Facultad de Ingeniería

Carrera de Ingeniería de Software

TRABAJO DE INVESTIGACIÓN

Análisis acústico del estado y bienestar emocional, mediante espectrogramas y redes neuronales: Revisión sistemática

Autores

Ortiz Orellana, Fabrizio Heiner (U21313581)

Rivera Calderón, Elkin Jenner (U19311223)

Docentes

Sanchez Atuncar, Giancarlo

Lima, Perú

2025

INTRODUCCIÓN

La salud mental es un área crítica de investigación debido al creciente impacto de los trastornos emocionales en la calidad de vida (World Health Organization, 2023). La depresión, en particular, presenta un desafío para los sistemas de salud, dado que su diagnóstico temprano es a menudo subjetivo (Nfissi et al., 2024). Ante esto, técnicas como los espectrogramas y las redes neuronales han surgido como herramientas prometedoras. Los espectrogramas visualizan las características del habla, lo que permite a las redes neuronales identificar patrones complejos asociados a los estados de ánimo (Issa et al., 2020).

Sin embargo, a pesar de los avances en el análisis de voz con inteligencia artificial, persisten importantes vacíos de conocimiento. Las variaciones en el idioma, entonación y entorno acústico influyen en el rendimiento de los modelos, dificultando el establecimiento de estándares (Ooba et al., 2025). Además, muchos estudios se limitan a poblaciones específicas, impidiendo una visión panorámica y la comparación de arquitecturas de redes neuronales y sus métricas de desempeño (Wang et al., 2025). Esta falta de sistematización en la literatura limita el aprovechamiento de estas tecnologías en aplicaciones reales para la detección temprana de depresión y el riesgo suicida.

En este contexto, es necesaria una Revisión Sistemática de Literatura (RSL) que consolide los hallazgos sobre el uso de espectrogramas y redes neuronales en la detección de trastornos emocionales. Esta RSL permitirá identificar patrones metodológicos, resaltar vacíos de investigación y sugerir mejores prácticas. La importancia de esta revisión radica en su impacto científico, al organizar un campo de estudio disperso; en su relevancia tecnológica, al guiar el diseño de soluciones de IA; y en su valor social, al contribuir a mejorar la detección temprana del riesgo suicida, un problema de urgencia global (Manikandan & Neethirajan, 2025). No se han encontrado revisiones recientes que sinteticen los avances en esta área, lo que justifica este estudio.

El objetivo principal de esta revisión sistemática de la literatura es analizar y clasificar los principales estudios que utilizan espectrogramas y redes neuronales para el reconocimiento de emociones, con el fin de identificar las arquitecturas más prometedoras y las mejores prácticas en el preprocesamiento de datos.

En tal sentido, este documento se organiza de la siguiente manera: la Sección 2, "Metodología", detalla el proceso de la RSL, incluyendo la estrategia de búsqueda. La Sección 3, "Resultados", presenta una síntesis de los hallazgos clave de los estudios seleccionados, enfocándose en las arquitecturas y métricas de desempeño. La Sección 4, "Discusión", analiza y compara los resultados, identificando desafíos y tendencias.

Finalmente, la Sección 5, "Conclusiones", resume los hallazgos principales de la revisión y sugiere futuras direcciones de investigación.

PROBLEMÁTICA:

A pesar del potencial del análisis acústico basado en espectrogramas y redes neuronales para ofrecer indicadores objetivos del estado emocional, la implementación efectiva de estas técnicas en contextos clínicos y de atención enfrenta desafíos concretos. En entornos de atención primaria, servicios de salud mental y líneas de ayuda, la dependencia de evaluaciones clínicas tradicionales y escalas psicométricas provoca demoras en la identificación temprana de la depresión y del riesgo suicida, lo que incrementa la probabilidad de intervenciones tardías y resultados adversos [1]. Estudios recientes que analizan llamadas de alto riesgo y grabaciones clínicas muestran resultados prometedores pero también limitaciones metodológicas que impiden la generalización inmediata de los modelos propuestos [2].

Además, la literatura evidencia una gran heterogeneidad metodológica: variaciones en el tipo de espectrograma utilizado (STFT, mel, log-mel, MFCC), diferentes estrategias de preprocesamiento y augmentación, diversidad en las arquitecturas de redes neuronales (CNN, RNN, Transformer, híbridos) y protocolos de validación inconsistentes entre estudios, lo cual dificulta la comparación y la replicación de resultados [3][4][5]. Esta falta de estandarización se agrava cuando los modelos se entrenan y evaluan en poblaciones y condiciones acústicas reducidas (idioma, calidad de grabación, ambiente controlado), lo que reduce la robustez y la capacidad de transferencia a escenarios del mundo real (por ejemplo, llamadas telefónicas o grabaciones en entornos ruidosos) [4][5].

Por otro lado, existen preocupaciones sobre la explicabilidad y la seguridad de los modelos de aprendizaje profundo en aplicaciones sensibles como la detección del riesgo suicida; la ausencia de métricas y protocolos comunes para reportar la validación externa y el manejo de falsos negativos limita la confianza clínica en estas herramientas [6]. En consecuencia, se identifica una necesidad urgente de sintetizar sistemáticamente la evidencia para: (a) mapear qué prácticas metodológicas producen resultados más reproducibles, (b) evaluar la robustez de los enfoques frente a variaciones de idioma y condiciones acústicas, y (c) establecer recomendaciones para la validación externa y la interpretación clínica.

En síntesis, la problemática central es la falta de una visión integrada y estandarizada que permita evaluar comparativamente el rendimiento de métodos basados en espectrogramas y redes neuronales para la detección de depresión y riesgo

suicida, y determinar su aplicabilidad práctica en distintos contextos sanitarios y sociales.

JUSTIFICACIÓN:

Ante la problemática descrita, se propone realizar una Revisión Sistemática de Literatura (RSL) que consolide y organice los hallazgos sobre el uso de espectrogramas y redes neuronales profundas en la detección de trastornos emocionales. Desde una perspectiva científica, esta RSL permitirá identificar patrones metodológicos recurrentes, limitar sesgos metodológicos documentados y determinar lagunas de investigación que requieran estudios primarios adicionales [7][8].
Desde el punto de vista tecnológico, la revisión orientará a ingenieros e investigadores sobre qué tipos de representaciones espectrográficas, arquitecturas de redes y técnicas de augmentación han mostrado mayor capacidad predictiva y transferencia entre datasets, facilitando la implementación de soluciones de software robustas y reproducibles en salud digital [3][5]. Además, la sistematización de prácticas de validación y métricas favorecerá la comparabilidad entre estudios y la generación de benchmarks que aceleren la madurez del campo.
En términos sociales y clínicos, mejorar la detección temprana de la depresión y el riesgo suicida mediante herramientas no invasivas basadas en voz puede traducirse en intervenciones más oportunas, apoyo psicosocial tempran o derivaciones a servicios especializados, lo que potencialmente reduce morbilidad y mortalidad asociada a estos trastornos [1][2]. Finalmente, dado que no se ha identificado una revisión reciente que integre específicamente las evidencias centradas en espectrogramas + redes neuronales para depresión y riesgo suicida, la RSL propuesta llenará un vacío relevante y generará recomendaciones prácticas para futuros estudios y aplicaciones clínicas [6].

Metodología

Pregunta PICO y sus componentes

RQ1: ¿En personas con depresión y/o riesgo suicida (2015–2025), cómo afecta el uso de espectrogramas analizados mediante redes neuronales profundas en comparación con métodos sin aprendizaje profundo o sin representaciones espectrográficas a la exactitud diagnóstica (AUC, F1, sensibilidad, especificidad) y la detección temprana?

- **P:** ¿Qué poblaciones (edad, idioma, contexto clínico vs. no clínico) han sido estudiadas y qué criterios se usan para etiquetar depresión/riesgo?
- **I:** ¿Qué tipos de espectrogramas y técnicas de preprocesamiento son más usados?
- **C:** ¿Qué diferencia de desempeño existe entre enfoques DL+espectrograma y ML clásico con features acústicos?
- **O:** ¿Qué métricas y protocolos de validación se reportan con más frecuencia?
- **T:** ¿Qué tendencias temporales (2015–2025) se observan en técnicas y métricas?

Tabla 1

Elementos PICO

RQ	¿En personas con depresión y/o riesgo suicida (2015–2025), cómo afecta el uso de espectrogramas analizados mediante redes neuronales profundas en comparación con métodos sin aprendizaje profundo o sin representaciones espectrográficas a la exactitud diagnóstica (AUC, F1, sensibilidad, especificidad) y la detección temprana?
RQ1	¿Qué poblaciones (edad, idioma, contexto clínico vs. no clínico) han sido estudiadas y qué criterios se usan para etiquetar depresión/riesgo?
RQ2	¿Qué tipos de espectrogramas y técnicas de preprocesamiento son más usados?
RQ3	¿Qué diferencia de desempeño existe entre enfoques DL+espectrograma y ML clásico con features acústicos?
RQ4	¿Qué métricas y protocolos de validación se reportan con más frecuencia?
RQ5	¿Qué tendencias temporales (2015–2025) se observan en técnicas y métricas?

Ilustracion 1

Palabras clave

P	I	C	O
	spectrogram, mel	machine learning,	AUC, F1,

depression, major depressive disorder, suicidal, suicide risk, mental health, PHQ-9, BDI, C-SSRS, adults, adolescents, patient	spectrogram, log-mel, MFCC, STFT, deep learning, neural network, CNN, LSTM, GRU, Transformer, GNN	SVM, random forest, logistic regression, prosodic features, clinical assessment	accuracy, precision, recall, sensitivity, specificity, external validation, robustness, explainability

Tabla 2

Ecuación

(TITLE-ABS-KEY(speech OR voice OR audio OR vocal OR "phone call" OR call*) AND TITLE-ABS-KEY(spectrogram* OR "mel spectrogram" OR "log-mel" OR STFT OR MFCC OR wavelet) AND TITLE-ABS-KEY("neural network*" OR "deep learn*" OR CNN OR "convolutional neural network" OR LSTM OR GRU OR transformer* OR "graph neural network" OR GNN) AND TITLE-ABS-KEY(depress* OR "mood disorder" OR "mental health" OR "suicide risk" OR suicid*))

Ilustración 2

Búsqueda mediante palabras clave en Scopus

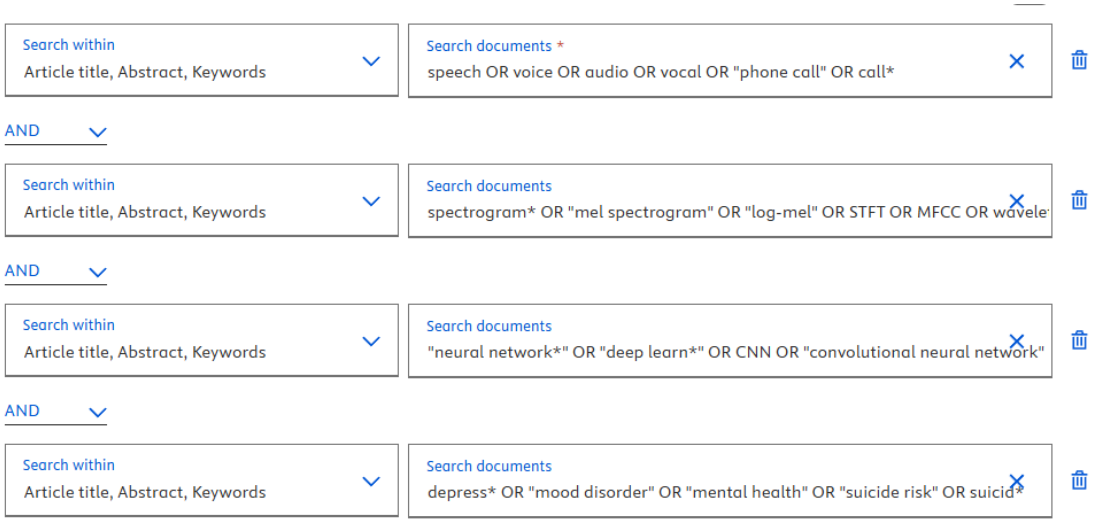


Ilustración 3

Cantidad de investigaciones encontradas en Scopus

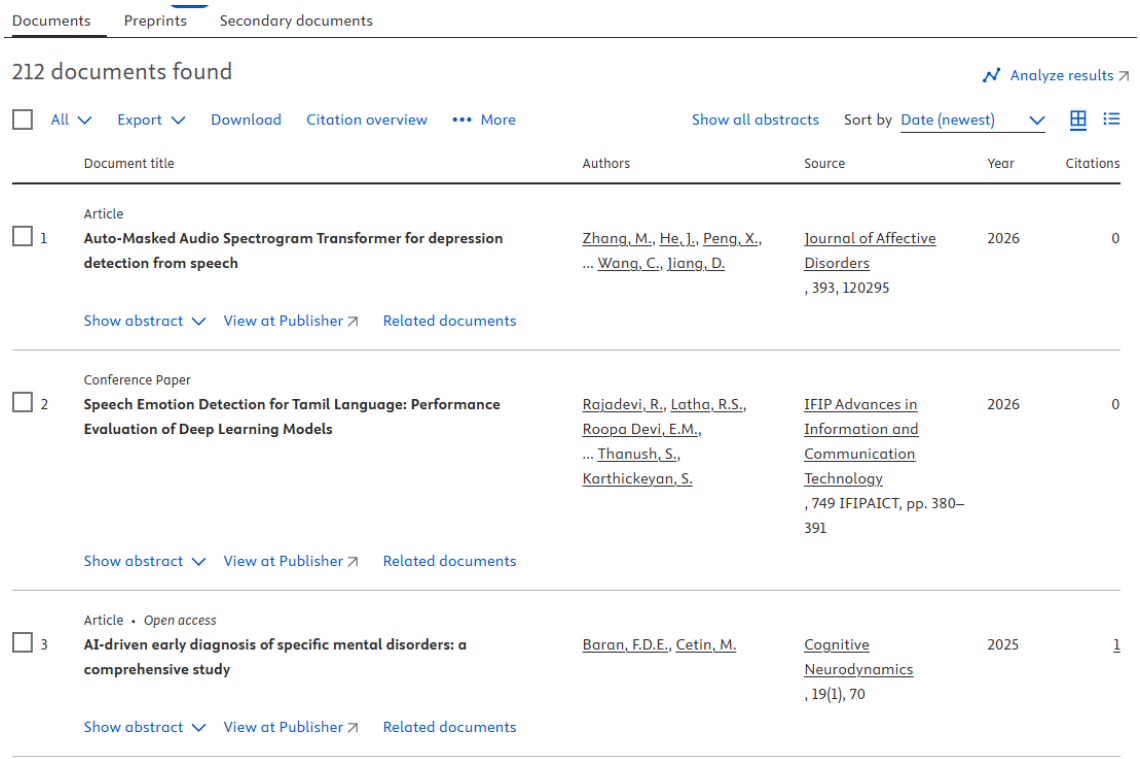


Tabla 3

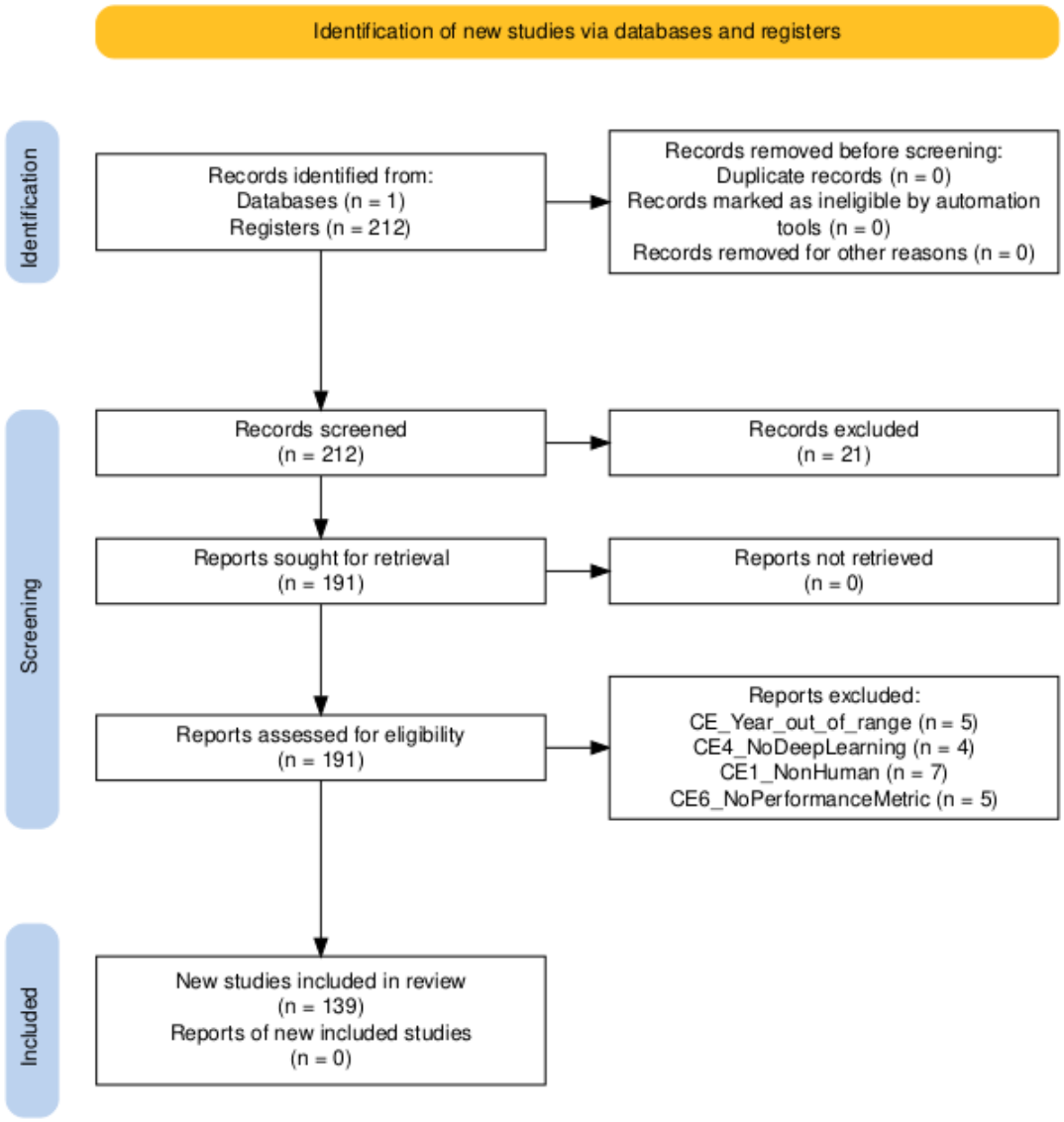
Criterios de Inclusión y Exclusión

Criterios de Inclusión	Criteriode Exclusión
CI1. Estudios publicados entre 2015 y 2025.	CE1. Estudios centrados en animales o en dominios no relacionados con salud mental (p. ej., acústica animal sin relación clínica).
CI2. Artículos en inglés (o en español, si se decide ampliar).	CE2. Trabajos sin datos empíricos (editoriales, opinión, cartas, posters sin texto completo).
CI3. Estudios empíricos (artículos originales, conferencias, revisiones sistemáticas) que utilicen señales de voz humana y representaciones espectrográficas (STFT, mel/log-mel, MFCC como imagen) y que apliquen redes neuronales profundas para clasificación o detección de depresión, trastornos del ánimo o riesgo suicida.	CE3. Estudios que no usan espectrogramas ni MFCC-imagen (por ejemplo solo análisis textual/NLP o solo imágenes médicas no relacionadas con audio).
CI4. Estudios que reporten al menos una métrica de desempeño (ej.: AUC, F1, accuracy, precision, recall, sensibilidad, especificidad).	CE4. Estudios que no utilizan redes neuronales profundas (siempre que no sean comparadores dentro del mismo estudio, en ese caso pueden ser incluidos como comparación).
CI5. Texto completo disponible (PDF	CE5. Documentos duplicados (se

accesible al equipo).	conserva la versión más completa / con mayor información).
CI6. Estudios que comparen la intervención (DL + espectrograma) con métodos alternativos, o que realmente evalúen la aplicabilidad/robustez del modelo.	CE6. Estudios sin métrica de desempeño reportada o sin información suficiente para evaluar resultados (salvo que sean SLRs relevantes que sí aporten síntesis útil).

Figura 1

Diagrama PRISMA



BIBLIOGRAFIA

- **World Health Organization. (2023).** *World mental health report: Transforming mental health for all.* Geneva: World Health Organization. World Health Organization
- **Nfissi, A., Bouachir, W., Bouguila, N., & Mishara, B. L. (2024).** Unlocking the emotional states of high-risk suicide callers through speech analysis. En *2024 IEEE 18th International Conference on Semantic Computing (ICSC)* (pp. 33–40). IEEE. https://doi.org/10.1109/ICSC59802.2024.00012. ResearchGate+1
- **Issa, D., Demirci, M. F., & Yazici, A. (2020).** Speech emotion recognition with deep convolutional neural networks. *Biomedical Signal Processing and Control*, *59*, 101894. https://doi.org/10.1016/j.bspc.2020.101894. ScienceDirect+1
- **Ooba, H., Maki, J., & Masuyama, H. (2025).** Voice analysis and deep learning for detecting mental disorders in pregnant women: a cross-sectional study. *Discover Mental Health Research.* (Artículo disponible en PubMed Central). PMC+1
- **Wang, Y., Qu, T., Zhu, W., Wang, Q., Cao, Y., & Gui, R. (2025).** A hybrid model using multimodal feature perception and multiple cross-attention fusion for depressive episodes detection. *Information Fusion*, *124*, 103354. https://doi.org/10.1016/j.inffus.2025.103354. ScienceDirect+1
- **Manikandan, V., & Neethirajan, S. (2025).** AI-Powered vocalization analysis in poultry: systematic review of health, behavior, and welfare monitoring. *Sensors*, *25*(13), 4058. https://doi.org/10.3390/s25134058. *(Esta revisión sobre vocalizaciones animales la citamos como ejemplo de revisión sistemática en acústica; su metodología y discusión sobre espectrogramas/DL es útil para técnicas aplicadas).*
- **Page, M. J., McKenzie, J. E., Bossuyt, P. M., et al. (2021).** The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ, 372:n71.* https://doi.org/10.1136/bmj.n71. BMJ+1
- **Kitchenham, B., & Charters, S. (2007).** Guidelines for performing Systematic Literature Reviews in Software Engineering (EBSE Technical Report). (Guía clásica adaptada en SE; útil para aspectos metodológicos de la RSL). ACM Digital Library+1
- **Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001).** The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine*, *16*(9), 606–613. (Instrumento de cribado clínico citado como criterio de población). PMC+1
- **Beck, A. T., Steer, R. A., & Brown, G. K. (1996).** *Manual for the Beck Depression Inventory-II (BDI-II).* The Psychological Corporation. (Referencia del BDI-II). SpringerLink

- **Posner, K., Brown, G. K., Stanley, B., et al. (2011).** The Columbia-Suicide Severity Rating Scale (C-SSRS): initial validity and internal structure. *American Journal of Psychiatry*, *168*(12), 1266-1277. (Escala citada en la población P).