VOLUMEN 48
XXXI Verano De la Ciencia
ISSN 2395-9797
www.jovenesenlaciencia.ugto.mx

Desarrollo de un Modelo de Soporte al Diagnóstico de Depresión Mediante Procesamiento Digital de Señales de Voz

Developing a Diagnostic Support Model for Depression Through Digital Voice Processing

Teodoro CORDOVA FRAGA*, Huetzin Aaron PEREZ OLIVAS**, Luis Mauricio LÓPEZ JARAMILLO***
División de Ciencias e Ingenierías, Universidad de Guanajuato campus León
Loma del Bosque 103, Lomas del Campestre, 37150 León, Guanajuato, México.
*theo@fisica.ugto.mx, **ha.perez@ugto.mx, *** lm.lopezjaramillo@ugto.mx

Blanca Olivia MURILLO ORTIZ*
Unidad de Investigación Médica, Unidad Médica de Alta Especialidad (UMAE) No. 1 Hospital de Especialidades Centro Médico Nacional del Bajío, OOAD Guanajuato, Instituto Mexicano del Seguro Social (IMSS). Av. Insurgentes y Blvd. Adolfo López Mateos s/n, Col. Los Paraísos, C.P. 37328, León, Guanajuato, México.
*bomo907@hotmail.com

Aldo Paul Barrientos Velázquez[1]
[1]División de Ciencias e Ingenierías
ap.barrientosvelazquez@ugto.mx

Neil Granados Macías[1]
[1]División de Ciencias e Ingenierías
n.granadosmacias@ugto.mx

Leonardo Antonio Almada Malagón[1]
[1]División de Ciencias e Ingenierías
la.almadamalagon@ugto.mx

Resumen

El diagnóstico de la depresión ha dependido tradicionalmente de test escritos y evaluaciones que pueden estar cargadas de un alto grado de subjetividad. Una alternativa es la presente investigación, la cual demuestra que la voz humana contiene información capaz de revelar el estado de salud mental de forma cuantitativa y no invasiva. Mediante el análisis de grabaciones de voz capturadas con teléfonos celulares, se evaluaron distintos parámetros acústicos en pacientes con diagnóstico clínico de depresión y en sujetos sanos. Los resultados estadísticos confirman que el uso de la fuente o frecuencia fundamental (F_0) presentan un sesgo de género, haciéndolo eficaz para un diagnóstico. Por otro lado, el filtro o los Coeficientes Cepstrales en las Frecuencias de Mel (MFCC), demostraron no tener dependencia al género de los hablantes. Además, 8 de los 13 coeficientes evaluados lograron detectar diferencias estadísticamente significativas entre la población sana y la enferma. Se concluye que los MFCC son el biomarcador objetivo ideal para entrenar futuros modelos de IA, abriendo la puesta a diagnósticos accesibles.

Palabras clave: Depresión, Diagnóstico, F_0, MFCC, Biomarcador, IA.

Introducción

La depresión es una enfermedad con alta presencia poblacional, cuya evaluación diagnóstica tradicional (test escritos) presenta un grado elevado de subjetividad por parte del paciente, así como del evaluador **[1,2]**. Investigaciones recientes se han propuesto encontrar métricas cuantitativas y objetivas que favorezcan a su diagnóstico **[3]**. El análisis de voz ha surgido como una alternativa cuantitativa, una herramienta digital clave, prometedora y no invasiva para la evaluación del riesgo suicida y la detección temprana de la depresión **[4,5]**. La evidencia demuestra que las características paralingüísticas, acústicas y la dinámica del habla de los pacientes se alteran de forma medible frente a la presencia de enfermedades mentales y trastornos afectivos severos **[6,7]**.

Estudios en diversas poblaciones confirman la viabilidad de encontrar una asociación estadística entre el estado depresivo mayor y variaciones en los parámetros de la voz de los pacientes, tales como la velocidad de dicción o las variaciones en el tono. Físicamente, la mecánica del habla se analiza separando la fuente y el filtro. Para capturar y digitalizar las características computacionales de este filtro, los algoritmos de

reconocimiento emplean los Coeficientes Cepstrales en las Frecuencias de Mel (MFCC), dada su capacidad de simular la percepción auditiva humana.

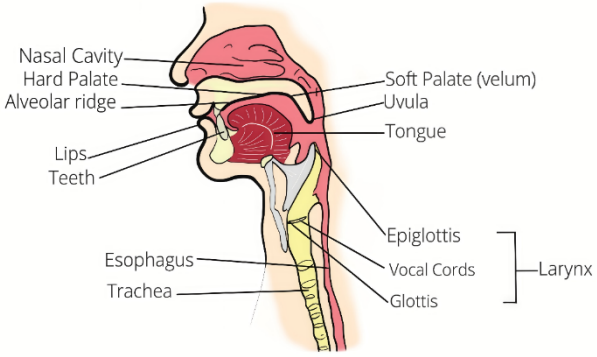


Figura 1. *Tracto vocal: mecánica del habla* **[17]**.

La implementación de la IA, a través de redes neuronales de Machine Learning y Deep Learning, ha validado consistentemente su alta precisión diagnóstica al ser entrenada con muestras de voz y registros de lectura **[8, 9]**. Estas arquitecturas automatizan de forma eficiente la detección y cuantificación de perfiles emocionales y depresivos **[10, 11]**.

El uso de dispositivos comunes para grabar registros acústicos de lectura condicional permite una recolección de datos temprana, altamente accesible y escalable, apoyando el desarrollo sostenible de sistemas médicos de bajo costo **[14, 15]**. La intersección entre el proceso acústico mediante teléfonos celulares, el control de variables de confusión (MFCC y F_0), y el entrenamiento de modelos de Machine Learning consolidan el marco científico para el desarrollo de herramientas objetivas e imparciales en el diagnóstico temprano de la depresión **[16]**.

Marco Teórico

Históricamente, se han tratado de cuantificar estos fenómenos centrándose en las variaciones de la frecuencia fundamental (F_0), la cual se origina en la vibración de las cuerdas vocales **[6]**. Autores como Darby (1977) y Hollien (1980) reportaron que los pacientes deprimidos mostraban consistentemente una reducción en el tono, un menor rango de variación de la F_0 y disminución en la velocidad del habla **[5]**. En esta línea, Nilsonne y Sundberg (1985) demostraron mediante pruebas auditivas que los contornos de F_0 contenían información paralingüística suficiente para identificar si un hablante pasaba por depresión **[5]**.

El uso de estos coeficientes ha demostrado ser superior en el modelado computacional. Estudios de Cummins (2011) confirmaron que los MFCC exhiben las características discriminatorias más fuertes para clasificar con exactitud la presencia de depresión en señales de voz **[5]**. Estas evidencias sustentan la premisa de que los MFCC capturan de forma objetiva la falta de coordinación articulatoria fina provocada por la enfermedad, aislando efectivamente el sesgo de género y proveyendo un vector de características óptimo para entrenar clasificadores de Inteligencia Artificial.

Objetivos

Objetivo general: Demostrar que los Coeficientes Cepstrales en las Frecuencias de Mel (MFCC) funcionana como un biomarcador objetivo frente a la Frecuencia fundamental (F_0), para la detección de depresión clínica, sentando las bases teóricas para el futuro desarrollo de modelos de Inteligencia Artificial.

Objetivos específicos:

- Captura de registros de voz de pacientes con diagnóstico clínico de depresión mediante lectura condicional.

- Realizar un análisis estadístico comparativo para identificar un biomarcador acústico objetivo entre los Coeficientes Cepstrales en las Frecuencias de Mel (MFCC) y la frecuencia fundamental (F_0), para mitigar sesgos de género.
- Realizar un análisis estadístico comparativo de las características extraídas por el biomarcador objetivo de sujetos sanos y sujetos con depresión.
- Proponer el uso del biomarcador valido como el estándar para entrenar modelos de IA, justificándolo con evidencia científica reciente.

Materiales y Métodos

Toma de datos (Fase clínica)

La toma de muestras acústicas se realizó a pacientes con diagnóstico clínico de depresión que acuden a las instalaciones de la Unidad Médica de Alta Especialidad (UMAE) No. 1 Bajío del IMSS. Se consideraron los registros de pacientes adultos mayores de 18 años que otorgaron su consentimiento informado por escrito.

La grabación de la voz de los pacientes se llevó a cabo en un consultorio al interior del IMSS, esto garantizó condiciones de bajo ruido de fondo. Posteriormente, fue explicado el procedimiento del registro, sin mencionar explícitamente el objetivo de la investigación para evitar que el paciente condicione o altere conscientemente la entonación de su lectura. Finalmente, se grabó utilizando un teléfono celular que se colocó a una distancia fija entre 15 y 20 cm de la boca del paciente, mientras leía un texto condicional proporcionado.

Procesamiento de señales (Fase computacional)

Los registros grabados fueron codificados y anonimizados para su procesamiento y cálculo de la frecuencia fundamental (F_0) y Coeficientes Cepstrales en las Frecuencias de Mel (MFCC).

Cálculo de la frecuencia fundamental (F_0) promedio:

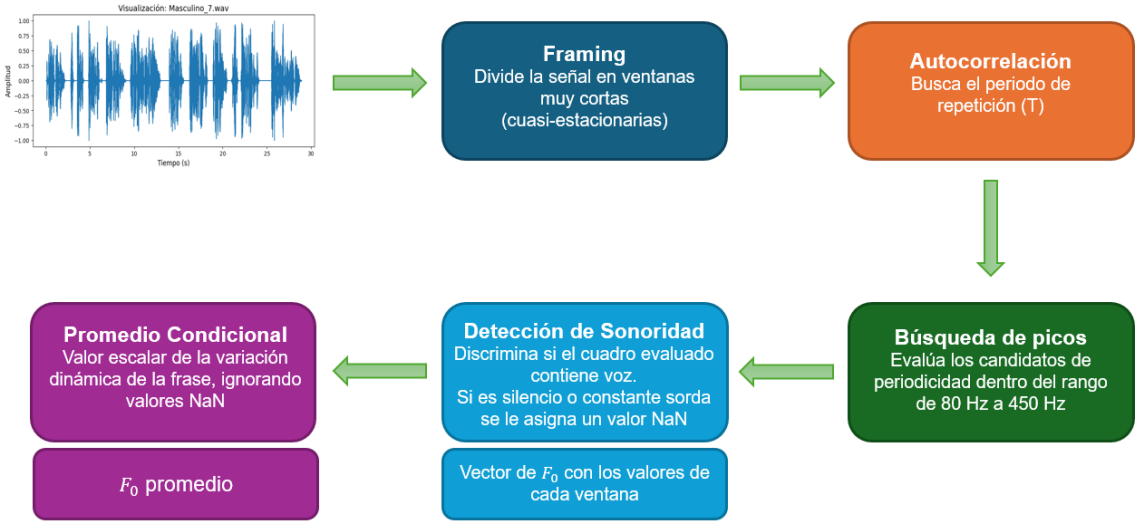


Figura 2. *Cálculo de la frecuencia fundamental promedio: Digitalización de la señal, enventanado en segmentos cuasi-estacionarios* (20 ms *a* 40 ms)*, se encuentran coincidencias máximas que indican qué tan rápido vibran las cuerdas vocales, restringiendo la búsqueda entre 80* Hz *y 450* Hz (*rango fisiológico que abarca la frecuencia de hombres y mujeres*) *para eliminar ruido, discriminación de ventanas donde no hay presencia de la vibración de las cuerdas vocales (silencio o constantes sordas*) *y cálculo del promedio* **[5,6]**.

Cálculo de los Coeficientes Cepstrales en las Frecuencias de Mel (MFCC):

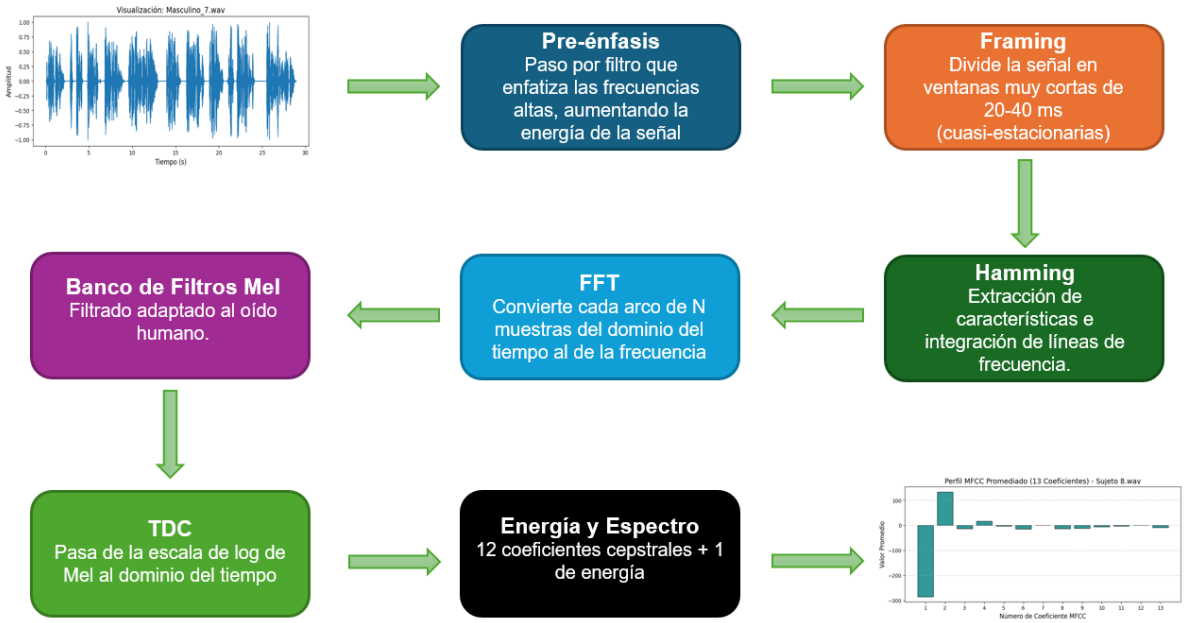


Figura 3. *Cálculo de los MFCC: Digitalización de la señal, filtrado pasa altas, enventanado en segmentos cuasi-estacionarios* (20 ms *a* 40 ms), *suavizado Hamming en cada ventana para evitar errores en los bordes al pasar a frecuencia, aplicación de la Transformada Rápida de Fourier* (FTT), *mapeo de la frecuencia lineal a la escala de Mel (logarítmica) que modela la audición humana, aplicación de la Transformada Discreta del Coseno* (TDC) *que separa la envoltura espectral (la forma de la boca y garganta) del tono de las cuerdas vocales, se obtiene un vector acústico de* 13 *valores que describen la geometría del tracto vocal al hablar, gráficamente se visualizan como un diagrama de barras* [18,19].

Análisis estadístico (Fase estadística comparativa)

Para evaluar las variables cuantitativas, su relación con el género y estado clínico, se utilizó estadística inferencial. Las pruebas estadísticas permitieron comparar las medias y medianas para identificar el grado de distinción entre sujetos sanos y sujetos con depresión, así como para validar la invariancia de género de los MFCC, consolidando estadísticamente el biomarcador objetivo antes de entrenar el algoritmo.

Análisis y Modelado (Fase de Inteligencia Artificial)

Con el biomarcador acústico validado (MFCC), se procedió a desarrollar el modelo de aprendizaje automático (Machine Learning) con estructura de clasificación y distinción de patrones acústicos entre los registros correspondientes al estado de depresión.

Resultados

Respuesta de los teléfonos celulares a señales de audio

Los registros de la voz fueron con un smartphone, para lo cual se evaluó la capacidad del dispositivo, estos teléfonos celulares (smartphones) para registrar señales acústicas capturadas de las grabaciones de voz de lectura condicional realizadas a pacientes con depresión.

Las señales obtenidas de los pacientes con diagnóstico de depresión y las obtenidas de un grupo de control (sanos) fueron representadas gráficamente y seccionadas por estado de salud (sano o con depresión) y por género de los sujetos para la preparación estadística.

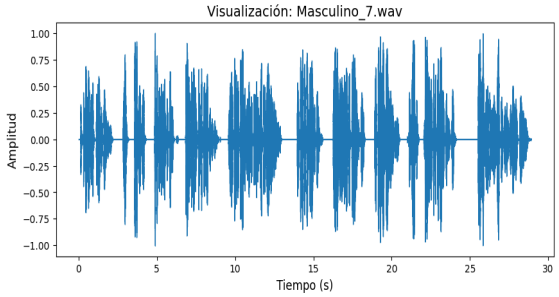


Figura 4. *Digitalización de la señal voz de pacientes con depresión* **[16]**.

Mitigación de sesgos de género

Se comparó la F_0 entre sujetos masculinos y sujetos femeninos, sin diferenciar si presentan depresión o no. Se obtuvo un $p - value = 1.45 \times 10^{-14}$, demostrando una diferencia significativa en la frecuencia fundamental entre géneros. Por lo tanto, si se utiliza como biomarcador para detectar depresión, es estadísticamente incorrecto evaluar a hombres y mujeres en el mismo grupo.

Posteriormente, se compararon los MFCC de los grupos. Se obtuvieron los siguientes $p - valores$ de los coeficientes:

Tabla 1. *Estadística comparativa de los MFCC entre géneros.*

MFCC	1	2	3	4	5	6	7	8	9	10	11	12	13
P-value	0.96	0.56	0.97	0.45	0.53	0.62	0.47	0.61	0.09	0.41	0.75	0.86	0.30

Para los coeficientes del MFCC 1 al MFCC 13 presentan una invariancia respecto al género. Por lo tanto, a diferencia de la fuente, los MFCC no presentan diferencia significativa entre géneros, abriendo la posibilidad de entrenar un modelo de IA agnóstico al sexo, enfocándose en la dinámica de la articulación del habla.

Comparativa entre sujetos sanos y con depresión.

Se comparó la F_0 entre sujetos sanos y con depresión. Se obtuvo un $p - valor = 0.819$, demostrando que no hay un diferencia significativa en la frecuencia a la que vibran las cuerdas vocales entre ambos grupos.

Posteriormente, se compararon los MFCC de los grupos. Se obtuvieron los siguientes $p - valores$ de los coeficientes:

Tabla 2. *Estadística comparativa de los MFCC entre sujetos sanos y con depresión.*

MFCC	1	2	3	4	5	6	7	8	9	10	11	12	13
P-value	<0.05	<0.05	**0.07**	<0.05	<0.05	**0.07**	<0.05	**0.18**	<0.05	<0.05	**0.11**	**0.99**	<0.05

8 de los 13 coeficientes mostraron diferencias estadísticamente significativas entre sanos y enfermos. Por los tanto, matemáticamente, capturaron una deformación consistente en el tracto vocal que ocurre de manera exclusiva en los pacientes con depresión (retardo psicomotor y tensión articulatoria).

Planteamiento del modelo de Inteligencia Artificial

La estadística reveló que 8 de los 13 MFCC lograron captar las deformaciones consistentes en el tracto vocal de los pacientes con depresión, ignorando por completo si la voz era de un hombre o de una mujer. Los MFCC son un biomarcador objetivo libre de sesgos de género y altamente sensible a los síntomas físicos de la enfermedad, teniendo el estándar ideal para plantear y entrenar un modelo predictivo de IA.

Revisiones sistemáticas recientes han confirmado que la IA posee una precisión diagnóstica cuando se entrena con muestras de voz en entornos clínicos **[1]**. Lo más revolucionario es el hecho de que estos algoritmos ya no necesitan equipos de estudio de grabación costosos. Investigaciones recientes han demostrado que es posible detectar perfiles depresivos analizando algo tan cotidiano como las notas de voz enviadas por WhatsApp **[12]**, o capturando registros directamente desde los micrófonos de los teléfonos celulares **[13]**.

Por lo tanto, el uso de los MFCC como biomarcador de este modelo asegura que la IA no solo sea accesible, rápida y económica, sino fundamental como modelo de soporte al diagnóstico de enfermedades como la depresión. Esto convierte al análisis computacional de la voz a través de dispositivos celulares en una herramienta de prevención en salud mental.

Conclusión

La investigación confirma que es posible encontrara marcadores físicos y medibles de la depresión clínica a través del análisis del habla. El estudio revela que las métricas, como el análisis de la fuente, fallan al intentar crear un modelo de diagnóstico debido al sesgo de género del paciente.

No obstante, el hallazgo central demuestra que los Coeficientes Cepstrales en las Frecuencias de Mel (MFCC) actúan como el biomarcador acústico ideal. Las pruebas estadísticas probaron que estos coeficientes son invariables antes el sexo biológico, garantizando un análisis equitativo. Además, se logró identificar la deformación presente en el tracto vocal debido al retardo psicomotor y la tensión articulatoria que ocurren exclusivamente en los pacientes que padecen de depresión.

Al demostrar que estos patrones acústicos pueden ser capturados con la misma eficacia utilizando teléfonos celulares, se establecen las bases teóricas y estadísticas para el desarrollo de sistemas de IA. La integración de este biomarcador con el aprendizaje automático promete transformar el campo de la salud mental, convirtiendo los teléfonos celulares en una herramienta de soporte diagnóstico rápido, económico y de alto alcance.

Agradecimientos

Se extiende un sincero agradecimiento a la Unidad Médica de Alta Especialidad (UMAE) No. 1 Hospital de Especialidades Centro Médico Nacional del Bajío, pertenecientes al Instituto Mexicano del Seguro Social (IMSS) en León, Guanajuato.

Los autores agradecemos también al comité organizacional del XXXI Verano de la Ciencia UG por el apoyo y sustento para realizar la investigación.

Referencias

[1] Bauer, J. F., *et al*., (2024). Validation of machine learning-based assessment of major depressive disorder from paralinguistic speech characteristics in routine care. *Depression and Anxiety*, **2024**(**1**): 9667377.

[2] Lombardo, C., *et al*., (2025). Speech analysis and speech emotion recognition in mental disease: A scoping review. *Frontiers in Psychology*, **16**: 1645860.

[3] Rodrigo, I., & Duñabeitia, J. A. (2025). Listening to the mind: Integrating vocal biomarkers into digital health. *Brain Sciences*, **15**(**7**): 762.

[4] Coliñir Olea, N., Figueroa Saavedra, C., & Jara Cabrera, G. (2023). Conducta suicida, riesgo suicida y los parámetros acústicos de la voz y el habla: Revisión sistemática. *Revista Argentina de Ciencias del Comportamiento*, *15*(2), 1-14.

[5] Cummins, N., Scherer, S., Krajewski, J., Schnieder, S., Epps, J., & Quatieri, T. F. (2015). A review of depression and suicide risk assessment using speech analysis. *Speech Communication*, **71**: 10-49.

[6] Dumpala, S. H., *et al*., (2026). Acoustics of depression. *Journal of Affective Disorders*, 121958.

[7] Zhao, M., **et al.**, (2025). Decoupled multi-perspective fusion for speech depression detection. *IEEE Transactions on Affective Computing*, **16**(**3**): 1772-1786.

[8] Liu, L., et al. (2024). Diagnostic accuracy of deep learning using speech samples in depression: A systematic review and meta-analysis. *Journal of the American Medical Informatics Association*, **31**(**10**): 2394-2404.

[9] Uddin, M. A., Joolee, J. B., & Sohn, K.-A. (2022). Deep multi-modal network based automated depression severity estimation. *IEEE Transactions on Affective Computing*, **14**(**3**): 2153-2167.

[10] Herrera Maldonado, L. Á. (s.f.). Detección de depresión con técnicas de inteligencia artificial.

[11] Yadav, P., Jain, P., & Jain, T. (2023). Depression detection through audio analysis using machine learning models ensuring sustainable development of mankind. *CEUR Workshop Proceedings*.

[12] Otani, V. H., *et al.*, (2026). ML-based detection of depressive profile through voice analysis in WhatsApp™ audio messages of Brazilian Portuguese speakers. *PLOS Mental Health*, **3**(**1**): e0000357.

[13] Sharma, S. K., *et al.*, (2025). Early detection of mental health disorders using machine learning models using behavioral and voice data analysis. *Scientific Reports*, **15**(**1**): 16518.

[14] Brenes, L. F., *et al.*, (2025). Data collection for automatic depression identification in Spanish speakers using deep learning algorithms: Protocol for a case-control study. *JMIR Research Protocols*, **14**(**1**): e60439.

[15] Callejas, F. E. S., *et al.*, (2025). Detección y cuantificación de emociones en población mexicana mediante el análisis de expresiones faciales y reconocimiento de voz.

[16] Wei, Y., *et al.*, (2025). Acoustic-based machine learning approaches for depression detection in Chinese university students. *Frontiers in Public Health*, **13**, 1561332.

[17] Seikel, J. A., Drumright, D. G., & Hudock, D. J. (2021). *Anatomy & physiology for speech, language, and hearing* (6ta ed.). Plural Publishing.

[18] Elamvazuthi, I. (2010). Voice recognition algorithms using mel frequency cepstral coefficient (MFCC) and dynamic time warping (DTW) techniques. *Computing Research Repository*.

[19] Saputro, W. T., Fadlil, A., & Murinto, M. (2025). Voice-based depression pattern recognition using mel-frequency cepstral coefficients feature extraction. *Jurnal PROCESSOR*, *20*(2).