora.

Dentro las siguientes descripciones se hará espacio a secciones donde se explicará como se implementó para las diferentes versiones de WGAN-GP, ya sea WGAN-GP(CNN1D) o cWGAN-GP(CNN1D). En las secciones donde no se haga esta distinción, querrá decir que los modelos comparten estas características.

- **Configuración del Modelo WGAN-GP** La configuración global del modelo se estableció con los siguientes hiperparámetros clave:
  - **Parámetros de Entrenamiento**:
    * `n_epochs = 3000`: Número de épocas de entrenamiento.
    * `batch_size = 6`: Tamaño de los lotes de datos.

    * `lr = 0.0001`: Tasa de aprendizaje para el optimizador Adam.
    * `b1 = 0.5` y `b2 = 0.999`: Parámetros beta del optimizador Adam, que controlan las tasas de decaimiento de los momentos.
    * `sample_interval = 5`: Intervalo de épocas para guardar puntos de control y muestras.
  - **Parámetros del Modelo**:
    * `latent_dim = 50`: Dimensionalidad del espacio latente (vector de ruido).
    * `n_class = 1 ó 2`: Número de clases, dependiendo si se trata de la WGAN-GP o la cWGAN-GP.
    * `vec_size = 512`: Tamaño de la dimensión de salida (se refiere al tamaño del vector de características).
    * `lambda_gp = 10`: Coeficiente para la penalización de gradiente en WGAN-GP.
  - **Parámetros de Entrenamiento del Discriminador y Generador**:
    * `n_critic = 1`: Número de iteraciones de entrenamiento del discriminador por cada iteración del generador.
    * `delay_gen = 5`: Número de lotes que deben pasar antes de entrenar el generador.

  Esta configuración se seleccionó tras pruebas preliminares para equilibrar la capacidad de aprendizaje del modelo y mantener la estabilidad durante el entrenamiento.

- **Modelo del Generador** Como se explicaba en la sección del marco teórico, el objetivo del generador es mapear un vector de ruido latente y una etiqueta de clase a una muestra sintética que siga la distribución de los datos reales.
  - **Input:**

    * `z`: Vector de ruido latente de tamaño 50 `latent_dim`.
    * `labels`: Etiquetas de clase (0 para CN, 1 para AD).

- **Proceso de Generación**: Para el proceso de generación primero se genera un *embedding* de la etiqueta, luego se concatena el vector de ruido Z con el *embedding* de la etiqueta, estos dos pasos, solo son para las cWGAN-GP, de lo contrario ingresaría el vector de ruido a la red como tal, para pasar por último por la función de activación, que para todas es una `Tanh`. A continuación, se detalla cada uno de los pasos.

    * **Embedding de Etiquetas**: Se utiliza una capa de *embedding* para transformar las etiquetas discretas en vectores continuos de tamaño 14 para la cWGAN-GP(CNN1D).
    * **Concatenación**: El vector de ruido z se concatena con el *embedding* de la etiqueta, formando un vector de entrada combinado.
    * **Arquitectura de red:**:
        · Se emplean múltiples capas de convolución transpuesta 1D (`ConvTranspose1d`) o capas lineales (`nn.Linear`) según sea el caso, con el fin de aumentar la dimensionalidad del tensor.
        · Cada capa está seguida de una normalización por lotes `BatchNorm1d` y una función de activación `LeakyReLU`.
        · Las dimensiones se incrementan progresivamente, permitiendo al modelo generar datos de mayor complejidad.
    * **Función de Activación Final**: Se utiliza una función `Tanh` para asegurar que los valores de salida estén en el rango [-1, 1], compatible con los datos preprocesados.

- **Detalle de Capas WGAN-GP (CNN1D) y cWGAN-GP (CNN1D)**:

    * **Capa 1**: Una capa de convolución traspuesta unidimensional `ConvTranspose1d` de entrada (64 canales) a una salida de 512*9 canales.

    - **Capas Intermedias**: Se reduce gradualmente el número de canales mientras se aumenta la longitud del vector. Cada capa se le aplica normalización `BatchNorm1d` con momentum de 0.8, seguido de una función de activación `nn.LeakyReLU` con un `negative_slope` de 0.2.
    - **Capa Final**: Una capa de convolución traspuesta unidimensional `ConvTranspose1d` que produce una salida con 9 canales, correspondiente al número de características extraídas.

- **Modelo del Discriminador** El modelo discriminador, busca distinguir entre muestras reales y generadas.

  - **Input:**
    - `Embedding_de_Audio:` El *embedding* del audio, que puede ser real (extraído utilizando *wav2vec*) o generado (producido por el modelo generador de la GAN, ya sea durante el entrenamiento o una vez entrenado).
    - `labels:` Las etiquetas correspondientes a las clases asociadas a los datos.
  - **Proceso de Discriminación:**
    - **Embedding de Etiquetas:** Similar al generador, se utiliza un *embedding* con `embed_size` = 512 en el caso de la cWGAN-GP (CNN1D), puesto que este vector se añade como un canal más en la entrada del Discriminador.
    - **Concatenación:** Para el caso de la cWGAN-GP (CNN1D), el *embedding* de la etiqueta se añade como un canal más de entrada, obteniendo así, 10 canales en total (9 canales, uno para cada estadístico y se le suma el *embedding* de clase en otro canal).
    - **Arquitectura de red:**

· Se aplican múltiples capas de convolución 1D (nn.Conv1d) convolución transpuesta 1D (`ConvTranspose1d`) o capas lineales (`nn.Linear`) según sea el caso, que reducen la dimensionalidad de la matriz o vector ingresado.

· Cada capa utiliza una función de activación `LeakyReLU` para mantener el flujo de gradientes y evitar el problema de gradientes muertos.

* **Salida:** Se obtiene un valor escalar que representa la validez de la muestra, indicando si es real o generada.

– **Detalle de Capas WGAN-GP (CNN1D) y cWGAN-GP (CNN1D):**

* **Capa 1:** Una capa convolucional unidimensional `Conv1d` que toma una entrada con 10 canales, 9 canales para cada uno de los 9 vectores de estadísticos que representan la distribución temporal de las 512 características al cual se le suma el *embedding* de etiqueta en otro canal más, esto para la cWGAN-GP (CNN1D); en el caso de la WGAN-GP (CNN1D) solo son 9 canales, como salida de esta capa se tienen un valor de 24 canales.

* **Capas Intermedias:** Se incrementa el número de canales de salida, como tantas capas haya, mientras se reduce la longitud del vector de salida.

* **Capa Final:** Una capa convolucional unidimensional `Conv1d` que produce una salida escalar.

• **Funciones de Pérdida y Optimización**

– **Función de Pérdida del Discriminador:** La función de pérdida del Discriminador está basada en la diferencia entre las puntuaciones asignadas a muestras reales y generadas, a la cual se añade la penalización de gradiente para asegurar la condición de Lipschitz, como se veía en la sección

del marco teórico.

- **Función de Pérdida del Generador:** Busca maximizar la capacidad del generador para engañar al discriminador. Se utiliza la negación de la puntuación promedio del discriminador sobre las muestras generadas únicamente.
- **Optimizadores:** Se emplea el optimizador Adam para ambos modelos, con tasas de aprendizaje y parámetros β definidos en un inicio de la sección.

Procedimiento de Entrenamiento de la WGAN-GP

A continuación, se describe el proceso de entrenamiento paso a paso de los variantes de WGAN-GP implementadas.

- **Preparación de Datos** Los datos se cargan en un dataload de la siguiente forma:
  - **Recopilación de Archivos:** Se recorren los directorios especificados para obtener rutas de archivos .pt con los vectores de los *embeddings*.
  - **Asignación de Etiquetas:** Se extraen las etiquetas de clase basadas en los nombres de los subdirectorios ('AD' o 'CN'). Se seleccionan los datos de acuerdo al modelo de WGAN-GP que se está trabajando.
- **Bucle de Entrenamiento** El bucle de entrenamiento sigue estos pasos:
  - **Iteración sobre Épocas y Lotes:** Se recorre el número de épocas definido como (`n_epochs`). En cada época, se itera sobre los lotes de datos proporcionados por el `DataLoader`.
  - **Entrenamiento del Discriminador:** Por cada lote, se entrena el discriminador `n_critic` veces antes de actualizar el generador.

    Generación de Muestras Falsas: Se generan muestras falsas utilizando el generador con vectores de ruido y etiquetas de clase.

Evaluación del Discriminador: Se calculan las puntuaciones del discriminador para las muestras reales y generadas.

Cálculo de la Pérdida: Se calcula la pérdida del discriminador como:

$$L_D = -\mathbb{E}_{x\sim P_r}[D(x)] + \mathbb{E}_{\hat{x}\sim P_g}[D(\hat{x})] + \lambda_{gp} \cdot GP$$

donde GP es la penalización de gradiente.

- **Actualización de Parámetros:** Se realiza la retropropagación y se actualizan los pesos del discriminador.
- **Penalización de Gradiente:** Asegura que el discriminador satisfaga la condición de Lipschitz y su cálculo se realiza interpolando muestras reales y generadas, para luego calcular los gradientes de las puntuaciones del discriminador respecto a las muestras interpoladas, con la siguiente fórmula:

$$GP = \mathbb{E}_{\hat{x}\sim P_{\hat{x}}}\left[(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2\right]$$

- **Entrenamiento del Generador:** Después de `n_critic` iteraciones del discriminador, se actualiza el generador.
- **Evaluación del Generador:** Se generan nuevas muestras y se obtienen las puntuaciones del discriminador.
- **Cálculo de la Pérdida:** La pérdida del generador es:

$$L_G = -\mathbb{E}_{\hat{x}\sim P_g}[D(\hat{x})]$$

- **Actualización de Parámetros:** Se realiza la retropropagación y se actualizan los pesos del generador.
- **Monitoreo y Registro:** Se almacenan las pérdidas y métricas relevantes en un `DataFrame` para un posible análisis posterior del comportamiento de entrenamiento, en caso de ser necesario.
- **Guardado de Modelos y Puntos de Control:** Se guardan puntos de control cada cierto número de épocas o al cumplir condiciones específicas. Los

puntos de control incluyen los estados de los modelos y optimizadores, permitiendo reanudar el entrenamiento si es necesario.

- **Manejo de Datos Especiales** Se implementan funciones para eliminar filas y columnas con valores NaN o infinitos. Esto garantiza la integridad de los datos y evita errores durante el entrenamiento.

Para elegir la época de entrenamiento de la cual se tomarían los pesos de los modelos WGAN-GP (CNN1D) y cWGAN-GP (CNN1D), se verificó la gráfica de entrenamiento de estos y se tomó aproximadamente el punto de convergencia de la función de pérdida del generador y discriminador. A continuación se encuentra la gráfica de entrenamiento de la WGAN-GP (CNN1D).

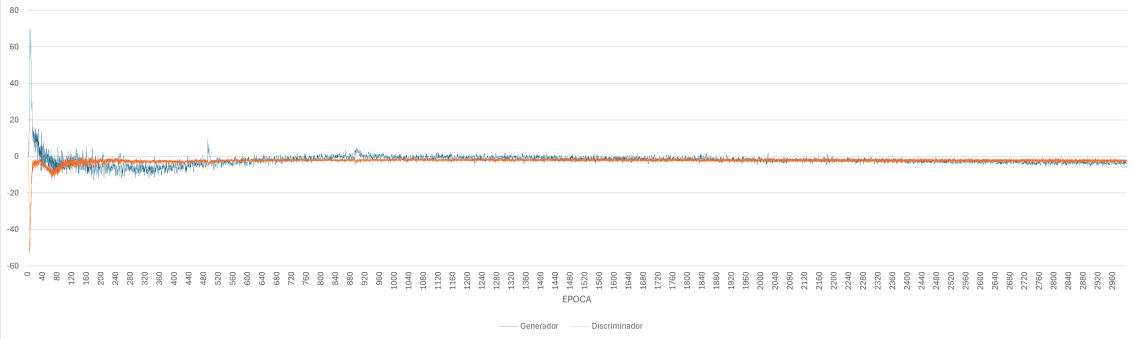


Figura C.1: Gráfica de entrenamiento de la WGAN-GP(CNN1D)

En la Figura C.1 la convergencia se observa alrededor de las 2200 épocas y este *checkpoint* fue el utilizado para generar los datos sintéticos. Ahora podemos apreciar la siguiente gráfica de entrenamiento de la cWGAN-GP (CNN1D).

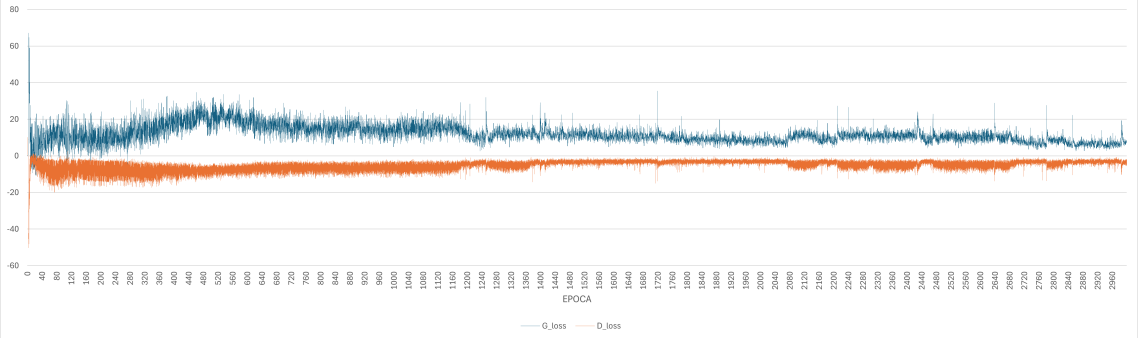


Figura C.2: Gráfica de primer entrenamiento de la cWGAN-GP (CNN1D)

Para la Figura C.2 se observa que luego de 3000 épocas, no se obtuvo la convergencia de las funciones de perdida, es por esto que se entrena el modelo 3000 épocas extra y esto lo podemos observar en la siguiente gráfica.

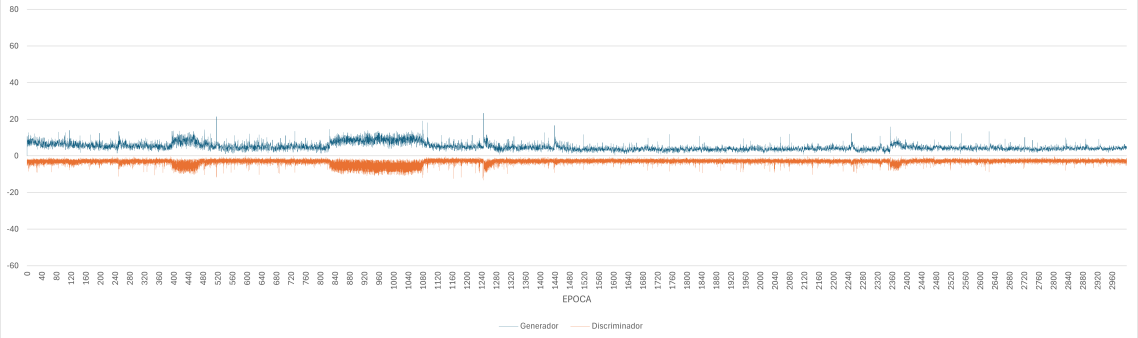


Figura C.3: Gráfica de segundo entrenamiento de la cWGAN-GP (CNN1D)

Como pudimos apreciar en la anterior Figura C.3, no se llegó a la convergencia de las funciones de pérdida, probablemente, por qué no hay suficientes elementos de la clase minoritaria para poder generalizar ambas clases, o las clases son muy similares y no logra generalizarlas. Por lo cual, se tomó la época 2900 de esta segunda parte del entrenamiento para generar los datos sintéticos con el modelo cWGAN-GP (CNN1D).

Generación de Datos Sintéticos

La generación de datos sintéticos se realiza siguiendo estos pasos:

- **Carga del Modelo Entrenado:** Se carga el generador desde el punto de control guardado utilizando `torch.load`.
- **Configuración para la Generación:**
  Parámetros:
  opt.gen_class: Especifica la clase para la que se generarán muestras (0 para CN, 1 para AD).
  opt.n_pruebas: Número de pruebas o iteraciones de generación.

- **Proceso de Generación:**

  - Generación del elemento sintético: Se genera un vector de ruido z y se pasa al generador junto con las etiquetas de clase en caso de ser una cWGAN, de no ser así, solo se envía el vector de ruido. Luego de pasar estos datos por el generador, se obtienen muestras sintéticas que siguen la distribución aprendida.

  - Filtrado con Clasificador: Las muestras generadas se evalúan utilizando un clasificador preentrenado (por ejemplo, SVM o *AdaBoost*). Solo se conservan las muestras que son clasificadas correctamente.

  - Almacenamiento de Muestras: Las muestras clasificadas correctamente se almacenan en un `DataFrame`. Se registra información adicional como el número de prueba y clasificador utilizado.

- **Integración con el Conjunto de Datos:** Los datos sintéticos se agregan al conjunto de datos de entrenamiento del Dataset SubADReSSo, aumentando la cantidad y diversidad de las muestras.

- **Evaluación con el nuevo Conjunto de Datos:** Se evalúa el nuevo conjunto de entrenamiento, cada vez que se añade un dato sintético nuevo; para esto se entrenan varios clasificadores con este nuevo conjunto de entrenamiento con el fin de ver con cuál clasificador se obtiene el mejor comportamiento, evaluando el conjunto de Test SubADReSSo con diferentes métricas.

- **Finalización para la Generación de Datos Sintéticos:** Las iteraciones continúan hasta que se asegura que el conjunto de entrenamiento resultante esté equilibrado en términos de clases.

Apéndice D

Artículos publicados

Los artículos derivados de esta tesis se listan a continuación.

- *Detectando el deterioro cognitivo a través del habla espontánea Detecting cognitive impairment through spontaneous speech.* Migan G. Galban Pineda, Francisco Iván González Hernández, Hugo Jair Escalante, Luis Villaseñor Pineda. Abstraction & Application Vol. 42, Universidad Autónoma de Yucatán, Yucatán, México, 2023.