



e-book exclusivo

Panorama da evolução recente de biomarcadores vocais na Psiquiatria

Panorama da Evolução Recente de Biomarcadores Vocais na Psiquiatria

Introdução

Transtornos mentais comuns, como depressão e ansiedade, afetam centenas de milhões de pessoas mundialmente ¹. Apesar da carga global dessas doenças, a psiquiatria historicamente carece de exames objetivos de rotina – diferentemente de outras áreas da medicina, não há “biomarcadores” amplamente utilizados em diagnóstico ou acompanhamento ². O diagnóstico baseia-se sobretudo em entrevistas clínicas e autorrelatos, sujeitos a vieses de memória e julgamento ³. Nesse contexto, *biomarcadores vocais* – isto é, características da voz e da fala associadas a estados mentais – emergiram nos últimos anos como uma promessa para suprir essa necessidade de medidas objetivas em saúde mental ⁴. A análise de fala oferece uma ferramenta *não invasiva, acessível e de baixo custo*: a voz pode ser captada via smartphone praticamente em qualquer lugar, inclusive de forma passiva no cotidiano do paciente ⁴. Além disso, por envolver processos neuromotores e cognitivo-linguísticos complexos, a produção de voz é sensível a diversas patologias neuropsiquiátricas ⁵ – por exemplo, alterações no timbre, ritmo ou entonação podem refletir mudanças no humor, pensamento ou nível de ativação psicomotora. Nos últimos anos, pesquisas em *Machine Learning (ML)* e *inteligência artificial (IA)* aplicadas à voz têm avançado rapidamente, alimentando a expectativa de que a voz possa se tornar um “sinal vital” digital na psiquiatria para detecção, triagem e monitoramento de transtornos mentais.

Avanços Científicos Recentes na Detecção de Biomarcadores Vocais

Nas últimas duas décadas – e em especial de 2018 em diante – houve um aumento exponencial de pesquisas explorando sinais vocais associados a transtornos psiquiátricos. Quase metade dos estudos focam em depressão maior, mas há trabalhos significativos voltados a esquizofrenia, transtorno bipolar, transtorno de estresse pós-traumático (TEPT) e outros ⁶. De forma geral, os resultados têm demonstrado que *parâmetros acústicos da fala diferem significativamente entre indivíduos com transtornos psiquiátricos e controles saudáveis*. Uma revisão sistemática publicada em 2025 identificou diversos estudos onde medidas de voz distinguiam pacientes deprimidos de pessoas sem depressão com AUC (área sob a curva) variando de 0,71 até 0,93, e acurácia de classificação entre ~78% e 96.5% ⁷ ⁸. Nesses trabalhos, características **prosódicas** (como pitch médio e variabilidade tonal), **espectrais** (padrões de frequências da voz) e **de qualidade vocal** (perturbações como jitter e shimmer) mostraram correlação significativa com a gravidade da depressão medida por escalas clínicas padronizadas ⁷. Em outras palavras, pacientes depressivos tendem a apresentar voz mais **monótona**, **“sem energia”** e **com lentificação** – descrição clínica clássica que encontra suporte em dados objetivos ⁹. Estudos de 2023-2024 confirmam essas observações: indivíduos em episódio depressivo frequentemente falam mais devagar, com mais pausas longas, mais hesitações e irregularidades sutis na voz, em comparação a pessoas saudáveis ¹⁰.

Além da depressão, *outros transtornos psiquiátricos também exibem possíveis biomarcadores vocais característicos*. Na **esquizofrenia**, por exemplo, há evidências de que a fala reflete sintomas como *afeto embotado*, *alogia* (pobreza de fala) e *disfunções do pensamento*. Um estudo de 2023 combinou medidas acústicas e de patologia vocal para classificar pacientes com esquizofrenia versus depressão e

controles, alcançando acurácia acima de 90% em cada comparação ¹¹ ¹². As características de voz mais importantes para distinguir a esquizofrenia envolveram *redução na coordenação articulatória, maior número de pausas e menor variabilidade na fala* – padrões compatíveis com a lentificação psicomotora e o empobrecimento da fluência verbal típicos do transtorno ¹³. Já no **transtorno bipolar**, a mudança de humor entre depressão e mania produz mudanças dramáticas na voz: durante *episódios maníacos*, os pacientes costumam falar muito mais rápido, em volume alto e com pressão de fala, enquanto na *depressão bipolar* ocorre o oposto (fala lenta, baixa, monotônica) ¹⁴. Pesquisas recentes aproveitam essas diferenças para identificar o estado de humor de pacientes bipolares a partir da voz. Por exemplo, um estudo de 2024 com gravações longitudinais de pacientes bipolares relatou que modelos de IA conseguiram diferenciar automaticamente trechos de fala gravados em estado depressivo versus maníaco com cerca de 80% de acurácia, atingindo aproximadamente 77% de sensibilidade para detectar depressão e 55% para mania ¹⁵. Isso demonstra potencial para usar a voz como um **indicador objetivo do estado afetivo** em transtorno bipolar, embora a detecção de mania ainda apresente desafios (sensibilidade mais baixa para mania, possivelmente por menor quantidade de dados ou maior variabilidade individual) ¹⁵.

No caso do **TEPT**, avanços notáveis também foram registrados. Um estudo pioneiro conduzido com veteranos de guerra mostrou que algoritmos de *aprendizado de máquina* podem identificar com alta acurácia quem tem TEPT analisando apenas características de sua fala em entrevistas clínicas. Nesse estudo, um classificador de *Random Forest* treinado em milhares de atributos de voz atingiu **89% de acurácia** ao distinguir veteranos com TEPT de controles, confirmando quantitativamente observações antes apenas anedóticas – por exemplo, pacientes com TEPT tendem a falar com um *tom “metálico” e pouco claro*, algo que clínicos já associavam ao transtorno ¹⁶ ¹⁷. Esses achados reforçam que a voz carrega marcas detectáveis de diversas condições psiquiátricas.

É importante notar que, embora os resultados recentes sejam promissores, nem todos os estudos reportam desempenhos elevados indistintamente. A eficácia pode variar conforme o transtorno e as condições de coleta de dados. Por exemplo, um trabalho que examinou simultaneamente depressão, bipolar e esquizofrenia encontrou boa discriminação entre pacientes e controles (AUC entre 0,75 e 0,92), mas teve **dificuldade para distinguir depressão unipolar de transtorno bipolar** apenas pela voz (AUC ~0,50, equivalente ao acaso) ¹⁸. Isso sugere que episódios depressivos em bipolar e em depressão maior podem produzir padrões vocais semelhantes, limitando a especificidade do biomarcador nesses casos ¹⁸. Mesmo assim, o conjunto de pesquisas dos últimos 5 anos consolidou a prova de conceito de que **biomarcadores vocais são reais e detectáveis**, associados a sintomas como humor deprimido, ansiedade, pensamento desorganizado e outros – abrindo caminho para aplicações clínicas inovadoras.

Principais Abordagens e Técnicas Utilizadas

A detecção de biomarcadores vocais psiquiátricos combina técnicas de **processamento de sinais de voz** com algoritmos de **inteligência artificial**. Em termos gerais, há duas abordagens que frequentemente se complementam:

- **Extração de características acústicas manuais:** Muitos estudos começam extraindo dezenas ou centenas de *features* da voz por meio de ferramentas de análise de áudio. Essas características incluem medidas **prosódicas** (e.g., frequência fundamental média – *pitch* – e sua variação; duração de pausas; velocidade de fala), medidas de **qualidade vocal** (e.g., *jitter* e *shimmer*, que quantificam microvariações na frequência e amplitude da voz indicando instabilidade vocal; razão harmônico-ruído que indica sopro), e medidas **espectrais** (e.g., coeficientes cepstrais de frequência Mel – MFCCs – que capturam a forma do espectro sonoro,

formantes vocais, entropia espectral etc.). Por exemplo, em depressão, encontram-se frequentemente valores elevados de jitter e shimmer (indicando uma voz menos estável), diminuição na variabilidade do *pitch* e fala mais pausada em relação a indivíduos saudáveis ¹⁰. Essas *features* servem de insumo para algoritmos de ML que aprendem a distinguir padrões de pacientes vs. controles. Modelos clássicos incluem máquinas de vetor de suporte (SVM), florestas aleatórias (*random forest*) e redes neurais artificiais de porte moderado. Um caso ilustrativo foi o do estudo sobre TEPT, no qual os pesquisadores **alimentaram 40.526 atributos de voz** extraídos de entrevistas de veteranos a um algoritmo random forest para “ensinar” o modelo a reconhecer combinações de traços vocais indicativos do transtorno ¹⁹ ²⁰. Nesse processo, o algoritmo identificou como relevantes sinais como *falta de clareza na articulação e tom de voz “sem vida”* característicos de quem sofre TEPT ¹⁷.

- **Aprendizado profundo e modelos de IA de ponta:** Em anos recentes, houve uma transição para métodos baseados em *deep learning*, que podem dispensar a seleção manual de atributos e aprender diretamente os padrões relevantes a partir das gravações de áudio brutas ou de seus espectrogramas. Redes neurais recorrentes (RNNs) e arquiteturas derivadas, como **LSTM** (Long Short-Term Memory) e **GRU** (Gate Recurrent Unit), já foram aplicadas para detecção de depressão e detecção de humor bipolar, muitas vezes combinando dados de voz com transcrições de fala. Por exemplo, um estudo de 2024 empregou um modelo GRU pré-treinado em fala espontânea em língua chinesa, adaptado para classificar estados de humor de pacientes bipolares, obtendo melhor desempenho do que modelos tradicionais ²¹ ¹⁵. Já empresas como a Ellipsis Health relatam o uso de arquiteturas de *deep learning* multimodais que **combinam padrões acústicos e linguísticos**: modelos que analisam tanto as propriedades sonoras da voz quanto as escolhas de palavras e conteúdo do discurso ²². Essa abordagem integrada pode capturar indicadores como hesitações, monotonia (sinal acústico) junto com fala negativa ou desorganizada (sinal semântico) para aprimorar a acurácia na detecção de depressão e ansiedade. Conforme descrito pelo fundador da Ellipsis, esses modelos avançados aprendem características diretamente dos dados de voz e texto transcrito, sem depender apenas de *features* pré-determinadas, utilizando técnicas modernas de aprendizagem por transferência para melhorar a robustez ²².

Outra técnica relevante é o uso de *representações de voz pré-treinadas*: modelos de IA genéricos (como *wav2vec*, da área de reconhecimento de fala) que são ajustados posteriormente para tarefas psiquiátricas com amostras rotuladas. Isso permite aproveitar grandes quantidades de dados não rotulados de fala para aprender representações acústicas úteis, que depois discriminam melhor, por exemplo, entre fala depressiva e não-depressiva mesmo com conjuntos de treinamento menores.

Além dos aspectos acústicos, algumas abordagens incorporam análise de **conteúdo verbal** (NLP). Especialmente em transtornos como esquizofrenia ou outros quadros psicóticos, a forma como a linguagem é organizada pode fornecer pistas – e.g., perda de coerência no discurso, uso de frases desconexas ou neologismos. Modelos de processamento de linguagem natural podem quantificar a coesão semântica ou peculiaridades lexicais, funcionando como biomarcadores complementares aos sinais puramente vocais. De fato, algoritmos híbridos vêm sendo propostos: analisando “*tons, pausas e entonações*” em conjunto com aspectos semânticos da fala transcrita ²³. No entanto, vale ressaltar que a maior parte dos biomarcadores de voz em psiquiatria até agora concentra-se em traços acústicos (prosódia e voz), já que o conteúdo verbal pode ser muito dependente do idioma e cultura.

Em suma, os *principais avanços tecnológicos* incluem: (1) **extração massiva de parâmetros de voz** via ferramentas como OpenSMILE, Praat etc., permitindo caracterizar matematicamente a fala; (2) **algoritmos de ML** que filtram quais combinações de parâmetros distinguem pacientes de controles (e até predizem gravidade do transtorno); (3) **redes neurais profundas** aprendendo automaticamente

padrões complexos em dados de áudio; e (4) **sistemas híbridos áudio-texto** que unem sinais vocais e de linguagem. Essa convergência de processamento de sinais e IA tornou possível detectar padrões sutis não acessíveis ao ouvido humano destreinado – por exemplo, variações minúsculas na estabilidade da voz (jitter) ou na ênfase de certas frequências – elevando o conceito de biomarcadores vocais a um novo patamar científico. No entanto, esses métodos sofisticados trazem consigo desafios de *explicabilidade* e *generalização*, discutidos adiante na seção de limitações.

Aplicações Clínicas Atuais e em Desenvolvimento

As aplicações das tecnologias de biomarcadores vocais na psiquiatria já começam a se delinear, embora ainda estejam em fases de pesquisa, validação ou uso piloto. Podemos destacar três grandes frentes de aplicação clínica:

- **Triagem e rastreamento populacional:** Uma das utilizações mais promissoras é incorporar análises de voz em *screenings* de saúde mental, especialmente em atenção primária ou mesmo via aplicativos de bem-estar. Por exemplo, na atenção básica é recomendada triagem universal para depressão, mas estima-se que menos de 5% dos pacientes são de fato rastreados rotineiramente ²⁴. Ferramentas de voz podem automatizar esse processo de forma quase invisível, sem aumentar a carga de trabalho dos profissionais. Um estudo recente avaliou a eficácia de um algoritmo de voz (Kintsugi Voice) em detectar depressão moderada a grave a partir de **apenas 25 segundos de fala livre** coletada via aplicativo móvel ²⁵ ²⁶. Com uma amostra de ~15 mil adultos, o modelo alcançou cerca de **71% de sensibilidade e 73% de especificidade** para identificar indivíduos com PHQ-9 \geq 10 (sintomas depressivos significativos) ²⁶. Esses números, embora abaixo da perfeição, sugerem que a triagem vocal poderia *complementar* os questionários tradicionais, servindo como um alerta inicial. Integrada em plataformas de telemedicina ou até em centrais de atendimento telefônico, a análise vocal poderia sinalizar quais pessoas possivelmente precisam de avaliação aprofundada – ampliando a cobertura do rastreio de transtornos como depressão e ansiedade sem exigir que cada paciente preencha formulários. Empresas já estão explorando essa frente: a Sonde Health, por exemplo, lançou em 2021 um app chamado “Sonde Mental Fitness” no qual o usuário fala por 30 segundos respondendo perguntas genéricas, e o sistema retorna um escore de bem-estar mental, indicando níveis de possível depressão, ansiedade ou fadiga ²⁷. *Startups* como a Kintsugi e Ellipsis visam integrar suas APIs de biomarcador vocal em plataformas de saúde digital para triagem rápida em clínicas, planos de saúde ou mesmo em assistentes virtuais.
- **Diagnóstico assistido e apoio à decisão clínica:** Outro campo em desenvolvimento é o uso de biomarcadores vocais como *ferramenta auxiliar* para psiquiatras, psicólogos e outros profissionais durante o diagnóstico. Não se propõe substituir a avaliação clínica, mas oferecer dados objetivos que possam corroborar ou quantificar impressões clínicas. Por exemplo, um algoritmo pode, em segundo plano, analisar a fala do paciente durante uma consulta (presencial ou por teleconferência) e gerar indicadores numéricos – como um escore de “identificação psicomotora vocal” ou um nível de probabilidade de depressão com base na voz. Esses indicadores poderiam ajudar a identificar casos sutis ou confirmar a gravidade relatada. **Importante:** há um debate se diagnosticar transtornos (como “depressão: sim/não”) via IA é realmente útil aos clínicos, dado que profissionais já conseguem diagnosticar pela entrevista. Alguns pesquisadores argumentam que a utilidade maior estaria em medir *sintomas específicos e sua evolução* ²⁸. De todo modo, já existem iniciativas de diagnóstico assistido: a Ellipsis Health desenvolveu um sistema que, após o paciente falar ~1 minuto sobre qualquer assunto, gera automaticamente escores de gravidade de depressão e ansiedade. Essa tecnologia está sendo pilotada em mais de 9 estados nos EUA e em projetos internacionais, inclusive por uma grande seguradora (Cigna) que a utilizou para criar a ferramenta “StressWaves”, a qual visualiza o nível

de estresse de uma pessoa a partir da voz e sugere exercícios de bem-estar ²⁹. Vale ressaltar que, até 2025, nenhum desses sistemas obteve aprovação regulatória plena (FDA) como dispositivo diagnóstico ³⁰ – no máximo são utilizados como *support tools* experimentais. Assim, a aplicação diagnóstica ainda é emergente. A expectativa é que, com mais validação, esses algoritmos possam realizar triagens automáticas (por exemplo, em salas de espera, identificar quem possivelmente tem depressão não diagnosticada) ou apoiar clínicos indicando sinais objetivos – por exemplo, informar que “o padrão de voz deste paciente está 80% similar ao de indivíduos com depressão severa” – cabendo ao profissional integrar essa informação ao conjunto diagnóstico.

- **Monitoramento e manejo longitudinal:** Talvez a aplicação mais consolidada seja o **monitoramento de sintomas ao longo do tempo**. Transtornos psiquiátricos são dinâmicos – os sintomas flutuam, podem agravar-se ou melhorar com tratamento – e a voz pode oferecer um meio de captar essas mudanças de forma contínua entre as consultas médicas. Projetos como o *PRIORI* (Universidade de Michigan) demonstraram que smartphones podem coletar automaticamente amostras de voz de pacientes bipolares (por ex., durante ligações) e alertar para mudanças que precedem recaídas maníacas ou depressivas ¹⁴ ³¹. No *PRIORI*, um aplicativo instalado no telefone gravava fala de forma segura e algoritmos detectavam alterações nos elementos físicos da fala (quantidade de fala, pitch, ritmo, volume), comparando com a linha de base do próprio paciente ³². Isso permitiu identificar tendências de virada de humor antes mesmo que o paciente percebesse, possibilitando intervenção precoce. De forma análoga, para depressão unipolar, já se vislumbra o uso de *apps* que, ao longo do dia ou semanalmente, coletam breves amostras de voz (ou analisem ligações voluntariamente) e forneçam ao paciente e ao médico um histórico objetivo do estado emocional. Um estudo de 2023 testou um aplicativo de “fitness mental” com 104 pacientes psiquiátricos ambulatoriais: o app gerava um escore de biomarcador vocal (calculado a partir de 8 parâmetros chave da voz) várias vezes por semana, e os resultados mostraram que esse escore acompanhava de perto a flutuação global de sintomas dos pacientes ³³ ³⁴. Quando o escore vocal se elevava indicando piora, o risco relativo de o paciente apresentar sintomas psiquiátricos severos (segundo uma escala clínica de referência) chegava ao dobro do normal; e nos pacientes que usavam o app quase diariamente, o risco de piora detectado chegou a ser **8,5 vezes maior** nos períodos indicados pelo biomarcador vocal elevado ³⁴. Ou seja, o monitoramento frequente permitiu antecipar agravamentos com muito mais sensibilidade. Os participantes também relataram achar a ferramenta útil e fácil de usar no cotidiano, evidenciando boa aceitabilidade ³⁵. Esse tipo de solução abre a porta para um acompanhamento *digital* contínuo: o psiquiatra poderia receber alertas se a voz do paciente começar a indicar recaída, auxiliando na decisão de contactar o paciente ou ajustar o tratamento antes que a situação piore significativamente.

Outras aplicações em desenvolvimento incluem o uso de biomarcadores vocais em **contextos específicos**, como: triagem de estresse pós-traumático em militares retornados de combate; detecção precoce de *delirium* ou sofrimento mental em pacientes hospitalizados (pela voz ao responder perguntas diárias); ferramentas para terapeutas acompanharem a melhora de pacientes sessão a sessão (voz do paciente ao relatar semana pode indicar melhora antes mesmo da autoavaliação mudar, por exemplo); e até integração com *assistentes virtuais* (p.e., a Alexa ou Google Assistant identificando sinais de depressão na voz do usuário – algo já explorado em patentes e protótipos por grandes empresas de tecnologia). Em resumo, as aplicações clínicas dos biomarcadores vocais abrangem **prevenção (triagem populacional), diagnóstico assistido e tratamento (monitoramento e intervenção precoce)**. Atualmente, a maioria está na fase de estudos piloto ou produtos iniciais, mas os próximos anos serão cruciais para validar sua eficácia e segurança antes de uma adoção ampla.

Vantagens e Limitações das Abordagens Atuais

Como toda tecnologia emergente, os biomarcadores vocais na psiquiatria apresentam **vantagens atraentes**, mas também **limitações e desafios** a serem superados.

Vantagens e Potenciais Benefícios:

- **Não-invasivo, acessível e de baixo custo:** A coleta de dados de voz não requer exame físico, laboratorial ou aparelhos caros – *basta um microfone*. Praticamente todo smartphone pode servir de dispositivo de captura, o que torna a solução escalável globalmente ⁴. Isso é especialmente vantajoso em saúde mental, onde muitos pacientes não têm acesso fácil a especialistas; a triagem vocal poderia ocorrer remotamente, ampliando o alcance dos serviços.
- **Aplicável a ambientes naturais e monitoramento contínuo:** Diferente de avaliações no consultório (momentâneas), a voz pode ser coletada ao longo do dia ou em diferentes contextos da vida real do paciente, inclusive de forma passiva. Esse monitoramento longitudinal permite detectar variações sutis no estado mental (*p.ex.*, piora gradual da depressão) que poderiam passar despercebidas em avaliações espaçadas. Estudos indicam que agregados de dados vocais ao longo de semanas melhoram significativamente a detecção de mudanças clínicas ³⁴. Em suma, a voz viabiliza uma espécie de *sensor contínuo* do bem-estar mental.
- **Objetividade e quantificação de sintomas:** Os algoritmos fornecem medidas numéricas padronizadas (escores) relacionados a sintomas, reduzindo a subjetividade. Por exemplo, o ritmo de fala ou extensão de pausas pode ser quantificado objetivamente, refletindo o grau de lentificação psicomotora de um paciente depressivo ¹³. Isso pode ajudar a avaliar respostas a tratamentos de forma mais granular do que apenas impressão clínica. O uso de medidas objetivas pode igualmente diminuir vieses de observador – algo relevante considerando que diagnósticos psiquiátricos podem variar entre profissionais.
- **Detecção precoce e pró-atividade:** Biomarcadores vocais podem captar sinais iniciais de transtornos antes que se tornem graves o bastante para diagnóstico formal. Startups alegam que certos padrões sutis na entonação e cadência podem surgir **meses ou anos antes** de um episódio depressivo declarado ³⁶. Se comprovado, isso possibilitaria intervenções precoces preventivas. Mesmo em contextos agudos, a rapidez da análise (basicamente em tempo real após coleta da voz) pode agilizar decisões – por exemplo, triando automaticamente pacientes de alto risco de ideação suicida na fila de espera de emergências psiquiátricas, direcionando-os prioritariamente.
- **Escalabilidade com IA e telemedicina:** A incorporação em plataformas digitais permite que *um algoritmo realize milhares de avaliações simultâneas*, algo impossível apenas com mão de obra humana especializada. Em cenário de escassez global de psiquiatras, tais ferramentas podem ajudar a filtrar casos que necessitam de atenção imediata ³⁷ e oferecer suporte a provedores menos especializados (clínicos gerais, etc.), guiando-os quando encaminhar ao especialista ³⁸. Em telepsiquiatria, analisadores de voz podem rodar em segundo plano durante as consultas virtuais sem interferir, adicionando uma camada extra de informação clínica.
- **Melhoria contínua via aprendizado de máquina:** À medida que mais dados são coletados, os modelos podem ser refinados para aumentar a acurácia e abranger populações diferentes. Ou seja, trata-se de uma tecnologia “viva” que tende a ficar mais precisa com o tempo e com a

diversidade de uso, eventualmente podendo atingir níveis de desempenho muito altos para certas tarefas.

Limitações e Desafios Atuais:

- **Precisão diagnóstica limitada e variação nos resultados:** Apesar de muitos estudos relatando acurácia acima de 80-90%, meta-análises sugerem que, em média, o desempenho dos algoritmos atuais ainda fica aquém de instrumentos convencionais em alguns cenários. Por exemplo, uma revisão de 2024 observou performances médias de sensibilidade ~78% e AUC ~0,78 para detecção de depressão via voz – resultados *inferiores aos do questionário clínico PHQ-9* aplicado pelos próprios pacientes ³⁹ ⁴⁰. Isso indica que os modelos ainda geram *falsos negativos* e *falsos positivos* em número significativo. Em contextos críticos (como risco de suicídio), depender de uma ferramenta com 80% de acerto pode ser arriscado. Além disso, há grande heterogeneidade entre estudos: diferentes bases de dados, línguas, métodos de extração e algoritmos dificultam comparar resultados e estabelecem muita variação. Essa falta de consistência sugere que a tecnologia não está totalmente madura – alguns modelos funcionam muito bem em seu conjunto de teste, mas podem falhar ao generalizar fora dele.
- **Problemas de generalização e viés:** Uma preocupação séria é se os biomarcadores vocais treinados em um determinado conjunto populacional vão funcionar em outros contextos. A voz humana varia por idioma, sotaque, gênero, faixa etária, cultura e até características individuais da personalidade. Modelos treinados majoritariamente com um perfil de falantes podem apresentar viés e erro sistemático ao analisar vozes distintas. Casos reais já foram documentados: a empresa Winterlight Labs, por exemplo, descobriu que seu algoritmo inicialmente promissor só funcionava bem para falantes de um dialeto específico do inglês canadense – quando aplicado a pessoas de outros sotaques, o desempenho caiu drasticamente ⁴¹. Outro estudo revelou disparidades preocupantes: um sistema identificava falsamente sinais de depressão em até 40% de pessoas negras que na verdade eram saudáveis, devido a vieses nos dados de treinamento ⁴². Esse tipo de viés algorítmico pode amplificar desigualdades e gerar **erros éticos graves**, rotulando indevidamente minorias ou grupos não representados. Portanto, superar o problema de generalização requer treinar IA com **dados amplos e diversos**, representando múltiplos idiomas, etnias, idades e condições, e ainda assim realizar validações externas rigorosas. Atualmente, muitos estudos usam bases limitadas (algumas com poucas dezenas de pacientes) ou corpora coletados em condições artificiais, o que restringe a aplicabilidade universal.
- **Falta de padronização de protocolos:** Não há consenso ainda sobre *quais tarefas de fala ou quais parâmetros* são ideais para se extrair. Diferentes estudos usam métodos distintos – alguns pedem ao participante para ler um texto padronizado, outros usam respostas livres a perguntas, outros analisam conversas espontâneas. Cada abordagem pode enfatizar aspectos vocais diferentes. Essa falta de padronização dificulta a replicabilidade e também a tradução para prática clínica. Será necessário definir protocolos ótimos (por exemplo, um conjunto breve de tarefas de fala padronizadas que maximizem a revelação de biomarcadores para certo transtorno) para que diferentes clínicas e sistemas usem as mesmas referências. Só assim resultados de um local poderão ser comparados a de outro. Instituições como NIH e consórcios internacionais já discutem a criação de *datasets* compartilhados e desafios (challenges) padronizados para acelerar esse alinhamento, mas até 2025 isso ainda está em progresso.
- **Questões regulatórias e validação clínica insuficiente:** Até o momento, nenhuma ferramenta de biomarcador vocal foi aprovada como dispositivo médico diagnóstico pelas agências reguladoras (FDA nos EUA, EMA na UE, Anvisa no Brasil etc.) ³⁰. Para que isso ocorra, serão

necessárias evidências robustas de eficácia e segurança em estudos clínicos amplos. Muitos dos resultados publicados vêm de estudos de prova de conceito ou coortes limitadas; poucos foram testados em ambientes clínicos reais controlados. Falta demonstrar, por exemplo, que usar o biomarcador vocal melhora de fato os desfechos do paciente (como identificar mais gente que precisa de tratamento e assim reduzir sintomas médios na população). Sem essa validação, essas ferramentas permanecerão restritas a pesquisa ou ao uso não regulamentado de bem-estar. A aprovação regulatória exigirá padronização, provas de que o algoritmo funciona para os subgrupos populacionais indicados sem viés, e planos de vigilância pós-mercado. Trata-se de um caminho que deve ser percorrido nos próximos anos.

- **Privacidade e aceitação ética:** A voz é um dado pessoal sensível – pode revelar não apenas emoções, mas também identidade, origem geográfica, gênero e estado de saúde. O uso de biomarcadores vocais levanta preocupações de privacidade: pacientes podem temer estar sendo “vigiados” ou analisados sem consentimento (por exemplo, aplicativos coletando voz em segundo plano). É fundamental garantir que haja consentimento informado claro e que os dados de voz sejam armazenados de forma segura (idealmente anonimizados) para evitar usos indevidos. Além disso, existe um **risco de mau uso** dessas tecnologias fora do contexto médico: imaginemos empregadores, seguradoras ou até sistemas jurídicos tentando inferir estado mental de alguém pela voz, sem controle clínico – isso seria eticamente problemático. Especialistas alertam para a possibilidade de “industrialização da saúde mental”, em que decisões automatizadas a partir da voz possam ocorrer em escolas, empresas ou triagens legais, estigmatizando indivíduos ⁴³. Portanto, junto ao avanço técnico, será necessário desenvolver diretrizes éticas e legais claras para evitar discriminação e preservar a autonomia e dignidade das pessoas avaliadas.
- **Interpretação clínica e integração nos fluxos de trabalho:** Mesmo que uma ferramenta vocal seja acurada, é preciso que seja útil e utilizada pelos profissionais. Alguns clínicos podem resistir em adotar uma “caixa-preta” algorítmica sem entender completamente como ela toma decisões. Embora um estudo (HAL, 2024) argumente que a falta de explicabilidade não é a barreira principal ⁴⁴, ainda assim é importante fornecer *outputs* intuitivos. Por exemplo, um psiquiatra talvez valorize mais um feedback do tipo “indicador de lentificação de fala aumentou 30% em relação à visita anterior” do que um vago “score de depressão = 0.65”. Integrar esses sistemas ao prontuário eletrônico de maneira que se encaixem no fluxo de trabalho (em vez de acrescentar burocracia) também será um fator decisivo na adoção. A tecnologia só trará benefício real se for **aceita pelos usuários finais (médicos e pacientes)** e se agregar valor claro na tomada de decisão.

Resumindo, *as abordagens atuais de biomarcadores vocais são promissoras*, mas enfrentam obstáculos. A precisão dos modelos ainda deve melhorar e se tornar mais consistente entre diferentes populações; é preciso mitigar vieses e provar generalização; estabelecer protocolos uniformes; obter validações clínicas robustas e aprovações regulatórias; e abordar as preocupações éticas e de usabilidade. **Vantagens como acessibilidade, não-invasividade e potencial de monitoramento contínuo conferem enorme apelo à tecnologia**, mas **as limitações atuais impedem seu uso irrestrito na clínica**. Os próximos passos da pesquisa e desenvolvimento buscarão justamente transformar esses desafios em oportunidades de aprimoramento.

Estudos e Publicações Recentes Relevantes (2023–2025)

Para ilustrar o estado da arte, a seguir estão exemplos de estudos e artigos recentes (até 2024/25) que marcaram o campo de biomarcadores vocais em saúde mental, com seus achados principais:

Estudo (Ano)

Resumo e Resultados Principais

Briganti & Lechien (2025) – Revisão Sistemática de IA aplicada à voz na depressão ⁴⁵ ⁷ .

Analisou 12 estudos (16.872 participantes, incluindo >1.500 com depressão). Constatou que medidas de fala diferenciam consistentemente depressão de controles saudáveis (AUC entre 0,71–0,93; acurácias 78–96%) ⁷ . Medidas prosódicas (entonação, ritmo), espectrais e de perturbação vocal (jitter, shimmer) mostraram **correlação** com escores padronizados de depressão ⁷ . Conclui que a voz é um **biomarcador promissor para detecção e monitorização da depressão**, mas destaca a **heterogeneidade metodológica e problemas de generalização**. Seis dos 12 estudos tinham alto risco de viés (especialmente na seleção de pacientes), e contextos de gravação variavam (clínica vs. smartphone) ⁸ ⁴⁶ . Recomenda padronizar métodos e validar modelos em amostras independentes antes da adoção clínica em larga escala ⁴⁶ .

Donaghy et al. (2024) – Revisão Narrativa de ML para detectar voz na depressão ⁴⁷ ³⁹ .

Revisou 19 estudos publicados de 2019 a 2022 que aplicaram aprendizado de máquina em voz para identificar depressão. 17 trabalhos faziam classificação (depressão vs. não) e 2 regressão (predizer escore de depressão). Houve grande variedade de métricas utilizadas; em média, alcançaram **sensibilidade ~0,78, especificidade ~0,78 e AUC ~0,78**, dependendo do estudo ³⁹ . Apesar de promissores, esses índices ficaram abaixo da acurácia típica do questionário PHQ-9 (usado como benchmark) ⁴⁰ . A revisão salienta a dificuldade de comparar estudos devido à diversidade de abordagens e métricas, e sugere diretrizes para aumentar a **generalização** – por exemplo, sempre testar modelos treinados em dados novos de outro centro populacional ⁴⁸ . Em suma, os autores veem potencial nos biomarcadores vocais de depressão, porém o **desempenho atual não supera métodos tradicionais**, indicando necessidade de melhorias e padronização.

Pan et al. (2023) – Estudo Comparativo Multigrupos (Front. Psychiatry) ⁴⁹ ¹⁸ .

Investigou se características vocais conseguem distinguir depressão de outros transtornos e controles. Incluiu 106 pacientes depressão, 32 bipolar, 20 esquizofrenia e 114 controles. Extraiu **i-vetores** de MFCC (técnica de representação da fala) e treinou regressão logística para diversas tarefas de classificação (par a par). Resultados: modelos distinguiram bem **qualquer transtorno vs. controle** (AUC ~0,90) e **esquizofrenia vs. controle** (AUC ~0,92), assim como depressão vs. controles (AUC ~0,85) ¹⁸ . Porém, ao tentar separar **depressão vs. bipolar**, o desempenho foi nulo (AUC = 0,50, equivalente ao acaso) ¹⁸ . Também foi difícil diferenciar bipolar vs. esquizofrenia (AUC ~0,75). Conclui que **traços vocais têm valor discriminativo** – confirmando poder para identificar depressão e esquizofrenia frente a saudáveis – mas **não são totalmente específicos a cada diagnóstico** (especialmente depressão vs. mania). Os autores sugerem aprofundar a investigação dos mecanismos pelos quais voz reflete sintomas em cada transtorno e desenvolver modelos de ML mais sofisticados ou multi-modais para melhorar a distinção.

Ji et al. (2024) –
*Detecção de Humor
Bipolar por Voz*
(Front. Neurology)

50 15 .

Estudo visando identificar **episódios depressivos vs. maníacos** em pacientes com transtorno bipolar através da voz. Coletou **diários de voz reais** dos pacientes relatando seu humor. Extraiu 22 atributos acústicos; 21 deles apresentaram diferenças significativas entre estados de humor. Testou quatro modelos (incluindo **GRU pré-treinado** em fala chinesa, GRU simples, BiLSTM e LDA) usando validação leave-one-out. Melhor desempenho: modelo GRU pré-treinado, com **sensibilidade de 77,5% para detectar estado depressivo e 54,8% para mania**, especificidades de 86,1% (depressão) e 90,3% (mania) ¹⁵ . A acurácia geral foi 80,2% ¹⁵ . Discussão: a voz pode **diferenciar de forma confiável** fases de humor distintas, oferecendo abordagem mais objetiva para acompanhar transtorno bipolar ⁵¹ . No entanto, o menor acerto para mania sugere necessidade de mais dados ou features para capturar plenamente as particularidades da fala maníaca.

Watts et al. (2023) –
*Cohort Study “Mental
Fitness” (medRxiv)*

33 34 .

Avaliou um biomarcador vocal longitudinal em 104 pacientes psiquiátricos ambulatoriais diversos. Cada participante gravou segmentos breves de voz (~30s) várias vezes por semana via aplicativo móvel, que gerava um escore denominado **“Mental Fitness Vocal Biomarker (MFVB)”**. O escore combinava 8 *features* vocais previamente associadas a saúde mental. Foi utilizada como referência a escala validada M3 (que rastreia depressão, ansiedade, TEPT e bipolar) preenchida pelos pacientes. **Achados:** o escore vocal correlacionou-se com a gravidade global dos sintomas; pacientes com MFVB elevado apresentaram risco significativamente maior de sintomas moderados/graves na avaliação M3. Um único áudio previu risco relativo ~1,5 vezes maior de sintomatologia elevada ⁵² . Já agregados semanais de voz elevaram o RR para ~2,0 ³⁴ . Notavelmente, entre participantes altamente engajados (5-6 gravações/semana, 38% da amostra), um MFVB consistentemente alto se associou a um **risco 8,5 vezes maior** de sintomas intensos ³⁴ . Isso demonstra que o *monitoramento frequente amplifica a acurácia preditiva*. Feedback dos usuários indicou boa usabilidade e benefícios percebidos, reforçando o potencial do método como adjunto escalável, de baixo custo e preservação de privacidade no acompanhamento terapêutico ³⁵ .

Mazur et al. (2025) –
*Estudo de Triagem
com Kintsugi Voice*
(Ann. Fam. Med.) ²⁶

53 .

Pesquisa transversal avaliando a ferramenta de biomarcador vocal da startup **Kintsugi** para detectar depressão em contexto de cuidados primários. Recrutou 14.898 adultos nos EUA e Canadá via mídia social; cada participante gravou ≥ 25 segundos de fala livre (inglês) pelo app e respondeu ao PHQ-9. Definiu-se PHQ-9 ≥ 10 como depressão moderada a severa. **Resultados:** a análise de voz por IA identificou padrões compatíveis com depressão significativa com **71,3% de sensibilidade e 73,5% de especificidade** ²⁶ . Em outras palavras, a voz conseguiu sinalizar corretamente cerca de 71% dos casos com depressão moderada/alta (comparado ao PHQ) e tranquilizar 73% dos sem depressão – desempenho que os autores julgam promissor para ampliar a triagem na atenção primária. Concluem que a tecnologia de voz tem *potencial utilidade clínica* para ajudar a rastrear depressão em larga escala ⁵³ , mas recomendam mais estudos sobre efetividade em cenários reais e integração no fluxo de trabalho antes de adoção rotineira.

(Observação: PHQ-9 = Patient Health Questionnaire-9, questionário padrão de depressão; AUC = área sob a curva ROC; MFCC = Mel-frequency cepstral coefficients; RR = risco relativo.)

Tecnologias e Iniciativas de Destaque (Startups e Projetos)

O fervilhante campo de biomarcadores vocais aplicados à saúde mental tem atraído diversas empresas de tecnologia e projetos de pesquisa translacional. Abaixo estão alguns exemplos notáveis de *startups* e soluções tecnológicas, ilustrando o cenário atual (até 2025) e suas abordagens:

Tecnologia / Projeto	Descrição e Aplicação
Ellipsis Health (EUA)	Plataforma de biomarcadores vocais para saúde comportamental. Utiliza ~90 segundos de fala livre do usuário (pode ser uma resposta à pergunta “como você está se sentindo?”) para gerar escores de depressão e ansiedade validados clinicamente. A empresa destaca combinar algoritmos acústicos (tom, pausas, frequência vocal) com algoritmos semânticos (análise de palavras) treinados em grande base de dados ²³ . Estudos internos indicam forte concordância dos “escores de vitalidade mental” de Ellipsis com avaliações de clínicos ⁵⁴ . A solução tem sido pilotada em sistemas de saúde e seguradoras – por exemplo, a seguradora Cigna incorporou a tecnologia em 2021 para oferecer o teste “StressWaves” aos clientes, visualizando níveis de estresse pela voz e recomendando exercícios de bem-estar ²⁹ . Ellipsis captou financiamento significativo (US\$26 milhões em rodada Série A) e colabora com instituições acadêmicas para validação. Até 2025, é usada como ferramenta de suporte (não diagnóstica) e busca aprovação regulatória para medir severidade de depressão/ansiedade como dispositivo médico ⁵⁵ .
Kintsugi (EUA)	Startup focada em IA de voz para saúde mental. Desenvolveu o Kintsugi Voice – tecnologia que analisa trechos curtos (~20 segundos) de fala espontânea e identifica <i>sinais sutis de depressão e ansiedade</i> . Destaca-se por ser idioma-independente : o algoritmo extrai padrões acústicos universais, permitindo análise em qualquer língua ³⁶ . O núcleo do produto é oferecido como API chamada <i>KiVA</i> (Kintsugi Vocal Analysis), para integração em plataformas de telessaúde, call centers, etc. Segundo a empresa, seu modelo identifica depressão com acurácia acima de 80% (“acerta 4 a cada 5 casos”) ³⁶ e é capaz de sinalizar <i>indícios precoces</i> que podem surgir até 10 anos antes de métodos tradicionais detectarem ⁵⁶ . A Kintsugi originou-se de dados coletados via um app de terapia de conversa, com mais de 100 mil downloads, construindo um amplo banco de vozes diversas ⁵⁷ . Em 2022, levantou US\$20 milhões para escalar a solução ⁵⁸ ⁵⁹ , e em 2023 seu algoritmo foi avaliado positivamente em um estudo científico independente (Mazur et al., 2025 – vide tabela acima). A Kintsugi posiciona sua tecnologia como ferramenta para triagem em clínicas (detecção de sofrimento mental oculto) e monitoramento passivo em contextos como suporte ao paciente crônico ou linhas de prevenção ao suicídio.

Tecnologia / Projeto	Descrição e Aplicação
Sonde Health (EUA)	<p>Plataforma líder em biomarcadores vocais <i>multicondicionais</i>. Diferentemente de startups focadas só em saúde mental, a Sonde aborda vários domínios de saúde via voz – do bem-estar mental a doenças respiratórias. A empresa lançou o aplicativo Sonde Mental Fitness, que pede ao usuário que fale por 30 segundos (respondendo perguntas simples) e então analisa o áudio para sinais de depressão, ansiedade, estresse e fadiga ²⁷. O usuário recebe um <i>score</i> diário de “aptidão mental” e sugestões se o score indicar risco ⁶⁰. Paralelamente, a Sonde oferece APIs para parceiros incorporarem a análise vocal em seus produtos (por exemplo, planos de saúde podem usar para acompanhamento de membros). A tecnologia da Sonde se baseia em áudio DSP e ML, e a empresa alega possuir um dos maiores bancos de voz proprietários – <i>mais de 1 milhão de amostras de voz de ~100 mil indivíduos</i> – abrangendo múltiplos idiomas ⁶¹. Isso teria permitido criar modelos robustos e <i>gerais</i>, inclusive para condições como asma, COVID-19 (a Sonde lançou um teste vocal para risco de COVID durante a pandemia) e agora saúde mental. A Sonde tem parcerias anunciadas com fabricantes de dispositivos (para integrar voice biomarker em wearables) e estava conduzindo ensaios clínicos em 2021-2022 para validar detecção de depressão e até sinais precoces de Parkinson pela voz ⁶². É apontada como uma das únicas empresas com abordagem multi-condição e com planos de buscar aprovação regulatória para triagem de saúde ampla via voz.</p>
Winterlight Labs (Canadá)	<p>Empresa derivada de pesquisa acadêmica, conhecida inicialmente por desenvolver biomarcadores de fala para doenças neurodegenerativas (como Alzheimer). Seu software analisa gravações de pacientes descrevendo figuras ou narrando histórias, extraindo características linguísticas (coerência, complexidade lexical) e acústicas. Embora focada em demência, a Winterlight também explorou aplicação em <i>psiquiatria</i>, por exemplo, na detecção de <i>sintomas negativos</i> da esquizofrenia pela falta de variação emocional na voz. A startup ganhou atenção com um estudo publicado em 2017 mostrando alta acurácia na detecção de Alzheimer a partir da fala, mas enfrentou um desafio de generalização linguística: descobriu-se que seu modelo inicial funcionava bem apenas para falantes de um certo sotaque inglês do Canadá, falhando com outros sotaques e idiomas ⁴¹. Esse revés ilustra a importância de bases diversas. Desde 2019, a Winterlight se concentrou em colaborações estratégicas (por ex., com a Johnson & Johnson) para desenvolver marcadores de Alzheimer e Parkinson, e está conduzindo ensaios clínicos com parceiros como Genentech e Pear Therapeutics ⁶². No campo psiquiátrico, seus avanços são menos divulgados, mas a empresa continua a ser referência em análise de fala com fins diagnósticos. O caso Winterlight serve de alerta sobre vieses, ao mesmo tempo em que seus progressos em validação clínica inspiram outras startups a seguir rigor científico semelhante.</p>

Tecnologia / Projeto	Descrição e Aplicação
Vocalis Health (Israel)	<p>Nascida da fusão das pioneiras Beyond Verbal (emocion AI) e Healthymize, a Vocalis inicialmente focou em <i>biomarcadores vocais de bem-estar e carga emocional</i>. A Beyond Verbal já nos anos 2010 afirmava detectar emoções como tristeza ou vigor pela voz, e chegou a sugerir correlações com depressão. Após a fusão em 2019, a Vocalis explorou uso da voz para detecção de <i>doenças e condições variadas</i>. Com a chegada da pandemia, a empresa pivotou seu foco para COVID-19, desenvolvendo um teste vocal para indicar infecção pelo coronavírus em parceria com a Clínica Mayo ⁶³. Esse redirecionamento ilustra como a tecnologia de biomarcadores vocais é versátil, encontrando aplicações conforme as demandas (no caso, triagem de COVID remotamente via voz). Embora isso tenha reduzido a ênfase em psiquiatria, a Vocalis acumulou experiência e patentes em análise de estresse e sobrecarga mental, que podem futuramente se traduzir em produtos voltados à saúde mental novamente. A trajetória da Vocalis também evidencia o interesse de grandes instituições médicas no tema – e.g., a Clínica Mayo engajada em pesquisas de voz – o que indiretamente beneficia o avanço de biomarcadores vocais psiquiátricos por expansão do conhecimento e aceitação da ideia de “diagnóstico pela voz”.</p>

(Nota: a ausência de menção de determinada empresa não diminui sua importância; o ecossistema inclui outras iniciativas como: Cogito Corp (EUA), que analisou ligações telefônicas de veteranos para sinais de depressão e TEPT; AudioSense (UE), com foco em monitorar esquizofrenia; Tapping into Collective Voice (Reino Unido), etc. A lista visa apenas exemplificar a diversidade de abordagens.)

Perspectivas Futuras e Considerações Finais

Os progressos recentes sugerem que **biomarcadores vocais têm um futuro promissor na psiquiatria**, mas a jornada até a adoção clínica ampla depende de vários fatores. Olhando adiante para 2025 e além, algumas perspectivas e tendências se destacam:

- **Validação Clínica e Aprovação Regulatório:** Espera-se que nos próximos anos surjam resultados de estudos clínicos maiores testando essas ferramentas em cenários do mundo real – por exemplo, ensaios controlados avaliando se usar um biomarcador vocal no cuidado de depressão melhora taxas de detecção e resultados do paciente. Evidências positivas abririam caminho para aprovações oficiais (FDA, CE, etc.). Visto que pelo menos uma empresa (Ellipsis) já está seguindo diretrizes do FDA para aprovação de IA médica ⁵⁵, é plausível que até o fim da década tenhamos **os primeiros biomarcadores vocais aprovados** como dispositivos auxiliares, possivelmente inicialmente para monitoramento de gravidade (e não diagnóstico primário). Isso seria um marco para credibilidade e adoção.
- **Integração em Prontuários e Fluxos de Telemedicina:** No futuro, ferramentas de análise vocal podem se tornar parte integrada das plataformas de atendimento. Por exemplo, um paciente em teleconsulta pelo computador ou telefone pode, automaticamente, ter sua fala analisada em segundo plano, com o software alertando o médico de sinais relevantes (“indicadores vocais sugerem ansiedade alta hoje”). Em consultórios físicos, talvez o paciente responda a um curto questionário falado a um *kiosk* antes da consulta, e o resultado já conste no prontuário eletrônico. A integração fluida – garantindo que a informação vocal agregue valor sem sobrecarregar – será fundamental. Grandes players de tecnologia em saúde (Cerner, Epic, etc.)

possivelmente vão incorporar APIs de startups de biomarcadores vocais assim que houver produtos aprovados, o que aceleraria difusão.

- **Foco em Sintomas e Biomarcadores Granulares:** Como discutido, há um movimento conceitual no campo sugerindo mudar a ênfase de “diagnosticar transtorno X ou Y” para *quantificar sintomas específicos* via voz ²⁸. Essa mudança se alinha com a prática psiquiátrica centrada em sintomas (p.ex., medir severidade da anedonia, do humor deprimido, da agitação psicomotora). Espera-se que futuros algoritmos sejam treinados para **estimar escalas clínicas** (PHQ-9, HAM-D, YMRS, PANSS etc.) ou até domínios específicos (como o grau de *alogia* em esquizofrenia, ou nível de *ansiedade social* durante fala). Isso tornaria os biomarcadores mais úteis para acompanhamento de tratamento – por exemplo, se um antidepressivo está funcionando, o sistema poderia mostrar redução gradual no “índice vocal de depressão” do paciente semana a semana, correlacionado com a melhora clínica. Essa abordagem orientada a sintomas também contorna em parte a polêmica de tentar “diagnosticar por IA” (o que alguns clínicos veem com ceticismo); em vez disso, oferece um *termômetro objetivo de manifestações clínicas*.
- **Personalização e Aprendizado Contínuo:** No futuro, os sistemas poderão aprender o *padrão de voz basal de cada indivíduo* e detectar mudanças relativas a si mesmo. Em vez de depender apenas de diferenças absolutas em relação a um modelo populacional, o algoritmo pode criar um perfil vocal do paciente em estado saudável e então notar desvios quando aquele indivíduo em particular começa a falar muito mais devagar ou de forma mais tensa que seu normal. Essa personalização aumentaria a sensibilidade para mudança intraindividual (útil em monitoramento de recaídas) e reduziria falsos alarmes causados por diferenças interindividuais. Tecnicamente, isso exigirá IA capaz de treinar modelos específicos por usuário e atualizar continuamente conforme novos dados de voz chegam – algo factível com aprendizado federado e on-device, preservando privacidade.
- **Expansão para Múltiplos Idiomas e Contextos Culturais:** Um futuro desejável é que biomarcadores vocais funcionem não apenas em inglês ou chinês (idiomas mais presentes nos estudos até agora), mas em português, espanhol, árabe, diversas línguas – tornando a ferramenta global. Para isso, iniciativas colaborativas de coleta de voz multicultural serão cruciais. Alguns projetos já começam a incluir dados de países diferentes (por ex., há corpora de depressão em espanhol, em hindi etc. emergindo). Empresas que almejam mercado global, como Kintsugi, enfatizam que seus modelos são treinados para serem *agnósticos de idioma*, baseando-se apenas em acústica ³⁶, mas isso ainda precisa ser provado em prática ampla. Acredita-se que frameworks de aprendizagem auto-supervisionada (que aprendem representações universais de fala) e *transfer learning* ajudarão a alcançar modelos multilíngues robustos.
- **Combinação com outros Digital Biomarkers:** A voz é uma peça do quebra-cabeça do “fenótipo digital” em saúde mental. Futuramente, a maior eficácia pode vir de **combinar múltiplos sinais**: voz + expressão facial (análise de microexpressões em videochamadas), voz + dados de sono e atividade (de wearables), voz + texto de mídias sociais (análise de sentimento), etc. Por exemplo, para detectar risco de depressão, um sistema integrado poderia analisar simultaneamente a fala, o padrão de uso do smartphone, a qualidade do sono e o nível de atividade física – fornecendo uma avaliação muito mais precisa do estado mental do que qualquer marcador isolado. Portanto, os biomarcadores vocais provavelmente farão parte de **plataformas de monitoramento multimodal**, e pesquisas nessa direção (integrando dados) devem crescer. Entretanto, a voz muitas vezes se destaca por si só, pois captura tanto a esfera biológica (via

acústica) quanto a psicológica (via conteúdo da fala), então continuará sendo um componente central.

- **Superação de Ceticismo e Educação:** No campo clínico, a adoção plena também depende de vencer o ceticismo e mostrar utilidade prática. Com mais publicações demonstrando casos de uso bem-sucedidos – por exemplo, um estudo clínico mostrando que um app de voz ajudou a reduzir hospitalizações de pacientes bipolares ao alertar sobre mania iminente – os profissionais ganharão confiança. Organizações profissionais poderão eventualmente emitir diretrizes sobre uso de ferramentas de voz (por ex., recomendando triagem vocal anual em população de risco). A educação de novos psiquiatras e psicólogos provavelmente incluirá familiarização com tecnologia digital, incluindo biomarcadores vocais, para que a próxima geração já encare isso como parte do arsenal terapêutico.

Em conclusão, a evolução recente das tecnologias de biomarcadores de voz aplicadas à psiquiatria tem sido marcante – **dos primeiros estudos exploratórios a sistemas de IA capazes de detectar depressão, bipolaridade, esquizofrenia e TEPT através da fala com precisão considerável**. Os avanços em *machine learning* e processamento de sinal habilitaram a identificação de padrões vocais sutis ligados a sintomas psiquiátricos, e aplicações práticas começam a despontar em triagem e monitoramento clínico. **As vantagens de uma ferramenta não-invasiva, escalável e objetiva são claras**, mas **igualmente claros são os desafios técnicos, clínicos e éticos** que precisam ser enfrentados antes que esses biomarcadores se tornem rotina. A tendência é que a pesquisa prossiga em ritmo acelerado – impulsionada pelo interesse global em saúde mental digital – e que vejamos refinamentos importantes nos próximos 2 a 5 anos em termos de acurácia, abrangência populacional e validação. Se bem-sucedidos esses esforços, os biomarcadores vocais poderão se consolidar como um precioso aliado no combate ao sofrimento psíquico: ajudando a *identificar mais cedo quem precisa de ajuda, orientar intervenções personalizadas e acompanhar a recuperação* – tudo isso usando apenas a forma como falamos como janela para a mente. Em suma, a voz pode vir a ocupar um lugar de destaque como **ferramenta de auxílio clínico na psiquiatria do futuro**, tornando o cuidado em saúde mental mais proativo, acessível e data-driven, sem perder de vista a necessidade de humanidade e contexto na interpretação de qualquer dado que venha a fornecer ⁶⁴.

1 4 5 6 28 44 Why Voice Biomarkers of Psychiatric Disorders Are Not Used in Clinical Practice? Deconstructing the Myth of the Need for Objective Diagnosis
<https://hal.science/hal-04679802v1/file/2024.lrec-main.1531.pdf>

2 3 9 Detection of Major Depressive Disorder Using Vocal Acoustic Analysis and Machine Learning
<https://www.medrxiv.org/content/10.1101/2020.06.23.20138651v1.full.pdf>

7 8 45 46 Speech and Voice Quality as Digital Biomarkers in Depression: A Systematic Review - PubMed
<https://pubmed.ncbi.nlm.nih.gov/40410060/>

10 24 25 26 53 Evaluation of an AI-Based Voice Biomarker Tool to Detect Signals Consistent With Moderate to Severe Depression
<https://www.annfammed.org/content/annalsfm/23/1/60.full.pdf>

11 12 13 Relative importance of speech and voice features in the classification of schizophrenia and depression | Translational Psychiatry
https://www.nature.com/articles/s41398-023-02594-0?error=cookies_not_supported&code=743aed15-4e10-45ba-a89a-bdb3357df571

14 31 32 PRIORI - Longitudinal Voice Patterns in Bipolar Disorder | Prechter Program | Michigan Medicine | University of Michigan

<https://medicine.umich.edu/dept/prechter-program/priori-longitudinal-voice-patterns-bipolar-disorder>

15 21 50 51 Frontiers | Depressive and mania mood state detection through voice as a biomarker using machine learning

<https://www.frontiersin.org/journals/neurology/articles/10.3389/fneur.2024.1394210/full>

16 17 19 20 NYU School of Medicine Develops Tool that Diagnoses Post-Traumatic Stress Disorder by Voice Analysis | NYU Langone News

<https://nyulangone.org/news/nyu-school-medicine-develops-tool-diagnoses-post-traumatic-stress-disorder-voice-analysis>

18 49 Frontiers | Exploring the ability of vocal biomarkers in distinguishing depression from bipolar disorder, schizophrenia, and healthy controls

<https://www.frontiersin.org/journals/psychiatry/articles/10.3389/fpsyt.2023.1079448/full>

22 23 29 30 37 38 41 42 43 54 55 62 63 64 AI startups claim to detect depression from speech, but the jury's out on their accuracy | VentureBeat

<https://venturebeat.com/ai/ai-startups-claim-to-detect-depression-from-speech-but-the-jurys-out-on-their-accuracy/>

27 60 61 Sonde Health Debuts Mental Health Vocal Biomarker App and API - Voicebot.ai

<https://voicebot.ai/2021/10/21/sonde-health-debuts-mental-health-vocal-biomarker-app-and-api/>

33 34 35 52 Validating the efficacy and value proposition of Mental Fitness Vocal Biomarkers in a psychiatric population: prospective cohort study | medRxiv

<https://www.medrxiv.org/content/10.1101/2023.11.21.23298774v1.full>

36 56 57 58 59 Mental Health Vocal Biomarker Startup Kintsugi Raises \$20M - Voicebot.ai

<https://voicebot.ai/2022/02/28/mental-health-vocal-biomarker-startup-kintsugi-raises-20m/>

39 40 47 48 A Review of Studies Using Machine Learning to Detect Voice Biomarkers for Depression | Journal of Technology in Behavioral Science

<https://link.springer.com/article/10.1007/s41347-024-00454-2>

Leia mais em [metadoctors.com](https://www.metadoctors.com)!

Entre em contato com a equipe metaDoctors no [LinkedIn](#).

Siga o metaDoctors no [Facebook](#) ou [Instagram](#).

Inscreva-se no canal metaDoctors do [YouTube](#) para ter acesso a todos os vídeos sobre tendências, tecnologias e dispositivos que moldarão o futuro da medicina.



Uma Comunidade Médica que oferece Informação,
Tecnologia, Cursos e Benefícios.