

O emprego da inteligência artificial/machine learning na detecção de depressão por meio de voz: uma mini revisão integrativa

Anna Carolina Martins Bandeira¹, Ana Beatriz Alencar Agostinho¹, Isadora Morais Dias¹, Mileny Santos Pereira¹, Sarah Ingrid Verissimo da Silva Matos Dutra¹, George Martins Ney da Silva Junior²

1. Discente do curso de Medicina do Centro Universitário UniEVANGÉLICA.

2. Docente curso de Medicina do Centro Universitário UniEVANGÉLICA.

RESUMO: A depressão é uma condição psiquiátrica que afeta pessoas de diferentes faixas etárias. Métodos tradicionais de diagnóstico apresentam limitações. O uso de inteligência artificial (IA) tem sido promissor para auxiliar nos diagnósticos de desordens psiquiátricas. Esta mini revisão analisa estudos recentes, dos últimos 5 anos, que empregaram ferramentas de IA para a detecção de depressão por meio de recursos de voz. Analisar o emprego de inteligência artificial/machine learning como método auxiliar no diagnóstico de depressão por meio de recursos de voz. Trata-se de uma mini revisão integrativa que obteve dados nas plataformas: PubMed, MEDLINE e SciELO, utilizando os seguintes descritores: “Aprendizado de Máquina”, “Depressão”, “Detecção” e “Voz”. Foram incluídos estudos gratuitos publicados nos últimos cinco anos, na língua portuguesa e/ou inglesa, com relevância para a pesquisa e publicados em periódicos indexados. Literaturas destoantes da temática abordada foram excluídas. Deve-se notar que a inteligência artificial, mais precisamente o Machine Learning, é benéfico para a detecção de depressão por meio de recursos de voz quando comparados os grupos não deprimidos e deprimidos. A taxa de sucesso do uso desse mecanismo foi acima de 95% para qualquer um dos modelos de análise de voz por inteligência artificial, comprovando, assim, que essa ferramenta pode ser um grande aliado na detecção de depressão. Houve unanimidade no tocante à necessidade de novas pesquisas nesta área, que além de inovadora pode auxiliar o profissional médico no diagnóstico precoce de depressão. Ao final das pesquisas e após análise dos resultados obtidos os autores constatarem que a inteligência artificial (IA) pode ser uma grande aliada na detecção e diagnóstico da depressão por meio da voz em diferentes faixas etárias, porém ressaltam que apesar de seu caráter inovador ainda é uma tecnologia recente, sendo necessárias mais pesquisas com outros métodos de captação de recursos de voz e com amostragens maiores para que o uso da inteligência artificial possa ser considerado uma possibilidade para a detecção de depressão em âmbito mundial.

Palavras-chave:

Aprendizado de máquina. Depressão. Detecção. Voz.

INTRODUÇÃO

A depressão é uma condição psiquiátrica que afeta milhões de pessoas de diferentes faixas etárias em todo o mundo, de acordo com dados da Organização Mundial da Saúde, sendo esperado um aumento nesse número no futuro¹. Essa desordem acaba por impactar a economia global em cerca de 1

trilhão de dólares anualmente, justamente por acometer milhões de pessoas ao redor do planeta². O agravamento dos sintomas depressivos sem um diagnóstico pode causar danos aos indivíduos, desde a alterações no estilo de vida até mesmo levar ao autoextermínio, fator que tem surgido mais frequentemente em grupos mais jovens³. Quando diagnosticada corretamente, a depressão é uma doença tratável e seus sintomas podem ser aliviados⁴.

Atualmente o diagnóstico de depressão clínica é baseado principalmente no DSM-V e na Classificação Internacional de Doenças e Problemas Relacionados à Saúde, 10ª edição (CID-10), porém foi notado que esses manuais possuem limitações para o diagnóstico como o fator de que ambas as diretrizes se baseiam no número de sintomas e na duração destes, resultando em uma visão não dimensional da depressão⁵. Ademais, a depressão não possui biomarcadores específicos, além de ser uma condição clinicamente heterogênea, assim, um dos desafios é diagnosticar a depressão em tempo hábil⁴. Nesse cenário, a fim de superar as limitações impostas pelos manuais citados para o diagnóstico da depressão, os autores analisados conduziram pesquisas utilizando a Inteligência Artificial (IA) para a detecção de depressão por meio da captação de recursos de voz.

Alguns dos estudos analisados fizeram uso de tarefas de fala gravadas para extrair recursos vocais, enquanto outros utilizaram entrevistas para captar esses recursos¹⁻⁴⁻⁵. Os recursos de voz eram captados durante tarefas de fala e entrevistas e as gravações eram submetidas a programas de computador baseados em Inteligência Artificial, os quais utilizaram principalmente a curva ROC e outros parâmetros como regressão logística (RL) e multicamadas perceptron (MP)²⁻⁴⁻⁵. E também houve o uso de aplicativos ou modelos criados para serem testados e comparados com os modelos citados¹⁻³.

Uma vez que os métodos tradicionais disponíveis apresentam limitações, os autores encontraram no uso de IA, sob a ótica do profissional médico, um auxílio para o diagnóstico da depressão e, sob a ótica do paciente, uma oportunidade de acesso a um tratamento mais preciso. Os recursos de voz podem ser uma ferramenta útil para auxiliar o diagnóstico de depressão, isso se deve ao fato de que pacientes depressivos apresentam variações no tom de voz, no ritmo da fala, pronúncia das palavras, dentre outras características¹⁻⁵. Diante disso, o objetivo desta mini revisão integrativa é analisar o emprego de inteligência artificial/machine learning como método auxiliar no diagnóstico de depressão por meio de recursos de voz.

METODOLOGIA

Este estudo é uma mini revisão bibliográfica integrativa que pretende responder à seguinte pergunta: “Como o emprego de inteligência artificial/machine learning pode ser benéfico para detecção de depressão por meio de recursos de voz?”. Foram pesquisados artigos nas bases de dados PubMed, MEDLINE e SciELO (Scientific Electronic Library Online). Foram utilizados os seguintes descritores: “depressão”, “detecção”, “voz”, “aprendizado de máquina” acrescido do booleano AND.

Em primeira análise, foram encontrados dez artigos por meio da pesquisa com utilização dos descritores mencionados. Posteriormente, foram utilizados os seguintes critérios de inclusão: estudos publicados nos últimos cinco anos, na língua portuguesa e/ou inglesa, com relevância para a pergunta norteadora da pesquisa, com acessibilidade ao texto completo gratuitamente, apenas trabalhos publicados em periódicos indexados. Os critérios de exclusão foram: artigos de revisão, dissertações de mestrado e tese de doutorado. Dessa forma, foram selecionados para a mini revisão de literatura 5 artigos.

RESULTADOS

Nesta mini revisão integrativa, será descrita uma análise dos resultados apresentados pelos cinco artigos selecionados, além de apresentar um panorama geral por meio do Quadro 1. De uma forma geral, deve-se notar que a inteligência artificial, mais precisamente o Machine Learning, foi benéfica para a detecção de depressão por meio de recursos de voz, sendo capaz de discriminar os grupos de não deprimidos dos grupos de deprimidos.

De acordo com CANSEL, a taxa de sucesso do modelo desenvolvido na distinção dos diferentes transtornos foi de 96,943%. Alguns modelos de análise utilizados tiveram melhor desempenho para diagnosticar transtorno bipolar, mas desempenho pouco eficaz para detectar depressão (modelo kNN), outros modelos mostraram alto desempenho para detectar ansiedade e psicose, mas tiveram desempenho baixo na detecção de transtorno bipolar (modelo SVM). Os achados apoiam a utilização da análise de fala para distinguir os principais transtornos psiquiátricos por terem potencial de melhorar o processo de diagnóstico¹.

Segundo LIN, a taxa de conclusão do protocolo foi de 61% para ambos os grupos. O uso além do protocolo (falar mais do que 5 minutos) foi de 27% para o grupo positivo e 9% para o grupo negativo. O aplicativo Ellipsis Health mostrou uma AUC (estimativa de acurácia) de 0,82 para os grupos combinados quando comparado ao PHQ-8 e GAD-7 com um ponto de corte de 10. O desempenho foi alto para participantes seniores, bem como para as faixas etárias mais jovens².

Conforme ZHANG, que propuseram um modelo de fusão multimodal para detecção de depressão em adolescentes, seus resultados apresentaram melhor desempenho com o modelo proposto pelos autores quando comparados com modelos unimodais baseados em HSFs, baseados em espectrogramas, baseados em HSFs e espectrogramas, baseados em incorporação de palavras e baseados em incorporação de frases. Os autores salientaram que os recursos de voz podem ser utilizados para detecção de depressão em adolescentes porque é um dos recursos de áudio que mais carregam informações, quando comparados com texto³.

Já KIM, considerou que um único modelo demonstrou melhor desempenho na detecção de depressão por voz baseadas na língua coreana, o modelo CNN. A maior precisão alcançada com esse

modelo foi de 78,14%. Resultados desse estudo mostram que as características acústicas aprendidas em profundidade (em inglês: *deep learning*) levam a um melhor desempenho do que aqueles obtidos usando a abordagem convencional e modelos pré-treinados⁴.

Por fim, SHIN, apresentou que o estudo revelou ainda mais diferenças na voz em episódios depressivos e confirmou que grupos não deprimidos e participantes com depressão menor e maior poderiam ser distinguidos com precisão através do aprendizado de máquina⁵.

Quadro 1: Artigos incluídos na análise da mini revisão integrativa de literatura, separados por autor/ano, desenho de estudo, objetivos e conclusões em ordem crescente de publicação.

AUTOR/ANO	DESENHO DE ESTUDO	OBJETIVOS	CONCLUSÕES
SHIN <i>et al.</i> (2021)	Estudo transversal	Examinar se a voz pode ser utilizada como biomarcador para a detecção de transtorno de depressão maior e depressão menor.	O emprego do exame do tom de voz e das mudanças de tom revelou possibilidade de discriminação entre não deprimidos (ND) e deprimidos, porém, sem discriminar entre portadores de depressão maior (MDE) e portadores de depressão menor (mDE).
LIN <i>et al.</i> (2022)	Estudo transversal	Determinar a viabilidade de coletar amostras semanais de voz para a triagem de saúde mental e demonstrar portabilidade e melhor desempenho dos modelos de aprendizado de máquina da Ellipsis para pacientes de várias idades	O aplicativo Ellipsis Health demonstrou viabilidade no uso de gravações de voz para detectar depressão e ansiedade entre várias faixas etárias; os modelos de aprendizado de máquina usando a metodologia Transformer mantiveram o desempenho e melhoram a metodologia LSTM quando aplicados à população do estudo.
ZHANG <i>et al.</i> (2022)	Estudo transversal	Detectar a depressão em adolescentes sob o aspecto da fusão multimodal de dados, visando fornecer um método de triagem automatizado para detecção e intervenção precoce na depressão	Os resultados mostraram que a fusão de informações de áudio e texto são alternativas viáveis na detecção da depressão. Os resultados experimentais também comprovaram a eficácia do modelo e método de fusão proposto.
KIM <i>et al.</i> (2023)	Estudo transversal	Propor uma estrutura para detecção automática de depressão usando características acústicas em larga escala baseadas na língua coreana	A análise de dados de fala registrados durante a leitura de frases dependentes de texto pode ajudar a prever o status de depressão automaticamente, capturando as características da depressão, como a verificação do humor dos pacientes com transtorno depressivo maior por meio da detecção da consistência das descrições objetivas nas falas

			dos pacientes.
CANSEL <i>et al.</i> (2023)	Estudo transversal	Desenvolver um novo sistema de apoio à decisão clínica baseado em inteligência artificial, utilizando sinais de fala para distinguir entre transtornos depressivo, ansioso, espectro da esquizofrenia e espectro bipolar.	O método desenvolvido tem potencial de ajudar os psiquiatras a distinguirem seus pacientes de forma eficiente e confiável.

DISCUSSÃO

De acordo com os resultados apresentados, o uso de inteligência artificial, especialmente do machine learning, traz benefícios e pode ser uma ferramenta auxiliar para a detecção de depressão por meio da voz independente da faixa etária do indivíduo. De uma forma geral, os resultados dos estudos encontrados nessa revisão estão de acordo com a literatura, apesar de algumas discrepâncias.

Um exemplo de uso das gravações de fala no auxílio da detecção precoce e no monitoramento da depressão é abordado mediante uma análise longitudinal de aprendizagem automática personalizada acerca dos padrões e das atividades cotidianas realizadas pelos indivíduos, de modo a criar um modelo singular para cada paciente. As descobertas em questão vão ao encontro do estudo realizado sobre a implementação do aplicativo Ellipsis, o qual também utiliza um design centrado no paciente com o fito de melhorar o engajamento do usuário e a eficácia do diagnóstico. Os achados também corroboram com outro trabalho que afirma que um dos principais imbróglis dos modelos de detecção atuais é o uso de informações globais, levando, assim, a uma baixa precisão no diagnóstico da patologia estudada^{2,3,6}.

Dentre os estudos analisados, dois deles estavam em concordância por terem sido realizados no mesmo país e seguindo métodos semelhantes sendo que um deles criou uma ferramenta para detecção automática de depressão por voz usando a língua nativa juntamente com a verificação do humor dos participantes e a comparou com outros modelos já existentes, enquanto o outro trabalho usou dos modelos de machine learning já existentes para trazer novas conclusões sobre a detecção de depressão como a de que os participantes com sintomas depressivos apresentavam voz mais grave e maiores hesitações e pausas durante as entrevistas^{4,5}.

Porém, alguns estudos divergem no tocante a qual metodologia seria a melhor a ser utilizada para a coleta de dados por voz para a detecção da depressão mediante inteligência artificial. Um dos trabalhos aponta que o uso de tarefas cronometradas de três minutos não são um recurso efetivo para a detecção de depressão, uma vez que os indivíduos analisados já podem estar predispostos a estarem ansiosos ou com sintomas depressivos diante da participação na pesquisa, diferentemente da utilização

de entrevistas, as quais deixam os participantes mais livres para poderem se expressar verbalmente ou hesitarem durante um tempo mais longo, uma vez que as entrevistas tinham duração de trinta minutos⁵⁻⁸.

Ademais, outros dois trabalhos apresentam contrastes quanto ao uso da inteligência artificial na detecção de depressão. Enquanto um estudo utiliza da ferramenta tecnológica para a detecção precoce de sintomas psíquicos, quantificando a gravidade de forma longitudinal e coletando dados de voz de participantes sem relação com sintomas adquiridos, o outro trabalho tem como amostra participantes inseridos em uma clínica psiquiátrica, utilizando, assim, do recurso mencionado para determinar possíveis agravantes em diagnósticos já confirmados e conhecidos¹⁻⁶.

CONCLUSÃO

Ao final das pesquisas e após análise dos resultados obtidos os autores constatarem que a inteligência artificial (IA) pode ser uma grande aliada na detecção e diagnóstico da depressão por meio da voz em diferentes faixas etárias, porém ressaltam que apesar de seu caráter inovador ainda é uma tecnologia recente, sendo necessárias mais pesquisas com outros métodos de captação de recursos de voz e com amostragens maiores para que o uso da inteligência artificial possa ser considerado uma possibilidade para a detecção de depressão em âmbito mundial.

REFERÊNCIAS

¹CANSEL *et al.* A new artificial intelligence-based clinical decision support system for diagnosis of major psychiatric diseases based on voice analysis. **Psychiatria Danubina**. v. 35, n. 4, p. 489-499, 2023. DOI: <https://doi.org/10.24869/psyd.2023.489>. Disponível em: https://www.psychiatria-danubina.com/UserDocsImages/pdf/dnb_vol35_no4/dnb_vol35_no4_489.pdf. Acesso em: 2 de junho de 2024.

²LIN *et al.* Feasibility of a machine learning-based smartphone application in detecting depression and anxiety in a generally senior population. **Frontiers in Psychology**. v. 13, n. 811517, 2022. DOI: <https://doi.org/10.3389/fpsyg.2022.811517>. Disponível em: <https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2022.811517/full>. Acesso em: 2 de junho de 2024.

³ZHANG *et al.* Adolescent depression detection model based on multimodal data of interview audio and text. **International Journal of Neural Systems**. v. 32, n. 11, p. 225004, 2022. DOI: <https://doi.org/10.1142/S0129065722500459>. Disponível em: <https://www.worldscientific.com/doi/epdf/10.1142/S0129065722500459>. Acesso em: 2 de junho de 2024.

⁴KIM *et al.* Automatic depression detection using smartphone-based text-dependent speech signals: deep convolutional neural network approach. **Journal of Medical Internet Research**, v. 25, 2022. DOI: <https://doi.org/10.2196/34474>. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9909514/>. Acesso em: 2 de junho de 2024.

⁵SHIN *et al.* Detection of minor and major depression through voice as a biomarker using machine learning. **Journal of Clinical Medicine**. v. 10, n.14, p. 3046, 2021. DOI: <https://doi.org/10.3390/jcm10143046>. Disponível em: <https://www.mdpi.com/2077-0383/10/14/3046>. Acesso em: 2 de junho de 2024.

⁶DIAZ-RAMOS *et al.* Using wearable devices and speech data for personalized machine learning in early detection of mental disorders: Protocol for a participatory research study. **JMIR research protocols**. v. 12, 2023. DOI: <https://doi.org/10.2196/48210>. Disponível em: <https://www.researchprotocols.org/2023/1/e48210>. Acesso em: 2 de junho de 2024.

⁷McGINNIS *et al.* Giving voice to vulnerable child: machine learning analysis of speech detects anxiety and depression in early childhood. **IEEE Journal of Biomedical Health Informatics**. v.10, n.1109, 2019. DOI: <https://doi.org/10.1109/JBHI.2019.2913590>. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7484854/>. Acesso em: 2 de junho de 2024.

⁸HAQUE *et al.* Detection of child depression using machine learning methods. **Plos One**. v.16, n.12, 2021. DOI: <https://doi.org/10.1371/journal.pone.0261131>. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8675644/>. Acesso em: 2 de junho de 2024.