JDHBI - Journal of Digital Health and Biomedical Informatics, 2026, 1:1
doi: 10.5753/jdhbi.2026.6590
This work is licensed under a Creative Commons Attribution 4.0 International License.

Research Paper

Development of Predictive Risk Algorithms for Mental Health in Indoor Settings: a Cohort Study

Ana Laura de Sene Amâncio Zara [Universidade Federal de Goiás | *analaurazara@ufg.br*]
Mariana Maria Batista [Universidade Federal de Goiás | *marianamariabatista@egresso.ufg.br*]
Bianca Ribeiro Soares [Universidade Federal de Goiás | *soaresbianca@egresso.ufg.br*]
Myllena D'ávila Mariana Gomes Silva [Universidade Federal de Goiás | *myllena.mariana@discente.ufg.br*]
Joédan Silva Santos [Universidade Federal de Goiás | *joedansantos@egresso.ufg.br*]
Myrella Alves Bordado [Universidade Federal de Goiás | *myrellaalves@discente.ufg.br*]
Júlia Gomes Moreira [Universidade Federal de Goiás | *julia.moreira@egresso.ufg.br*]
Melyssa Mariana Gomes Silva [Universidade Federal de Goiás | *melyssamariana@discente.ufg.br*]
Antônio Emílio Pereira [Universidade Federal de Goiás | *antoniopereira@discente.ufg.br*]
Daniel Fazzioni [Universidade Federal de Goiás | *fazzioni@egresso.ufg.br*]
Murilo de Oliveira Guimarães [Universidade Federal de Goiás | *muriloguimaraes@egresso.ufg.br*]
Alex Victor Echeverria Ugarte Silva [Universidade Federal de Goiás | *ugarte_alex@egresso.ufg.br*]
Aldo André Díaz Salazar [Universidade Federal de Goiás | *aldo.diaz@ufg.br*]
Arthur Ricardo de Sousa Vitória [Universidade Federal de Goiás | *arthursvrr@gmail.com*]
Paula Geralda Barbosa Coelho [Universidade Federal de Goiás | *paulacoelho@ufg.br*]
Paulo Roberto D'Aurea Machado [Dynamic Mind Health | *paulo@dmhealth.com.br*]
Rodrigo da Silva Dias [Dynamic Mind Health | *rodrigo@dmhealth.com.br*]
Iwens Gervásio Sene Júnior [Universidade Federal de Goiás | *iwens@ufg.br*]
Renato de Freitas Bulcão-Neto [Universidade Federal de Goiás | *rbulcao@ufg.br*]

Instituto de Informática, Universidade Federal de Goiás, Alameda Palmeiras, s/n - Chácaras Califórnia, Goiânia-GO, 74690-900, Brazil.

Abstract. This article describes the methodology and characteristics of a cohort designed to support the development of intelligent predictive algorithms for anxiety and depression risk in indoor settings. This prospective cohort study included students and staff from a university. Symptoms of anxiety and depression were assessed using self-reported questionnaires (PROMIS®). Indoor environments were monitored using devices that measured air quality parameters, including humidity, temperature, and carbon dioxide levels. A total of 601 individuals were recruited, with a retention rate of 85.2%. The majority were male (55.3%), with a median age of 22 years. Dropouts and retained participants differed significantly in terms of sex, age, occupation, chronic illnesses, and general health status. Among participants, the prevalence of anxiety and depression was 54.7% and 41.8%, respectively. The observed prevalence rates of anxiety and depression exceed those typically reported in Brazilian academic settings, which, although concerning, provide a methodological advantage. Including participants with diverse clinical outcomes, particularly those with extreme scores, is crucial for enhancing the precision of predictive algorithms. Therefore, the ideal sample composition should prioritize clinical heterogeneity, reflecting the diversity of real-world healthcare scenarios and directly contributing to improved computational modeling accuracy. Differential sample attrition did not compromise the high prevalence of anxiety and depression, a key feature for developing predictive algorithms with the potential to accurately identify mental health risks in indoor settings.

Keywords: Mental Health, Machine Learning, Predictive Modeling, Artificial Intelligence, Longitudinal Studies

Received: 23 Mar 2026 • **Accepted:** 14 May 2026 • **Published:** 13 Nov 2026

1 Introduction

Mental health is defined by the World Health Organization (WHO) as a state of well-being in which individuals can utilize their own abilities, recover from routine stress, be productive, and contribute to their community (World Health Organization, 2025). Mental health disorders have a high global prevalence and result in over 7.2 trillion international dollars (Int$ 7.2 trillion) in annual economic losses (Arias *et al.*, 2022). Moreover, individuals affected by mental disorders are at increased risk for adverse health outcomes and premature mortality (GBD 2019 Mental Disorders Collaborators, 2022).

In 2019, approximately 12% of the global population was affected by some form of mental disorder, with depression and anxiety being the most prevalent. Due to the COVID-19 pandemic, the prevalence of depression increased by 27.6% and anxiety by 25.6% in 2020, impacting all age groups (Santomauro *et al.*, 2021). Brazil ranks among the countries with the highest prevalence of mental disorders (Institute for Health Metrics and Evaluation, 2021), with self-reported anxiety at 26.8% (95% CI: 24.3%–29.5%) and depression at 12.7% (95% CI: 11.4%–14.2%) (Hallal *et al.*, 2023).

One of the risk factors for mental disorders is poor indoor air quality and environmental conditions (Finell and Nätti, 2019, 2020). Extremely low or high indoor temperatures, inadequate ventilation, mold, humidity, outdoor pollution, chemical substances in construction materials, and bad cleaning practices contribute to poor air quality in indoor spaces (Wolkoff, 2018). Besides reducing comfort and increasing the risk of respiratory and allergic diseases, poor indoor air quality negatively affects mental states, impairs cognitive performance, and leads to workplace absenteeism or reduced productivity (Finell and Nätti, 2020; Wolkoff *et al.*, 2021).

The set of illnesses caused or exacerbated by indoor air pollution is referred to as Sick Building Syndrome (SBS). SBS is diagnosed when two or more symptoms occur at least twice a week while inside a building and subside upon leaving. These symptoms include headache, fatigue, lethargy, eye irritation, nasal and throat discomfort, skin problems, and difficulty concentrating (Joshi, 2008).

In 2010, the WHO issued guidelines promoting the creation and maintenance of environments that ensure thermal comfort, adequate ventilation, and support health and well-being (World Health Organization, 2010). Following the COVID-19 pandemic, indoor air quality has gained increased attention. To meet WHO recommendations, tools capable of assessing indoor environments are required. In this context, sensor-based devices, machine learning solutions, and mobile applications can enhance the ability to identify, diagnose, and monitor general and mental health conditions through digital technologies, enabling interactive and iterative integration of environmental data (Chojer *et al*., 2020; De Angel *et al*., 2022; Eisenstadt *et al*., 2021; Sheikh *et al*., 2021).

To support the development of machine learning algorithms capable of accurately predicting symptoms based on environmental data, epidemiological studies involving primary data collection on mental health, perceived air quality, and sensor-based environmental measurements are essential. This initiative aims to facilitate remote and timely healthcare support and clinical decision-making, representing a significant contribution to the field of health informatics.

Accordingly, the objective of the present study is to describe the methodological procedures for collecting primary data on the mental health of individuals who occupy indoor environments monitored by air quality sensors.

2 Method

2.1 Ethical Considerations

This is a prospective longitudinal study conducted across 41 indoor environments within a public university, between December 2022 and December 2023. The research is part of the project entitled "IoT and Intelligent Algorithms for Monitoring Air Quality in Indoor Settings", approved by the Research Ethics Committee (protocol no. 5.513.411) and conducted in accordance with Resolution No. 466, dated December 12, 2012, of the Brazilian National Health Council (Brazil Ministry of Health and National Health Council, 2012).

2.2 Sample

For cluster-based sampling, indoor environments with the highest concentration of staff, interns, and/or students were selected, provided they had access to electricity and Wi-Fi to enable installation of a hardware device equipped with sensors for air quality monitoring. This device was designed to continuously measure indoor temperature (°C), relative humidity (%), and carbon dioxide levels (parts per million – ppm). Eligible participants were individuals aged 18 years or older, of any sex, who remained in the monitored environments for at least two hours during any shift and voluntarily agreed to participate by signing the Informed Consent Form (ICF). Participant enrollment in the cohort occurred between May 15 and November 29, 2023.

2.3 Enrolment and Follow-up

Participants were invited in person to complete electronic questionnaires in two stages. In Stage 1 (Enrolment), they responded to a structured questionnaire designed to collect demographic, socioeconomic, and clinical data, as well as the Patient-Reported Outcomes Measurement Information System (PROMIS®) instrument (Pilkonis *et al*., 2011), validated in Brazil (Zumpano *et al*., 2017), for self-assessment of physical, mental, and social health.

In Stage 2 (Follow-up), after 30 minutes of exposure to the monitored environment, participants completed a questionnaire on perceived air quality (PQAr) (Silva *et al*., 2021) and two additional instruments for self-assessment of anxiety symptoms (PROMIS-Anxiety Short Form®) and depressive symptoms (PROMIS-Depression Short Form®) (HealthMeasures, 2023). Each participant could respond up to 15 times, once per day of presence in the monitored environment. Loss to follow-up was defined as failure to complete the Stage 2 questionnaire. Questionnaire links were sent via WhatsApp® or email using a messaging bot, which also delivered scheduled reminders based on the days and times participants reported frequenting the monitored space.

2.4 Data Collection Instruments and Variables

In Stage 1, sociodemographic and economic variables collected through the structured questionnaire included: date of birth, age, sex, education level, marital status, number of household members, race/ethnicity, professional activity, university role, average household income, and religion. Clinical variables included current pregnancy, smoking habits, presence of allergies and other chronic conditions, clinical diagnosis of mental disorders such as anxiety and/or depression, psychological and/or psychiatric follow-up, and continuous use of medication for anxiety and/or depression.

The PROMIS® Global Health questionnaire, version 1.2 (HealthMeasures, 2023), used in Stage 1, comprises 10 items assessing physical, social, and mental health. It includes two general items (Global01 and Global09) and two composite scales with four items each:

- Global Physical Health Scale, which includes physical health (Global03), physical functioning (Global06), pain intensity (Global07), and fatigue (Global08);
- Global Mental Health Scale, which includes quality of life (Global02), mental health (Global04), satisfaction with social activities and relationships (Global05), and emotional problems (Global10).

Responses were categorized using five-point Likert scales, except for pain intensity (Global07), which was rated on a scale from 0 to 10, where 0 indicates no pain and 10 indicates the worst imaginable pain. This item was recoded as follows: 0=5; 1–3=4; 4–6=3; 7–9=2; and 10=1. Items Global07, Global08, and Global10 are negatively worded and were reverse-coded so that higher scores reflect better overall health. Items Global01 (general health) and Global09 (social roles) were not included in the composite scale scores and were classified using raw scores: 5=excellent, 4=very good, 3=good, 2=fair, and 1=poor (HealthMeasures, 2023).

The sum of scale scores was converted into T-scores using the official conversion table (HealthMeasures, 2023), with higher scores indicating better health status. The Global Mental Health Scale was classified as: excellent (>56), very good (56–48), good (48–40), fair (40–29), and poor (<=29). The Global Physical Health Scale was classified as: excellent (>58), very good (58–50), good (50–42), fair (42–35), and poor (<=35) (HealthMeasures, 2023).

2.5 Perceived Air Quality and Symptom Assessment

In Stage 2, variables related to perceived air quality were obtained using a structured instrument developed by (Silva *et al*., 2021) which included: approximate time of entry into the monitored environment; assessment of indoor air quality; temperature and relative humidity, evaluated according to the thermal comfort scale Standard 55 (ANSI/ASHRAE, 2023); and odor or smell. In addition to sensory stimuli, the emergence of symptoms indicative of Sick Building Syndrome (SBS) was also assessed, including neurological, ophthalmological, respiratory, dermatological, and general clinical symptoms. These were classified on a five-point Likert scale: never (1), rarely (2), occasionally (3), frequently (4), and very frequently (5) (Joshi, 2008; Silva *et al*., 2021).

Anxiety and depression levels were assessed each time participants entered the monitored environment (Stage 2), using the PROMIS® Short Form-Anxiety and Short Form-Depression scales (HealthMeasures, 2023), each comprising eight items. The SF-Anxiety scale evaluates the intensity of feelings such as fear, anxious distress, hyperarousal, and somatic symptoms associated with arousal. The SF-Depression scale assesses the intensity of negative mood, self-perception, social cognition, and reductions in positive affect and engagement (HealthMeasures, 2023). Items were rated on a five-point Likert scale: never (1), rarely (2), sometimes (3), often (4), and always (5). Scores were calculated by summing the item values, multiplying by eight (total number of items), and dividing by the number of items answered. Higher scores indicate greater severity of anxiety or depression. Final scores were standardized into T-scores using the official conversion table (HealthMeasures, 2023) and classified as follows: normal (<55); mild (55.0–59.9); moderate (60.0–69.9); and severe (>=70.0). PROMIS® instruments are provided for non-commercial research use (HealthMeasures, 2023). All questionnaires were developed using the REDCap® platform (Harris *et al*., 2009).

2.6 Training and Calibration

During the pilot testing phase, conducted between December 13, 2022, and May 14, 2023, six interviewers were trained to use the REDCap® platform and to conduct in-person enrolment of participants. Virtual monitoring of participant adherence in both stages was performed using a real-time data dashboard (dashbot). The flowchart outlining the approach, enrolment, and follow-up procedures is presented in Figure 1. During this pilot phase, the questionnaires were refined within REDCap®, the bot was programmed for WhatsApp® and email communication, dashbot functionality was tested, mobile devices were installed in the selected environments, and an initial test was conducted with 30 participants

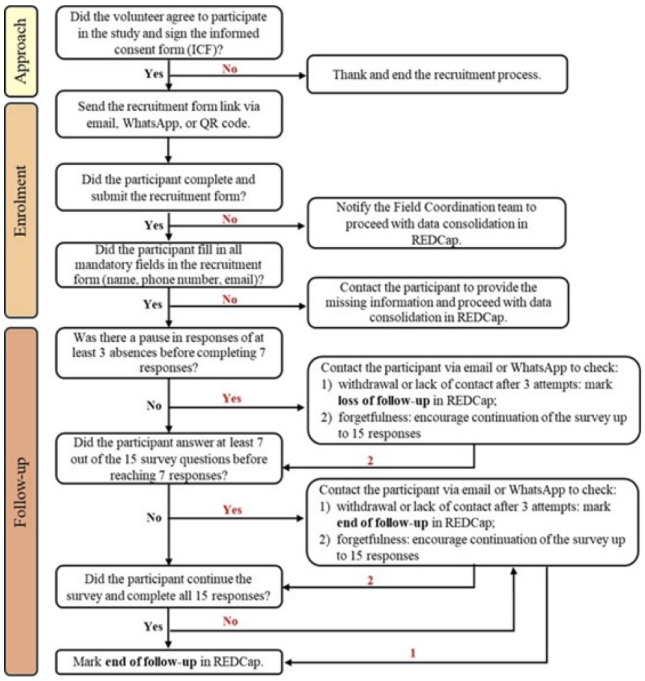


Figure 1. Flowchart of the participant approach, enrolment, and follow-up process in the study.

across two monitored environments. These individuals were not included in the final sample.

2.7 Data Analysis

Descriptive statistics were used to characterize the sample. Categorical variables are presented as absolute and relative frequencies, while continuous variables are reported as means and standard deviations or medians and interquartile ranges (IQR 25%–75%), depending on data distribution, which was assessed using the Shapiro-Wilk test (p>0.05).

To evaluate differential attrition, comparisons between participants and dropouts were conducted using the Chi-square test, Fisher's exact test, or the Chi-square test for trend, as appropriate for categorical variables. For continuous variables with normal distribution, differences were assessed using Student's t-test for independent samples, preceded by Levene's test for homogeneity of variances. The significance level for all tests was set at 5% (α=0.05). Data management and statistical analyses were performed using IBM® SPSS® Statistics software, version 25.

3 Results

Of the 637 eligible individuals invited to complete the questionnaires, 36 (5.7%) declined to participate. Among the 601 (94.3%) who agreed to participate, 43 (7.2%) did not complete the Stage 1 questionnaire, and 46 (7.3%) dropped out during Stage 2. This resulted in a final cohort of 512 participants (85.5%) with at least one day of follow-up, including responses on perceived air quality and mental health indicators (see Figure 2).

The sample was mostly composed of male individuals (n=308; 55.3%), aged from 18 to 72 years, with 16 or more years of education (n=404; 72.9%), non-Black (n=300; 54.2%), not living with a partner (n=332; 60.0%), affiliated with a religion (n=377; 67.6%), undergraduates (n=277; 49.6%), engaged in professional activities (n=422; 76.0%),

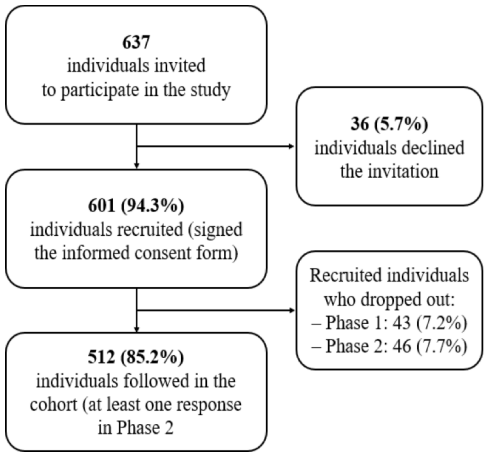


Figure 2. Flowchart of the sample: invited individuals, decliners, dropouts, and cohort participants.

and belonging to economic classes D or E (n=325; 59.2%) (see Table 1).

In Stage 1, a comparative analysis between participants (n=512) and dropouts (n=46) revealed differential distributions in certain sociodemographic characteristics. The dropout group was predominantly male (n=32; 69.6%). The median age was significantly lower among dropouts (p=0.005), with a median of 22 years (IQR 25%–75%: 20–25 years), compared to 27 years (IQR 25%–75%: 21–37 years) among participants. Most dropouts were undergraduate students (n=37; 80.4%; p<0.001) and were less likely to be professionally active (43.5% vs. 22.2%; p=0.001). Other variables, including education level, race/ethnicity, religion, cohabitation with a partner, number of household members, and average family income, did not show statistically significant differences between participants and dropouts (p>0.05) (see Table 1).

Among the total recruited participants, 41 individuals (7.4%) self-identified as smokers, and 38.2% (n=213) reported having had or currently experiencing some form of allergy, with respiratory allergies being the most prevalent (n=153; 71.8%). About chronic conditions, 44.4% (n=248) of the sample reported having one or more chronic diseases, with rhinitis and sinusitis being the most frequent, showing prevalence rates of 52.8% (n=131) and 41.5% (n=103), respectively.

A comparative analysis between the participant group (n=512) and the dropout group (n=46) revealed differential distributions in the variables "presence of chronic disease", "receiving psychological/psychiatric care", and "use of medication for anxiety and depression", all of which were more frequent among participants (p<0.05) (see Table 2).

Self-perceived general health was reported as very good or excellent by 45.5% (n=254) of the sample, with a statistically significant difference between participants and dropouts (p=0.002). In the dropout group, 67.4% (n=31) rated their general health as very good or excellent, compared to 43.6% (n=223) in the participant group (see Figure 3A).

Performance in frequent functional and social roles—including activities at home, work, and in the community, as well as responsibilities as a parent, child, spouse, employee, or friend—was rated as very good or excellent by 37.9% (n=211) of the sample (see Figure 3B), with no significant difference between participants and dropouts (p=0.414).

Global physical health was classified as excellent or very good in 37.1% of the sample (n=207). Although a higher proportion of dropouts (45.7%; n=21) reported excellent or very good physical health compared to participants (36.3%; n=186), this difference was not statistically significant (p=0.210) (see Figure 3C). Regarding global mental health, 32.4% of the sample (n=181) rated it as excellent or very good, with similar distribution between participants and dropouts (p=0.722) (see Figure 3D).

Regarding follow-up, a total of 4,511 responses were collected during Stage 2, ranging from 1 to 15 responses per participant, with a median of eight responses (IQR 25%–75%: 5–14 responses) (data not shown in figure). The median duration of cohort follow-up was 23 days (IQR 25%–75%: 16–28 days), ranging from 1 to 63 days (see Figure 4), with a median interval of 2.5 days (IQR 25%–75%: 2–4 days) between consecutive responses.

Considering mild to moderate scores, the prevalence of anxiety was 54.7% and depression was 41.8%, with a strong positive correlation between anxiety and depression scores (R2=0.788) (see Figure 3). Score levels varied: individuals with no symptoms of anxiety or depression were in the lower-left quadrant, while those with symptoms of anxiety and/or depression appeared in the other quadrants (see Figure 5).

4 Discussion

Cohort-based epidemiological studies are essential for the development of artificial intelligence algorithms capable of predicting mental health states based on the characteristics of indoor environments, given that predictive computational models rely on real-world data.

In addition to advancing knowledge in the field of health-oriented technology, the present study contributes to health promotion and well-being initiatives, in alignment with WHO recommendations for the care and maintenance of indoor environments (World Health Organization, 2010).

The development of innovative technologies applied to telehealth, authorized in Brazil since December 2022 (Brazil, 2022), enables closer interaction between healthcare professionals and patients, simplifying health monitoring with lower costs and greater efficiency, while also ensuring continuity of care (Araújo *et al*., 2023). The use of artificial intelligence in healthcare can streamline technical and logistical processes, providing instant and accurate notifications regarding changes in a patient's health status (World Health Organization, 2021).

The prevalence rates of anxiety and depression observed in this sample exceed those typically expected in academic settings in Brazil (Demenech *et al*., 2021). Although this epidemiological scenario is concerning, it presents a methodological advantage for the current study. The inclusion of participants with diverse outcomes — particularly those with extreme scores — is crucial for enhancing the accuracy of predictive algorithms (Duffy *et al*., 2020). Therefore, the ideal sample composition should not be determined solely by data volume, but also by the heterogeneity of clinical profiles, to reflect the reality of clinical care in the general population. This diversity directly contributes to improved accuracy in computational modeling.

Table 1. Distribution of sociodemographic and economic characteristics of the sample and comparison between participants and dropouts

Characteristic	Participants n/total (%) n=512	Dropouts n/total (%) n=46	Total n/total (%) n=558	p-value
Sex				
Male	276/511 (54.0)	32/46 (69.6)	308/557 (55.3)	0.042[a]
Female	235/511 (46.0)	14/46 (30.4)	249/557 (44.7)	0.042[a]
Age (years)	**n=500**	**n=42**	**n=542**	
Median (IQR)	27 (21–37)	22 (20–25)	26 (21–37)	0.005[b]
Minimum–maximum	18–72	19–63	18–72	0.005[b]
Education (years)				
≤ 8	10/508 (2.0)	–	10/554 (1.8)	0.805[c]
9–11	11/508 (2.2)	–	11/554 (2.0)	0.805[c]
12–15	113/508 (22.2)	16/46 (34.8)	129/554 (23.3)	0.805[c]
≥ 16	374/508 (73.6)	30/46 (65.2)	404/554 (72.9)	0.805[c]
Self-reported race/ethnicity				
Black	232/509 (45.6)	22/45 (48.9)	254/554 (45.8)	0.669
Non-black	277/509 (54.4)	23/45 (51.1)	300/554 (54.2)	0.669
Lives with a partner				
Yes	207/507 (40.8)	14/46 (30.4)	221/553 (40.0)	0.168[a]
No	300/507 (59.2)	32/46 (69.6)	332/553 (60.0)	0.168[a]
Religious affiliation				
Yes	348/512 (68.0)	29/46 (63.0)	377/558 (67.6)	0.494[a]
No	164/512 (32.0)	17/46 (37.0)	181/558 (32.4)	0.494[a]
University role				
Undergraduate student	240/512 (46.9)	37/46 (80.4)	277/558 (49.6)	< 0.001[a]
Graduate student	7/512 (1.4)	–	7/558 (1.3)	< 0.001[a]
Interns/scholarship holders	86/512 (16.8)	1/46 (2.2)	87/558 (15.6)	< 0.001[a]
Faculty/researcher	14/512 (2.7)	–	14/558 (2.5)	< 0.001[a]
Administrative staff	165/512 (32.2)	8/46 (17.4)	173/558 (31.0)	< 0.001[a]
Engaged in professional activity				
Yes	396/509 (77.8)	26/46 (56.5)	422/555 (76.0)	0.001[a]
No	113/509 (22.2)	20/46 (43.5)	133/555 (24.0)	0.001[a]
Number of household members	**n=506**	**n=45**	**n=551**	
Median (IQR)	3 (2–4)	3 (2–4)	3 (2–4)	0.548[b]
Minimum–maximum	1–9	1–6	1–9	0.548[b]
Monthly household income[e]				
A/B/C classes	468/510 (91.8)	41/46 (89.1)	509/556 (91.5)	0.576[d]
D/E classes	42/510 (8.2)	5/46 (10.9)	47/556 (8.5)	0.576[d]

Note: [a] Chi-square test. [b] Mann–Whitney test. [c] Chi-square test for trend. [d] Fisher's exact test. A significance level of 5% was adopted for all statistical tests. Bold p-values indicate statistically significant differences. [e] Household income classification: A/B/C classes $\geq$ R$1,965.87; D/E classes $<$ R$1,965.87.

Results from this application are presented in the article "A machine learning approach for anxiety and depression prediction using PROMIS® Questionnaires" (Vitoria *et al*., 2024). Machine learning algorithms using PROMIS® questionnaires have shown promise as a methodology for assessing anxiety and depression scores from the participants' perspective, aligning with their perceptions of well-being and workplace environmental quality (Vitoria *et al*., 2024). In that study, machine learning models with greater capacity to capture complex data patterns — such as multi-layer perceptron models — proved to be more suitable for predicting anxiety and depression scores.

The next step will be to develop machine learning algorithms based on other validated instruments, such as PHQ-9, GAD-7, and MBI (Sawaya *et al*., 2016; Maslach *et al*., 1997), to assess depressive symptoms, anxiety, and burnout, respectively. Additionally, the study aims to investigate the correlation between these symptoms and perceived air quality in indoor environments. Based on these advances, a mobile application is currently under development to support healthcare professionals through teleconsultation, assisting in more precise clinical decision-making for the treatment of individuals with mental disorders. The implementation of this technology primarily aims to prevent crises resulting from prolonged exposure to environments with poor air quality.

Table 2. Distribution of Clinical Characteristics of the Sample and Comparison Between Participants and Dropouts

Clinical Characteristics	Participants n/total (%)	Dropouts n/total (%)	Total n/total (%)	p-value[a]
Pregnant	2/235 (0.9)	1/14 (7.1)	3/249 (1.2)	0.322
Smoker	36/511 (7.0)	5/46 (10.9)	41/557 (7.4)	0.370
Has or had allergies				0.614[b]
Respiratory	140/197 (71.1)	13/16 (81.3)	153/213 (71.8)	0.565
Dermatological	62/197 (31.5)	4/16 (25.0)	66/213 (31.0)	0.781
Food-related	41/197 (20.8)	5/16 (31.2)	46/213 (21.6)	0.347
Ophthalmological	12/197 (6.1)	3/16 (18.8)	15/213 (7.0)	0.091
Drug-related	10/197 (5.1)	1/16 (6.2)	11/213 (5.2)	0.586
Chronic disease	236/512 (46.1)	12/46 (26.1)	248/558 (44.4)	**0.009**[b]
Rhinitis	124/236 (52.5)	7/12 (58.3)	131/248 (52.8)	0.695[b]
Sinusitis	99/236 (41.9)	4/12 (33.3)	103/248 (41.5)	0.766
Bronchitis	29/236 (12.3)	2/12 (16.7)	31/248 (12.5)	0.650
Hypertension	28/236 (11.9)	–	28/248 (11.4)	0.386
Obesity	23/236 (9.7)	–	23/248 (9.3)	0.520
Asthma	19/236 (8.1)	2/12 (16.7)	21/248 (8.5)	0.269
Hypothyroidism	17/236 (7.2)	3/12 (25.0)	20/248 (8.1)	0.062
Pharyngitis	10/236 (4.2)	1/12 (8.3)	11/248 (4.4)	0.427
Arthritis	8/236 (3.4)	–	8/248 (3.2)	> 0.999
Laryngitis	7/236 (3.0)	–	7/248 (2.8)	> 0.999
Cardiovascular disease	5/236 (2.1)	–	5/248 (2.0)	> 0.999
Diabetes	4/236 (1.7)	–	4/248 (1.6)	> 0.999
Rheumatism	4/236 (1.7)	–	4/248 (1.6)	> 0.999
Cancer	2/236 (0.8)	–	2/248 (0.8)	> 0.999
Osteoporosis	1/236 (0.4)	–	1/248 (0.4)	> 0.999
Clinical diagnosis of anxiety/depression	173/512 (33.8)	12/45 (26.7)	185/557 (33.2)	0.331[b]
Psychological or psychiatric follow-up	112/512 (21.9)	3/45 (6.7)	115/557 (20.6)	**0.016**[b]
Continuous use of medication for anxiety/depression	78/511 (15.3)	1/45 (2.2)	79/556 (14.2)	**0.016**[b]

Note: [a] Fisher's exact test ($p < 0.05$). [b] Chi-square test ($p < 0.05$). Bold p-values indicate statistically significant differences.

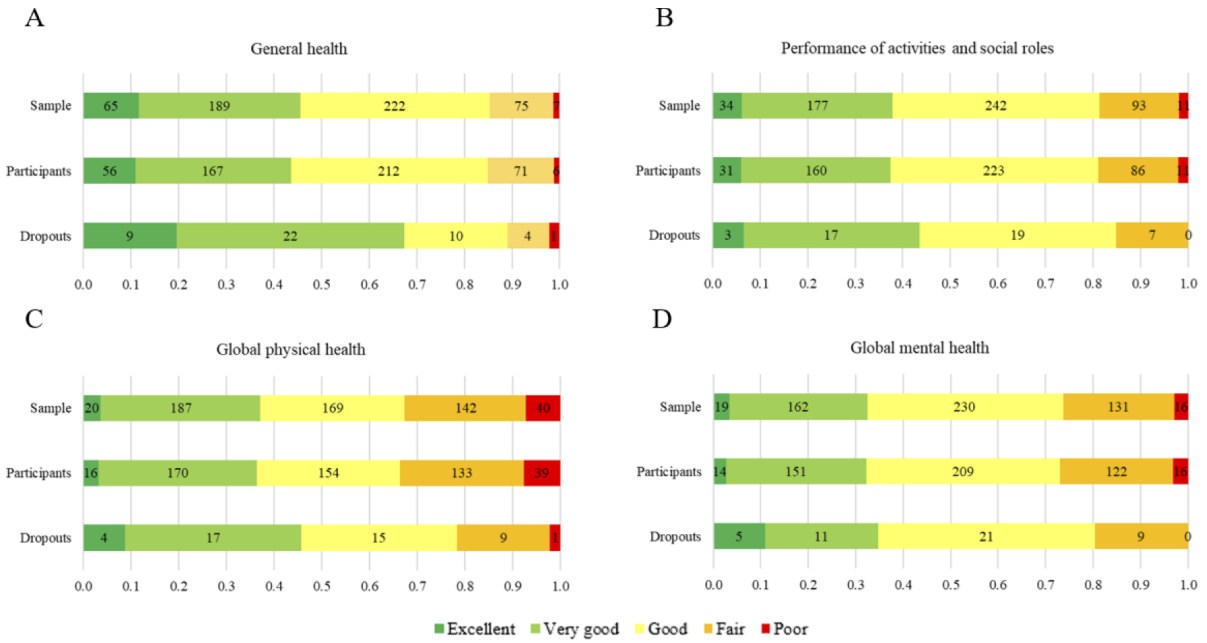


Figure 3. Classification of t-scores for general health (3A), functional and social role performance (3B), global physical health (3C), and global mental health (3D). Note: PROMIS® Global Health v.1.2 T-scores.

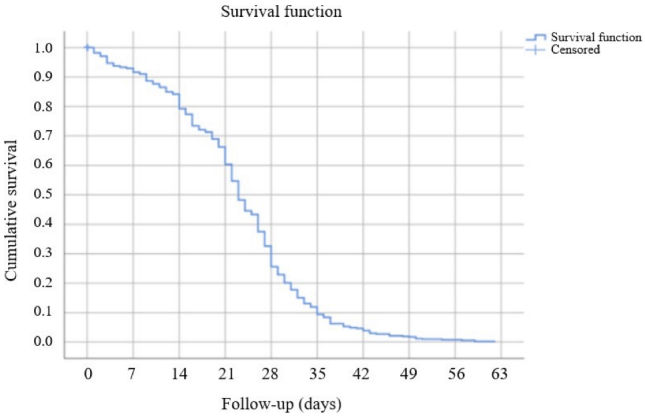


Figure 4. Survival function in the study according to follow-up duration (in days).

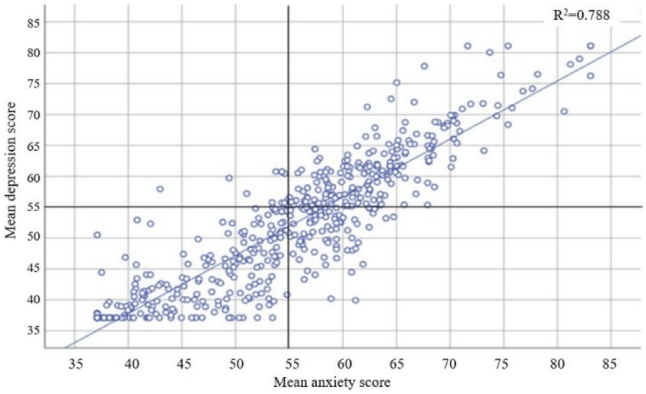


Figure 5. Correlation between mean anxiety and depression scores in the participant group (n=512).

5 Conclusion

This study presents a methodologically sound sample to support the development of machine learning algorithms capable of predicting risks associated with poor indoor air quality, which may impact individuals' quality of life and mental health. Despite differential follow-up losses related to certain sociodemographic characteristics, the sample remains valuable for investigation, as it provides data representative of high prevalence rates of anxiety and depression, with wide score variability, including extreme values. These features offer a solid foundation for building more robust predictive models, enhancing the ability to anticipate adverse clinical outcomes among individuals who study and/or work in indoor spaces.

Declarations

Funding

This research was financially supported by Dynamic Mind Health (DM Health), the Brazilian Micro and Small-Business Support Service (SEBRAE), and the Center of Excellence in Artificial Intelligence from the Federal University of Goiás (CEIA-UFG).

Authors' Contributions

Ana Laura de Sene Amâncio Zara: Methodology, Investigation, Writing – original draft, Writing – review and editing.
Mariana Maria Batista, **Bianca Ribeiro Soares**, **Myllena D'avila Mariana Gomes Silva**, **Joédan Silva Santos**, **Myrella Alves Bordado**, and **Júlia Gomes Moreira**: Investigation, Data curation.
Melyssa Mariana Gomes Silva and **Antônio Emílio Pereira**: Software, Data curation, Visualization.
Daniel Fazzioni, **Murilo de Oliveira Guimarães**, **Alex Victor Echeverria Ugarte Silva**, **Aldo André Díaz Salazar**, and **Arthur Ricardo de Sousa Vitória**: Formal analysis, Software, Methodology, Validation.
Paula Geralda Barbosa Coelho and **Paulo Roberto D'Aurea Machado**: Conceptualization, Funding acquisition.
Rodrigo da Silva Dias and **Iwens Gervásio Sene Júnior**: Conceptualization, Methodology, Investigation, Funding acquisition, Writing – original draft.
Renato de Freitas Bulcão-Neto: Supervision, Funding acquisition, Writing – review and editing.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

The datasets (and/or softwares) generated and/or analysed during the current study cannot be made available due to property rights.

Further relevant information

This study was conducted in accordance with the ethical standards of the institutional and/or national research committee and with Resolution No. 466, dated December 12, 2012, of the Brazilian National Health Council. The research protocol was approved by the Ethics Committee on Human Research from the Federal University of Goiás, Brazil (protocol no. 5.513.411). Parental/guardian informed consent was obtained for all participants included in the study.

References

ANSI/ASHRAE (2023). Thermal environmental conditions for human occupancy. Standard ANSI/ASHRAE Standard 55-2023, American Society of Heating, Refrigerating and Air-Conditioning Engineers, Atlanta, GA.

Araújo, H. P. A., Santos, L. C., and Alencar, R. A. (2023). Telehealth: The experience of health professionals in the sector. *Revista da Escola de Enfermagem da USP*, 57. DOI: 10.1590/1980-220X-REEUSP-2022-0374en.

Arias, D., Saxena, S., and Verguet, S. (2022). Quantifying the global burden of mental disorders and their economic value. *EClinicalMedicine*, 54:101675. DOI: 10.1016/j.eclinm.2022.101675.

Brazil (2022). Law no. 14.510, of december 27, 2022. `https://www.planalto.gov.br/ccivil_03/_ato2019-2022/2022/lei/l14510.htm`.

Brazil Ministry of Health and National Health Council (2012). Resolution no. 466, of december 12, 2012. `https://www.gov.br/conselho-nacional-de-saude/pt-br/atos-normativos/resolucoes/2012/resolucao-no-466.pdf/view`.

Chojer, H., Branco, P. T. B. S., Martins, F. G., Alvim-Ferraz, M. C. M., and Sousa, S. I. V. (2020). Development of low-cost indoor air quality monitoring devices: Recent advancements. *Science of the Total Environment*, 727:138385. DOI: 10.1016/j.scitotenv.2020.138385.

De Angel, V., Lewis, S., White, K., Oetzmann, C., Leightley, D., Oprea, E., *et al.* (2022). Digital health tools for the passive monitoring of depression: A systematic review of methods. *NPJ Digital Medicine*, 5(1):1–14. DOI: 10.1038/s41746-021-00548-8.

Demenech, L. M., Oliveira, A. T., Neiva-Silva, L., and Dumith, S. C. (2021). Prevalence of anxiety, depression and

suicidal behaviors among brazilian undergraduate students: A systematic review and meta-analysis. *Journal of Affective Disorders*, 282:147–159. DOI: 10.1016/j.jad.2020.12.108.

Duffy, A., Faurholt-Jepsen, M., and Ostacher, M. (2020). Using big data to advance mental health research. *Evidence-Based Mental Health*, 23(1). DOI: 10.1136/ebmental-2020-300143.

Eisenstadt, A., Liverpool, T. S., Infanti, E., Ciuvat, R. M., and Carlsson, C. (2021). Mobile apps that promote emotion regulation, positive mental health and wellbeing in the general population: Systematic review and meta-analysis. *JMIR Mental Health*, 8(11). DOI: 10.2196/31170.

Finell, E. and Nätti, J. (2019). The combined effect of poor perceived indoor environmental quality and psychosocial stressors on long-term sickness absence in the workplace: A follow-up study. *International Journal of Environmental Research and Public Health*, 16(24):4997. DOI: 10.3390/ijerph16244997.

Finell, E. and Nätti, J. (2020). Self-reported mould and long-term sickness absences from work. *Scandinavian Journal of Public Health*, 49(8):914–920. DOI: 10.1177/1403494820919561.

GBD 2019 Mental Disorders Collaborators (2022). Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990–2019: A systematic analysis. *Lancet Psychiatry*, 9(2):137–150. DOI: 10.1016/S2215-0366(21)00395-3.

Hallal, P. C., Sardinha, L. M. V., Aquino, E. C., Wehrmeister, F. C., and Paula, P. C. B. (2023). Covitel 2: Final report. Technical report, Vital Strategies and Umane. `https://www12.senado.leg.br/noticias/arquivos/2024/08/30/relatorio_covitel_2023.pdf`.

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., and Conde, J. G. (2009). Research electronic data capture (redcap). *Journal of Biomedical Informatics*, 42(2):377–381. DOI: 10.1016/j.jbi.2008.08.010.

HealthMeasures (2023). Patient-reported outcomes measurement information system (PROMIS). `https://www.healthmeasures.net/explore-measurement-systems/promis`.

Institute for Health Metrics and Evaluation (2021). Mental health. `https://www.healthdata.org/research-analysis/health-topics/mental-health`.

Joshi, S. (2008). The sick building syndrome. *Indian Journal of Occupational and Environmental Medicine*, 12(2):61. DOI: 10.4103/0019-5278.43262.

Maslach, C., Jackson, S. E., and Leiter, M. P. (1997). Maslach burnout inventory: Third edition. In Zalaquett, C. P. and Wood, R. J., editors, *Evaluating Stress: A Book of Resources*, pages 191–218. Scarecrow Education, Palo Alto.

Pilkonis, P. A., Choi, S. W., Reise, S. P., Stover, A. M., Riley, W. T., and Cella, D. (2011). Item banks for measuring emotional distress from PROMIS®: Depression, anxiety, and anger. *Assessment*, 18(3):263–283. DOI: 10.1177/1073191111411667.

Santomauro, D. F., Herrera, A. M. M., Shadid, J., Zheng, P., Ashbaugh, C., Pigott, D. M., *et al.* (2021). Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the covid-19 pandemic. *Lancet*, 398(10312):1700–1712. DOI: 10.1016/S0140-6736(21)02143-7.

Sawaya, H., Atoui, M., Hamadeh, A., Zeinoun, P., and Nahas, Z. (2016). Adaptation and initial validation of the PHQ-9 and GAD-7 in an arabic-speaking lebanese psychiatric outpatient sample. *Psychiatry Research*, 239:245–252. DOI: 10.1016/j.psychres.2016.03.030.

Sheikh, M., Qassem, M., and Kyriacou, P. A. (2021). Wearable, environmental, and smartphone-based passive sensing for mental health monitoring. *Frontiers in Digital Health*, 3:662811. DOI: 10.3389/fdgth.2021.662811.

Silva, M. M. G., Pereira, A. E., França, I. F., Cordeiro, B. M. O. S., Sene Junior, I. G., and Bulcão-Neto, R. F. (2021). Compatibility between indoor air quality and user perception. In *8th Regional School on Computer Applications to Health*, pages 78–81. DOI: 10.5753/ercas.2021.17444.

Vitoria, A. R. S., Guimaraes, M. O., Fazzioni, D., Diaz-Salazar, A. A., Zara, A. L. S. A., Sene-Junior, I. G., and Bulcao-Neto, R. F. (2024). A machine learning approach for anxiety and depression prediction using PROMIS questionnaires. In *Proceedings of the 19th Conference on Computer Science and Intelligence Systems (FedCSIS)*, volume 39, pages 187–194. DOI: 10.15439/2024F3586.

Wolkoff, P. (2018). Indoor air humidity, air quality, and health – An overview. *International Journal of Hygiene and Environmental Health*, 221(3):376–390. DOI: 10.1016/j.ijheh.2018.01.015.

Wolkoff, P., Azuma, K., and Carrer, P. (2021). Health, work performance, and risk of infection in office-like environments: The role of indoor temperature, air humidity, and ventilation. *International Journal of Hygiene and Environmental Health*, 233:113709. DOI: 10.1016/j.ijheh.2021.113709.

World Health Organization (2010). *WHO Guidelines for Indoor Air Quality: Selected Pollutants*. WHO, Geneva.

World Health Organization (2021). *Ethics and Governance of Artificial Intelligence for Health: WHO Guidance*. WHO, Geneva.

World Health Organization (2025). Mental health. `https://www.who.int/health-topics/mental-health`.

Zumpano, C. E., Mendonça, T. M. S., Silva, C. H. M., Correia, H., Arnold, B., and Pinto, R. M. C. (2017). Cross-cultural adaptation and validation of the PROMIS global health scale in the portuguese language. *Cadernos de Saúde Pública*, 33(1). DOI: 10.1590/0102-311X00107616.