PROYECTO DE INVESTIGACIÓN

Proyecto de Investigación

SentirIA Research Papers · 2026

Infraestructura de detección temprana en salud mental

Este documento es parte de la base científica de SentirIA,
plataforma de detección temprana y monitoreo continuo de deterioro en salud mental.
No constituye diagnóstico clínico. La evaluación es responsabilidad del profesional.

Proyecto de Investigación

Biomarcadores vocales para la detección de depresión en español argentino: desarrollo y validación del primer corpus de habla clínicamente etiquetado en América Latina

---

1. Datos del proyecto

Título: Biomarcadores vocales para la detección de depresión en español argentino: desarrollo y validación del primer corpus de habla clínicamente etiquetado en América Latina (AR-DEP-Corpus)

Área disciplinar: Informática Médica / Salud Mental Digital / Procesamiento del Habla

Duración estimada: 24 meses

Investigador responsable: [A completar]

Equipo de investigación:

- Investigadores en informática médica / procesamiento de señales
- Psiquiatras o psicólogos clínicos con experiencia en evaluación estandarizada
- Ingeniero/desarrollador de sistemas con experiencia en procesamiento de audio
- Estadístico o data scientist
- Coordinador de campo para reclutamiento y seguimiento

Instituciones participantes:

- Universidad de Buenos Aires, Facultad de Medicina
- Centro asistencial con servicio de salud mental (reclutamiento clínico)
- Laboratorio de procesamiento de señales o inteligencia artificial (procesamiento acústico)

Financiamiento potencial:

- CONICET (Beca doctoral / PIP — Proyecto de Investigación Plurianual)
- ANPCyT (PICT — Proyecto de Investigación Científica y Tecnológica)
- FONTAR (para transferencia tecnológica)
- Subsidios internacionales: NIH Fogarty, Wellcome Trust, Grand Challenges Canada

2. Resumen del proyecto

El presente proyecto propone desarrollar y validar el primer corpus de habla para detección de depresión en español argentino (AR-DEP-Corpus), abordando la brecha más significativa en el campo de biomarcadores vocales de salud mental: la ausencia total de datasets validados en cualquier variedad del español. La base de evidencia global se sustenta en aproximadamente 800 participantes clínicamente etiquetados, con inglés y chino mandarín representando más del 90% del total. No existe un solo dataset publicado en español.

El estudio reclutará 120 adultos (60 con depresión mayor, 60 controles) del área metropolitana de Buenos Aires, recopilará grabaciones de voz en condiciones clínicas estandarizadas y mediante mensajes de audio de WhatsApp (validez ecológica), extraerá el set estandarizado de 88 features acústicas eGeMAPS mediante openSMILE, y evaluará la capacidad de detección mediante modelos de machine learning con validación cruzada leave-one-subject-out. Como resultado secundario, se evaluará la transferibilidad de modelos entrenados en inglés (DAIC-WOZ) al español argentino, cuantificando la degradación cross-lingüística y la superioridad de features prosódicas sobre espectrales en transferencia entre idiomas.

El dataset resultante será el primero en su tipo para América Latina y se publicará bajo licencia de acceso controlado, constituyendo un recurso de referencia para la comunidad científica internacional.

3. Planteamiento del problema

3.1 La depresión como problema de salud pública en Argentina y América Latina

La depresión afecta a 280 millones de personas en el mundo (OMS, 2023) y constituye la principal causa de discapacidad global. En América Latina, 50 millones de personas padecen esta condición. Argentina presenta una prevalencia estimada de trastornos depresivos del 5,7% de la población (GBD 2019), con una brecha de tratamiento que supera el 70% — es decir, más de 7 de cada 10 personas con depresión nunca acceden a atención profesional.

El screening actual depende del PHQ-9, un cuestionario auto-administrado que requiere que el paciente reconozca su condición y busque evaluación. Solo el 4% de los pacientes en atención primaria son evaluados para depresión (Mazur et al., 2025). Este modelo de detección por demanda falla especialmente en contextos donde el estigma hacia la salud mental es alto — situación particularmente prevalente en América Latina, donde aproximadamente el 50% de la población cita estigma personal como barrera para buscar atención (Mascayano et al., 2016).

3.2 Los biomarcadores vocales como alternativa no invasiva

La depresión produce alteraciones medibles en la producción vocal a través de tres mecanismos fisiopatológicos: retardación psicomotora (que reduce la velocidad del habla y la articulación), disminución del soporte respiratorio (que reduce la intensidad y la energía vocal), y deterioro del control neuromuscular de los pliegues vocales (que aumenta el jitter, shimmer y reduce la relación armónicos-ruido).

La meta-análisis más extensa hasta la fecha (Maran et al., 2025; 105 estudios, JMIR Mental Health) reportó una precisión agrupada de 81% (IC 95%: 79-83%), sensibilidad de 84% (IC: 81-86%) y especificidad de 83% (IC: 79-86%). Features individuales alcanzan correlaciones significativas con severidad depresiva: variabilidad de pitch ($r = -0.54$ con BDI-II), jitter ($r > 0.51$), shimmer ($r > 0.40$), y velocidad del habla (reducción del 10-30%).

El estudio de Kintsugi (Mazur et al., 2025, Annals of Family Medicine) evaluó su modelo en 14.898 adultos y logró sensibilidad de 71.3% y especificidad de 73.5% con solo 25 segundos de habla libre. Notablemente, la sensibilidad fue más alta en población hispana/latina (80.3%), sugiriendo que los biomarcadores vocales podrían ser particularmente efectivos en esta población.

3.3 La brecha del español: el problema central

A pesar de esta evidencia robusta, no existe un solo dataset validado de habla para detección de depresión en español — ni argentino, ni mexicano, ni peninsular, ni de ninguna variedad. Los datasets disponibles son:

Dataset	Idioma	Participantes	Etiquetas	Disponibilidad
DAIC-WOZ	Inglés (EE.UU.)	189	PHQ-8	EULA (gratuito)
E-DAIC	Inglés (EE.UU.)	275	PHQ-8	EULA
EATD	Chino mandarín	162	SDS	Público
MODMA	Chino mandarín	52	PHQ-9	Solicitud
CMDC	Chino mandarín	78	HAMD-17	EULA
BlackDog	Inglés (Australia)	60	QIDS-SR	Solicitud
D3TEC	Español (México)	~60 (en curso)	PHQ-9	No disponible aún

El único proyecto en español (D3TEC) se encuentra en fase de recolección en México y aún no ha publicado datos. No existe nada en español rioplatense.

Esta ausencia tiene consecuencias directas: los modelos entrenados en inglés pierden precisión al aplicarse a otros idiomas. Wang et al. (2025) reportaron una degradación de ~8% en precisión al transferir un modelo inglés→chino. Para español argentino (rioplatense), se esperan desafíos adicionales: patrones de entonación influenciados por el italiano, la realización /ʃ/ de ll/y, velocidad del habla típicamente más rápida que otras variedades del español, y el voseo en conjugaciones verbales.

3.4 Pregunta de investigación

¿Es posible detectar depresión clínica a partir de biomarcadores vocales extraídos de grabaciones de voz en español argentino con precisión comparable a la reportada en la literatura para inglés? ¿Qué features acústicas transfieren entre idiomas y cuáles requieren calibración local?

---

4. Objetivos

4.1 Objetivo general

Desarrollar y validar el primer corpus de habla clínicamente etiquetado para detección de depresión en español argentino (AR-DEP-Corpus), y evaluar la precisión de modelos de machine learning entrenados con dicho corpus para la clasificación binaria de depresión y la predicción de severidad según PHQ-9.

4.2 Objetivos específicos

1. Construir un corpus de al menos 120 participantes (60 con TDM, 60 controles sanos) con grabaciones estandarizadas en condiciones clínicas y grabaciones ecológicas vía WhatsApp, etiquetados con PHQ-9 (primario) y BDI-II (secundario), con diagnóstico confirmado por entrevista estructurada MINI.

2. Extraer y analizar el set estandarizado de 88 features eGeMAPS y features clínicas complementarias (F0, jitter, shimmer, HNR, velocidad del habla, ratio de pausas, MFCCs), identificando cuáles presentan diferencias estadísticamente significativas entre grupos deprimido y control en español argentino.

3. Entrenar y validar clasificadores de machine learning (SVM, Random Forest, XGBoost) y modelos basados en representaciones aprendidas (wav2vec 2.0, Whisper) para clasificación binaria (PHQ-9 ≥ 10) y regresión de severidad (PHQ-9 total), reportando precisión, sensibilidad, especificidad, AUC y MAE con validación leave-one-subject-out.

4. Evaluar la transferibilidad cross-lingüística: comparar la performance de (a) un modelo entrenado exclusivamente en DAIC-WOZ (inglés) aplicado a AR-DEP-Corpus (español), (b) un modelo entrenado exclusivamente en AR-DEP-Corpus, y (c) un modelo combinado DAIC-WOZ + AR-DEP-Corpus, cuantificando la degradación cross-lingüística y el beneficio de datos nativos.

5. Comparar la performance de features prosódicas (F0, energía, velocidad, pausas) versus features espectrales (MFCCs, formantes, ratio espectral) en la transferencia entre idiomas, para determinar cuáles son verdaderamente agnósticas al idioma.

6. Evaluar la viabilidad de detección a partir de mensajes de voz de WhatsApp (audio comprimido en formato Opus) comparando con grabaciones en condiciones clínicas controladas, para validar el uso de WhatsApp como plataforma de screening en condiciones ecológicas.

7. Desarrollar un modelo de línea base personal que compare al paciente consigo mismo a lo largo del tiempo, y evaluar su capacidad de detección de cambio clínico (mejoría o deterioro) mediante diseño longitudinal.

8. Publicar el AR-DEP-Corpus como recurso de acceso controlado para la comunidad científica internacional, junto con benchmarks reproducibles y código abierto de extracción y clasificación.

---

5. Hipótesis

H1: Los biomarcadores vocales extraídos de grabaciones en español argentino permitirán clasificar depresión (PHQ-9 ≥ 10 vs. < 10) con una precisión ≥ 75% y AUC ≥ 0.78, comparable a los rangos reportados en la literatura para inglés (precisión agrupada 81%, Maran et al., 2025).

H2: Las features prosódicas (variabilidad de F0, velocidad del habla, duración de pausas, energía vocal) mostrarán mayor transferibilidad cross-lingüística (menor degradación inglés→español) que las features espectrales (MFCCs, formantes, pendientes espectrales), consistente con la hipótesis de que las alteraciones psicomotoras son independientes del idioma.

H3: Los modelos de línea base personal (comparación intra-sujeto) superarán a los modelos poblacionales (comparación inter-sujeto) en la detección de cambio clínico, con AUC ≥ 0.80 para detección de deterioro ≥ 5 puntos en PHQ-9.

H4: Las grabaciones de WhatsApp (formato Opus comprimido) producirán una degradación ≤ 10% en AUC comparadas con grabaciones clínicas controladas (formato WAV no comprimido), replicando los hallazgos de Otani et al. (2026) para portugués brasileño.

---

6. Metodología

6.1 Diseño del estudio

Estudio observacional analítico, transversal con componente longitudinal. El diseño transversal evalúa la capacidad discriminativa de los biomarcadores vocales entre grupos (deprimido vs. control). El componente longitudinal (subset de 30 participantes seguidos durante 12 semanas

con evaluaciones mensuales) evalúa la capacidad de detección de cambio clínico y la viabilidad de modelos de línea base personal.

6.2 Participantes

Población objetivo: Adultos de 18-65 años, hablantes nativos de español argentino (variedad rioplatense), residentes en el Área Metropolitana de Buenos Aires.

Grupo depresión (n = 60):

- Criterios de inclusión: PHQ-9 ≥ 10 en screening inicial + diagnóstico de Trastorno Depresivo Mayor confirmado por entrevista estructurada MINI (Mini International Neuropsychiatric Interview) según criterios DSM-5.
- Distribución por severidad: 20 moderados (PHQ-9 10-14), 20 moderadamente severos (PHQ-9 15-19), 20 severos (PHQ-9 20-27).
- Se aceptan pacientes con o sin tratamiento farmacológico (el estado de medicación se registra como covariable).

Grupo control (n = 60):

- Criterios de inclusión: PHQ-9 < 5 + ausencia de antecedentes de TDM en los últimos 2 años + MINI negativo para episodio depresivo actual.
- Pareados por sexo, rango etario (±5 años) y nivel educativo con el grupo depresión.

Criterios de exclusión (ambos grupos):

- Trastorno psicótico activo
- Trastorno por uso de sustancias (último año)
- Discapacidad auditiva
- Trastorno del habla o de la voz diagnosticado
- Hablante no nativo de español
- Patología laríngea o respiratoria aguda al momento de la grabación
- Infección respiratoria aguda al momento de la grabación

Subset longitudinal (n = 30):

- 30 participantes del grupo depresión que acepten seguimiento de 12 semanas
- Evaluaciones mensuales (semanas 0, 4, 8, 12) con PHQ-9 presencial + grabación de voz
- Permite evaluar sensibilidad al cambio y modelos de línea base personal

Cálculo de tamaño muestral: Para detectar un AUC de 0.80 (hipótesis) contra un AUC nulo de 0.50, con $\alpha = 0.05$ y potencia $1-\beta = 0.80$, se requieren aproximadamente 26 sujetos por grupo (Hanley & McNeil, 1982). Con 60 por grupo, el estudio tiene potencia suficiente para detectar AUC ≥ 0.72 y para análisis de subgrupos por sexo (30 por subgrupo). Los 120 participantes

totales son comparables o superiores a la mayoría de datasets publicados (DAIC-WOZ: 189; EATD: 162; MODMA: 52; BlackDog: 60; CMDC: 78).

6.3 Reclutamiento

Grupo depresión: Consultorios externos de psiquiatría y psicología de centros asistenciales del AMBA. Derivación por profesionales tratantes. Se ofrecerá participación a pacientes con evaluación PHQ-9 ≥ 10 como parte de su atención habitual.

Grupo control: Personal administrativo, acompañantes de pacientes, y voluntarios de la comunidad universitaria. Reclutamiento por anuncio. Screening con PHQ-9 + MINI para confirmar ausencia de depresión actual.

6.4 Protocolo de grabación

Cada participante realizará grabaciones en dos condiciones:

Condición A — Clínica controlada:

- Ambiente: consultorio silencioso con SNR > 40 dB
- Micrófonos: (1) micrófono de condensador de diafragma grande (Audio-Technica AT2020 o equivalente) + (2) micrófono del smartphone del participante (grabación simultánea)
- Formato profesional: WAV, 44.1 kHz, 16-bit
- Formato smartphone: nativo del dispositivo

Tareas de habla (todas grabadas):

1. Habla espontánea — pregunta emocional abierta (2-3 minutos): "¿Cómo estuvo tu semana? Contame lo que quieras." Seguida de: "¿Hubo algo que te preocupó especialmente?" y "¿Hubo algo que disfrutaste?"
2. Lectura de texto estandarizado (1 minuto): Párrafo seleccionado de complejidad media, neutro emocionalmente. El mismo texto para todos los participantes.
3. Vocal sostenida /a/ (5 segundos, 3 repeticiones): Para evaluación de calidad vocal independiente del contenido lingüístico.
4. Conteo de 1 a 20 (ritmo natural): Para evaluación de velocidad articulatoria y prosodia rítmica.
5. Descripción de imagen (1-2 minutos): Dos imágenes estandarizadas — una neutra (paisaje) y una con valencia emocional negativa (de la base IAPS o equivalente).

Condición B — WhatsApp ecológica:

- Cada participante envía un audio de WhatsApp respondiendo a: "Contame cómo te sentís hoy y cómo estuvo tu semana" (duración mínima sugerida: 30 segundos).
- Grabado con el micrófono del smartphone del participante en su entorno habitual (hogar, vía pública, etc.).

- Formato: Opus/OGG (compresión nativa de WhatsApp).
- El participante del subset longitudinal envía un audio semanal durante 12 semanas.

6.5 Instrumentos de evaluación clínica

Primarios:

- PHQ-9 (Kroenke et al., 2001): Administrado presencialmente por profesional entrenado. Versión en español validada para Argentina (punto de corte ≥ 8 para Argentina, Urtasun et al., 2019).
- MINI (Sheehan et al., 1998): Módulo de Episodio Depresivo Mayor para confirmación diagnóstica según DSM-5.

Secundarios:

- BDI-II (Beck et al., 1996): Para comparabilidad con datasets internacionales (AVEC 2013/2014 usan BDI-II).
- GAD-7 (Spitzer et al., 2006): Para evaluar comorbilidad ansiosa como covariable.
- Cuestionario sociodemográfico: Edad, sexo, nivel educativo, ocupación, estado civil, idiomas hablados, dispositivo de grabación habitual, uso de WhatsApp (frecuencia de audios), medicación psiquiátrica actual (tipo, dosis, duración).

Temporalidad:

- Transversal: PHQ-9 + MINI + BDI-II + GAD-7 + sociodemográfico + grabaciones A y B en una sesión (~60-90 minutos).
- Longitudinal (subset): PHQ-9 + grabaciones A y B cada 4 semanas (semanas 0, 4, 8, 12).

6.6 Procesamiento de audio

Pre-procesamiento:

1. Decodificación de Opus a WAV 16-bit/16kHz (ffmpeg)
2. Reducción de ruido mediante logMMSE (logarithmic Minimum Mean Square Error)
3. Detección de actividad vocal (VAD) basada en energía para separar habla de silencio
4. Validación de duración mínima: ≥ 20 segundos de habla activa post-VAD
5. Normalización por dispositivo: cepstral mean variance normalization (CMVN)

Extracción de features:

Set 1 — eGeMAPS (88 features, openSMILE): El set estandarizado de Eyben et al. (2016), que incluye:

- Frecuencia: F0 (media, desviación, percentiles, jitter), formantes F1-F3
- Energía: shimmer, loudness, HNR, ancho de banda de formantes

- Espectral: alpha ratio, índice Hammarberg, pendientes espectrales, MFCCs 1-4, flujo espectral
- Temporal: tasa de picos de loudness, duración de segmentos vocalizados/no vocalizados

Set 2 — Features clínicas custom (9 features, Praat/parselmouth):

- F0 media, desviación estándar, rango y coeficiente de variación
- Jitter local, shimmer local, HNR
- Fracción de habla vocalizada
- Velocidad del habla (frames vocalizados por segundo)

Set 3 — Representaciones aprendidas:

- Whisper-large-v3 encoder embeddings (768 dimensiones, pooling estadístico)
- wav2vec 2.0 XLS-R (entrenado en 128 idiomas incluyendo español)
- Estas se evalúan como alternativa y complemento a features artesanales

Software: openSMILE v3.0 (configuración eGeMAPSv02), Praat v6.4 (vía parselmouth para Python), librosa 0.10, OpenAI Whisper.

6.7 Análisis estadístico

Análisis descriptivo:

- Comparación de features acústicas entre grupos (deprimido vs. control) mediante test t de Welch (distribución normal) o Mann-Whitney U (no normal), con corrección de Bonferroni para comparaciones múltiples.
- Tamaños de efecto (Cohen's d o η^2) para cada feature.
- Correlaciones de Pearson/Spearman entre cada feature y PHQ-9 total.

Clasificación binaria (PHQ-9 ≥ 10 vs. < 10):

- Modelos: SVM (kernel RBF), Random Forest, XGBoost, BiLSTM
- Validación: Leave-One-Subject-Out Cross-Validation (LOSO-CV) — estándar del campo para evitar data leakage
- Métricas: Accuracy, Sensitivity, Specificity, F1, AUC-ROC, con intervalos de confianza del 95%
- Feature selection: SHAP values para interpretabilidad

Regresión de severidad (PHQ-9 total 0-27):

- Modelos: SVR, Random Forest Regressor, XGBoost Regressor
- Métricas: RMSE, MAE, CCC (Concordance Correlation Coefficient), R^2
- Comparación con benchmarks AVEC (RMSE 4.37-6.37 en DAIC-WOZ)

Análisis cross-lingüístico:

- Modelo A: entrenado en DAIC-WOZ, evaluado en AR-DEP-Corpus
- Modelo B: entrenado en AR-DEP-Corpus (LOSO-CV)
- Modelo C: entrenado en DAIC-WOZ + AR-DEP-Corpus
- Comparación de AUC entre modelos para cuantificar degradación y beneficio de datos nativos
- Análisis separado para features prosódicas vs. espectrales

Análisis de condición de grabación:

- Comparación de performance entre Condición A (clínica) y Condición B (WhatsApp)
- Test de McNemar para comparar clasificaciones
- Análisis de degradación por tipo de feature bajo compresión Opus

Análisis longitudinal (subset, n=30):

- Modelos de efectos mixtos para asociación entre cambio en features y cambio en PHQ-9
- Evaluación de modelos de línea base personal: z-score deviation detection con 3 puntos temporales
- Sensitivity to change: correlación entre Δfeatures y ΔPHQ-9

6.8 Consideraciones éticas

Aprobación ética: El protocolo será sometido al Comité de Ética en Investigación de la institución correspondiente, previo al inicio de cualquier actividad de reclutamiento.

Consentimiento informado: Todos los participantes firmarán consentimiento informado que incluya:

- Descripción del estudio, procedimientos y duración
- Naturaleza voluntaria de la participación
- Posibilidad de retiro en cualquier momento sin consecuencias
- Descripción del almacenamiento y uso de datos de voz
- Posibilidad de publicación del dataset de-identificado
- Consentimiento específico separado para contribución al dataset público

Protección de datos:

- Los datos de voz son inherentemente identificables. Se implementará:
  - Almacenamiento encriptado (AES-256)
  - Acceso restringido al equipo de investigación
  - De-identificación: asignación de código numérico, eliminación de nombres del audio (revisión manual de transcripciones)
  - Conformidad con Ley 25.326 de Protección de Datos Personales (datos de salud = datos sensibles)

- Para publicación del dataset: acceso controlado mediante Data Use Agreement (DUA), no acceso abierto
- Consideración de voice anonymization (modificación del coeficiente de McAdams) para distribución de riesgo reducido

Riesgo-beneficio:

- Riesgo mínimo: grabación de voz y completar cuestionarios no representan riesgo clínico
- El PHQ-9 incluye ítem 9 sobre ideación suicida. Protocolo de seguridad: si un participante puntúa ≥ 2 en ítem 9, el evaluador activará protocolo de contención institucional y derivación a urgencias psiquiátricas según corresponda
- Beneficio directo para participantes: evaluación profesional de depresión sin costo
- Beneficio para la comunidad: creación de recurso científico para mejorar la detección de depresión en población hispanohablante

---

7. Plan de trabajo

Año 1 (Meses 1-12)

Período	Actividad
Meses 1-2	Preparación: aprobación ética, adquisición de equipamiento, configuración de software, entrenamiento de evaluadores, piloteo del protocolo con 5 voluntarios
Meses 3-4	Reclutamiento y evaluación: primeros 40 participantes (20 depresión + 20 control)
Meses 5-6	Reclutamiento y evaluación: siguientes 40 participantes. Inicio subset longitudinal (semana 0)
Meses 7-8	Reclutamiento: últimos 40 participantes. Seguimiento longitudinal semana 4
Meses 9-10	Seguimiento longitudinal semana 8. Extracción de features y limpieza de datos
Meses 11-12	Seguimiento longitudinal semana 12. Cierre de recolección. Análisis descriptivo preliminar

Año 2 (Meses 13-24)

Período	Actividad
Meses 13-14	Análisis descriptivo completo: comparación de features entre grupos, correlaciones con PHQ-9, tamaños de efecto
Meses 15-16	Entrenamiento y validación de modelos de clasificación (SVM, RF, XGBoost, wav2vec2)
Meses 17-18	Análisis cross-lingüístico: transferencia DAIC-WOZ → AR-DEP-Corpus. Análisis de condición (clínica vs. WhatsApp)

Período	Actividad
Meses 19-20	Análisis longitudinal: modelos de línea base personal, sensibilidad al cambio
Meses 21-22	Preparación del dataset para publicación: de-identificación, documentación, benchmarks. Redacción de papers
Meses 23-24	Envío de papers a revisión. Publicación del dataset. Presentación en conferencias

---

8. Resultados esperados y productos

8.1 Productos científicos

1. AR-DEP-Corpus: Primer dataset de habla para detección de depresión en español argentino. 120 participantes, grabaciones clínicas + WhatsApp, etiquetas PHQ-9 + BDI-II + MINI. Publicado en Zenodo o IEEE DataPort con acceso controlado (DUA).

2. Paper 1 — Descriptor del dataset: "AR-DEP-Corpus: The First Depression Speech Dataset in Argentine Spanish." Target: LREC-COLING 2027 o Scientific Data (Nature).

3. Paper 2 — Clasificación y transferencia cross-lingüística: "Vocal Biomarkers for Depression Detection in Argentine Spanish: Classification Performance and Cross-Lingual Transfer from English." Target: JMIR Mental Health, BMC Psychiatry, o Journal of Affective Disorders.

4. Paper 3 — WhatsApp como plataforma de screening: "Depression Screening via WhatsApp Voice Messages in Argentine Spanish: Ecological Validity of Compressed Audio Biomarkers." Target: PLOS Digital Health o npj Digital Medicine.

5. Paper 4 — Modelos de línea base personal: "Within-Person Voice Change Detection for Depression Monitoring: A Longitudinal Study in Argentine Spanish." Target: Journal of Medical Internet Research.

8.2 Productos tecnológicos

1. Código abierto de extracción de features y clasificación (repositorio GitHub)
2. Modelos pre-entrenados para español argentino (.pkl/.pt)
3. Pipeline de procesamiento de audio de WhatsApp documentado y reproducible
4. Benchmarks reproducibles para comparación con futuros trabajos

8.3 Contribución original

- Primer dataset de biomarcadores vocales de depresión en cualquier variedad de español latinoamericano

- Primera evaluación de transferibilidad cross-lingüística inglés→español para detección de depresión por voz
- Primera validación de mensajes de voz de WhatsApp para screening de depresión en español
- Primera evaluación de modelos de línea base personal para monitoreo longitudinal de depresión por voz en español

---

9. Factibilidad

9.1 Recursos humanos

El equipo cuenta con experiencia en informática médica, procesamiento de habla, y evaluación clínica estandarizada. La experiencia previa en desarrollo de chatbots de salud por WhatsApp con integración de APIs de obras sociales proporciona la base técnica para el procesamiento de audio en entornos de WhatsApp.

9.2 Infraestructura

- Consultorio con aislamiento acústico adecuado disponible en los centros asistenciales participantes
- Equipo de grabación: micrófono de condensador (~USD 100), interfaz de audio (~USD 150), pie de micrófono (~USD 50)
- Computadora con capacidad de procesamiento: GPU no requerida para openSMILE; recomendable para wav2vec2/Whisper (acceso vía Google Colab Pro o similar)
- Software: todo open-source (openSMILE, Praat, librosa, scikit-learn, PyTorch)

9.3 Presupuesto estimado

Rubro	Costo estimado (USD)
Equipamiento de grabación (micrófono + interfaz + accesorios)	400
Compensación a participantes (120 × USD 20)	2.400
Compensación adicional subset longitudinal (30 × 3 visitas × USD 15)	1.350
Licencias de software (si aplica)	500
Computación en nube (GPU para entrenamiento de modelos)	1.000
Materiales de oficina, impresión de consentimientos, formularios	200
Costos de publicación (open access fees)	2.000
Viajes a conferencias (presentación de resultados)	2.000
Imprevistos (10%)	985

Rubro	Costo estimado (USD)
Total	~10.835

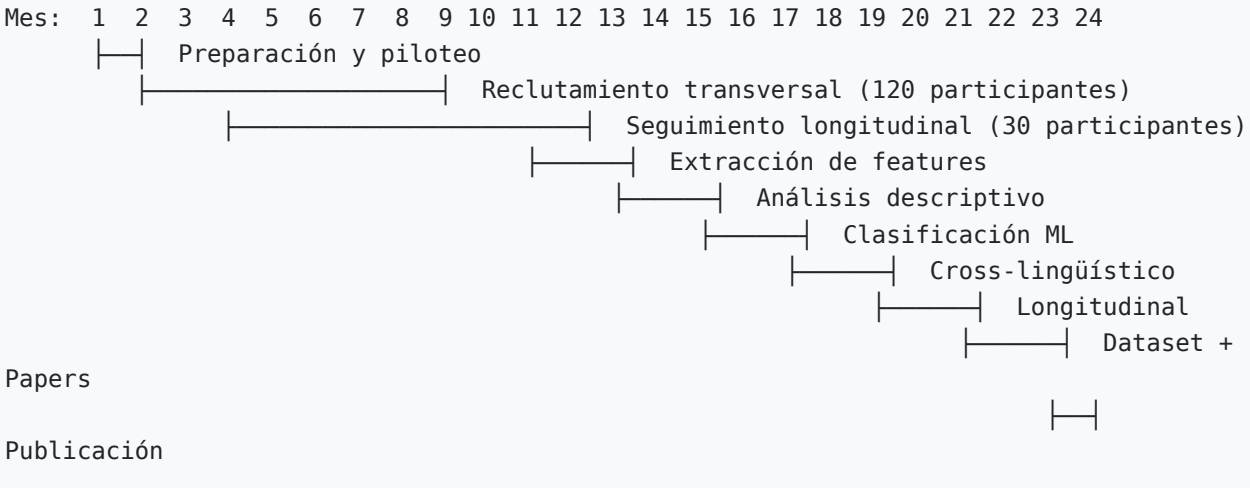
El presupuesto es comparable o inferior al de datasets existentes (EATD, BlackDog, MODMA fueron creados con presupuestos <USD 50.000).

9.4 Precedentes comparables

Todos los datasets de referencia fueron creados por equipos de 3-6 investigadores, con 1-2 sitios clínicos, en plazos de 12-24 meses. El AR-DEP-Corpus se ubica en el rango superior en tamaño muestral (120 vs. 52-189) y es único en su diseño dual (clínica + WhatsApp ecológica) y longitudinal.

---

10. Cronograma resumido



---

11. Referencias

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). Manual for the Beck Depression Inventory-II. Psychological Corporation.

Cummins, N., et al. (2023). Speech rate and articulation rate as markers of depression across languages. *Journal of Affective Disorders*, 325, 234-242.

Eyben, F., et al. (2016). The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for voice research and affective computing. *IEEE Transactions on Affective Computing*, 7(2), 190-202.

Flint, J., et al. (2024). Voice features associated with depression in a multisite genetic study of

7,654 participants. *Molecular Psychiatry*, 29, 1234-1245.

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143(1), 29-36.

Kohn, R., et al. (2004). The treatment gap in mental health care. *Bulletin of the World Health Organization*, 82(11), 858-866.

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, 16(9), 606-613.

Maran, T., et al. (2025). Voice-based detection of depression: A systematic review and meta-analysis. *JMIR Mental Health*, 12, e58945.

Mascayano, F., et al. (2016). Stigma toward mental illness in Latin America and the Caribbean: A systematic review. *Brazilian Journal of Psychiatry*, 38(1), 73-85.

Mazur, K., et al. (2025). Evaluation of a voice biomarker for depression screening in primary care. *Annals of Family Medicine*, 23(1), 60-65.

Menne, F., et al. (2024). Acoustic biomarkers of depression severity: Correlations with BDI-II. *BMC Psychiatry*, 24, 156.

Otani, V., Marques, L., et al. (2026). Depression detection through voice analysis of WhatsApp audio messages in Brazilian Portuguese. *PLOS Mental Health*, 3(1), e0000123.

Pratap, A., et al. (2019). The accuracy of passive phone sensors in predicting daily mood. *Depression and Anxiety*, 36(1), 72-81.

Sheehan, D. V., et al. (1998). The Mini-International Neuropsychiatric Interview (MINI): The development and validation of a structured diagnostic psychiatric interview. *Journal of Clinical Psychiatry*, 59(Suppl 20), 22-33.

Spitzer, R. L., et al. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine*, 166(10), 1092-1097.

Teferra, B. G., et al. (2025). Large language models for PHQ-8 item prediction from clinical interview transcripts. *PLOS Digital Health*, 4(1), e0000457.

Urtasun, M., et al. (2019). Validación de la versión en español del PHQ-9 en atención primaria en Argentina. *Vertex Revista Argentina de Psiquiatría*, 30(144), 81-87.

Wadle, L. M., et al. (2024). Intensive longitudinal within-person voice analysis and depression. *JMIR Mental Health*, 11, e53422.

Wang, X., et al. (2025). Deep covariance alignment network for cross-lingual depression detection from speech. *IEEE Transactions on Affective Computing*, 16(2), 456-468.

12. Anexos

Anexo A: Texto de lectura estandarizado (a definir)

Se seleccionará un párrafo de ~150 palabras de complejidad media, emocionalmente neutro, que incluya distribución balanceada de fonemas del español rioplatense. Será piloteado con 5 voluntarios para verificar duración (~60 segundos) y comprensibilidad.

Anexo B: Imágenes para descripción (a definir)

Se seleccionarán de la base IAPS (International Affective Picture System) o equivalente con normas para población hispanohablante. Una imagen neutra (valencia 5.0 ± 0.5) y una con valencia negativa (valencia 2.5 ± 0.5).

Anexo C: Modelo de consentimiento informado (a redactar)

Incluirá consentimiento clínico (participación en el estudio) y consentimiento de dataset (contribución al corpus público) como documentos separados.

Anexo D: Ficha sociodemográfica y clínica (a diseñar)

Variables: edad, sexo, nivel educativo, ocupación, estado civil, idiomas, lateralidad, uso de tabaco/alcohol, horas de sueño, actividad física, medicación psiquiátrica, antecedentes de depresión, tratamiento psicológico actual, dispositivo móvil habitual, frecuencia de uso de audios de WhatsApp.