Application of Formal Concept Analysis for Identifying Depression Patterns in Adults in Brazilian National Health Survey

Diogo Rodrigo dos Reis[a], Mark Alan Song[b] and Cristiane Neri Nobre[c]

Institute of Exact Sciences and Informatics, Pontifical Catholic University of Minas Gerais, Dom José Gaspar, Belo Horizonte, Brazil

Keywords: Data Mining, Formal Concept Analysis, Dyadic Analysis, Depression, Adults.

Abstract: This study investigates the relationship between health status and lifestyle behaviors in relation to depression, utilizing data from the 2019 Brazilian National Health Survey (NHS). By applying Formal Concept Analysis (FCA) through dyadic analysis, we identify associations among health perceptions, health habits, social support, experiences of violence, and diagnosed conditions to explore depression patterns among Brazilian adults, including young adults, middle-aged adults, and older adults. The analysis reveals that a perception of good health is frequently linked to depression from the studied dataset. Gender-specific trends are also apparent, as females are more frequently diagnosed with depression compared to males, indicating a potential gender bias. Sleep problems and medicine to sleep were also associated with depression, as well as self-deprecating thoughts and some traces of violence experiences. Our findings emphasize the intricate interplay between health perceptions, health behaviors, and gender in understanding depression, highlighting the necessity for nuanced approaches in mental health assessments and interventions.

1 INTRODUCTION

Depression has emerged as a major public health issue, standing as the second most prevalent mental health disorder worldwide (Disease et al., 2018).

Data mining tools have been increasingly employed in the field of health to investigate depression and mental illnesses (Hassan et al., 2023), (Zhang et al., 2022), (Biilah et al., 2022). These methods enable researchers to uncover complex patterns and correlations within datasets, providing valuable insights into the factors that contribute to mental health conditions. The application of data mining tools offers significant potential for enhancing early detection and diagnosis, particularly in cases of depression, where undiagnosed conditions remain a pressing concern (Williams et al., 2017).

Depression is particularly relevant in Brazil, and it has been investigated through the current National Health Survey (NHS) data (Harding et al., 2022) which is a comprehensive household survey conducted in Brazil, aimed at assessing the health conditions, lifestyle habits, and access to healthcare services among the Brazilian population.

The first edition of the NHS was conducted in 2013, followed by a subsequent survey in 2019. Both provide information evolving health landscape of Brazil, offering a robust database for policymakers, healthcare professionals, and researchers. They reveal significant trends and shifts in public health indicators, reflecting the socio-economic changes and the impact of health policies over the years.

The 2019 survey aimed to update the data collected in 2013 and explore new health challenges that emerged during this period. The scope of the survey was expanded to include additional questions on mental health, use of healthcare services, violence experiences, and preventive health measures, among others. With data from approximately 108,000 households, the NHS 2019 offered an updated and more detailed portrait of health conditions across Brazil.

In this context, this study aims to employ Formal Concept Analysis (FCA), using the NHS 2019 dataset to uncover associations between health status, health habits, social support, experience of violence, and diagnoses conditions to examine depression patters in Brazilian adults (young adults, middle-aged adults, and older adults).

[a] https://orcid.org/0000-0001-7908-7476
[b] https://orcid.org/0000-0001-7315-3874
[c] https://orcid.org/0000-0001-8517-9852

Reis, D. R., Song, M. A. and Nobre, C. N.
Application of Formal Concept Analysis for Identifying Depression Patterns in Adults in Brazilian National Health Survey.
DOI: 10.5220/0013186100003911

In *Proceedings of the 18th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2025) - Volume 2: HEALTHINF*, pages 647-654
ISBN: 978-989-758-731-3; ISSN: 2184-4305

2 FORMAL CONTEXT AND CONCEPT

Formal Concept Analysis (FCA), introduced by Rudolf Wille in the early 1980s, is a methodology for data analysis and knowledge representation (Wille, 1982). FCA structures data into a hierarchy of concepts, allowing the exploration of inherent relationships within a dataset. In this way, this method found applications in numerous fields, including machine learning, software engineering, and information retrieval (Wille, 1982), (Poelmans et al., 2013).

At its core, FCA is based on the notion of a *formal context*, which is defined as a triplet (G,M,I), where G is a set of objects, M is a set of attributes, and I is a binary relation indicating which objects have which attributes (Ganter et al., 1997).

From the context, FCA derives *formal concepts*, each consisting of an extent (a set of objects) and an intent (a set of attributes), satisfying certain mathematical properties (Ganter et al., 1997). These concepts form the building blocks for a deeper understanding of the data, as they encapsulate meaningful relationships between objects and their attributes. Formally, a *formal concept* is defined as a pair (A,B), where $A \subseteq G$ and $B \subseteq M$ such that A is the set of all objects that share all attributes in B, and B is the set of all attributes common to all objects in A. Mathematically, this can be expressed as Equations 1 e 2:

$$A = \{g \in G \mid \forall m \in B, (g,m) \in I\} \quad (1)$$

$$B = \{m \in M \mid \forall g \in A, (g,m) \in I\} \quad (2)$$

The collection of all formal concepts within a given formal context can be organized into a complete lattice, known as the *concept lattice*, where each node in the lattice represents a formal concept, and edges represent the inclusion relationships between concepts.

One of the key advantages of FCA is its ability to handle incomplete and inconsistent data effectively. Unlike many traditional data analysis methods that require complete and consistent datasets, FCA can work with data that has missing or contradictory information (Ganter et al., 1997). This makes FCA particularly useful in real-world applications since the robustness of the formal context and concept lattice of FCA provides a reliable means of extracting valuable insights from imperfect data and interpretable patterns within data, especially when dealing with complex datasets where traditional methods may fall short.

Furthermore, FCA has a strong theoretical foundation that supports its practical applications. The mathematical rigor behind FCA ensures that the concepts and relationships it identifies are well-defined and consistent (Ganter et al., 1997).

In the context of FCA, *association rules* are used to express implications between formal concepts. Unlike traditional association rules, which does not require the formal concept structure, FCA rules explicitly does (Agrawal and Srikant, 1994), taking the form $X \to Y$, where X and Y are formal concepts such that $X \subseteq Y$. These rules allow us to express that whenever the objects in concept X possess certain attributes, they are also likely to possess the attributes in concept Y. Two key metrics are used to evaluate the quality of association rules in FCA: *support* and *confidence*, such as:

$$\text{Support}(X \to Y) = \frac{|X \cap Y|}{|G|} \quad (3)$$

where $|X \cap Y|$ represents the number of objects that belong to both the antecedent and the consequent, and $|G|$ is the total number of objects in the formal context

$$\text{Confidence}(X \to Y) = \frac{|X \cap Y|}{|X|} \quad (4)$$

where $|X|$ is the number of objects in the antecedent. A higher confidence value indicates a stronger association between the antecedent and the consequent,

The formal concept lattice enables a more structured and hierarchical interpretation of association rules, which can be particularly valuable in applications where the relationships between attributes are complex and interdependent, being particularly useful for tasks such as concept learning, ontology construction, and the analysis of formal contexts in social and biological sciences (Roscoe et al., 2022).

3 RELATED WORKS

Concerning depression condition, Silva et al. (2014) investigated depression association with adults and detected a significant association between depression and health behaviors, particularly with smoking and physical activity. The associations found using the PHQ were similar to those observed with a single question about depressive mood. The study's results highlighted the importance of assessing the presence of depression and the frequency and severity of symptoms when implementing actions to promote healthy behaviors.

Lopes et al. (2022) found a significant increase in the prevalence of depressive symptoms over the six years between NHS 2013 and NHS 2019 surveys. The finding that the group of younger and unemployed

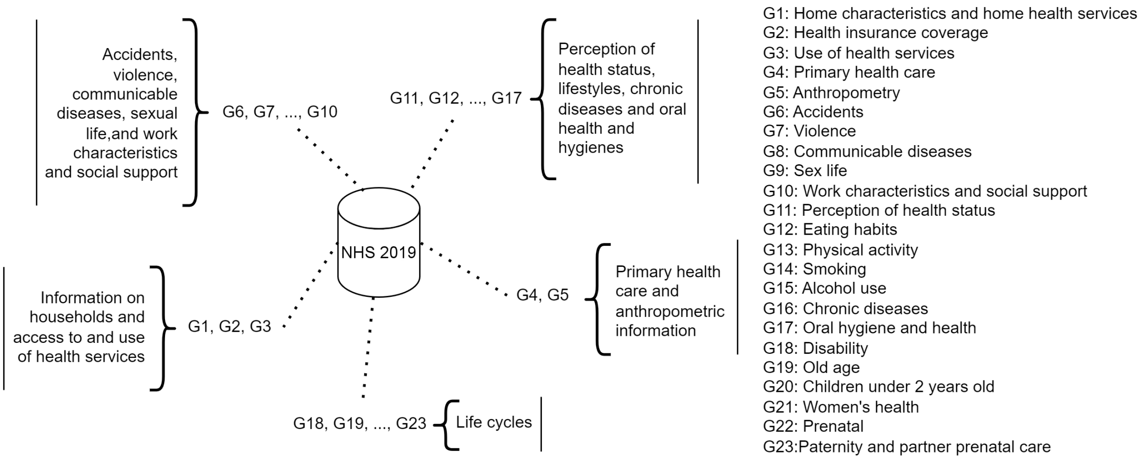

Figure 1: Organization of the groups of variables in the dataset in different dimensions of health.

men showed the highest variation in the prevalence of depressive symptoms.

Hintz et al. (2023) investigated associated factors to depression. The study included 88,531 participant records with 10.27% diagnosed with depression. The study showed the following factors associated with depression: age, brown and white race/skin color, female sex, poor, very poor, or regular self-reported health condition, diagnosis of cardiovascular disease, work-related musculoskeletal disorder, history of smoking habit, and macroeconomic region.

4 MATERIALS AND METHODS

4.1 Databases Description

This study utilizes data from the 2019 edition of the National Health Survey conducted in Brazil, aiming at assessing the health conditions, lifestyle habits, and access to healthcare services among the Brazilian population through a probabilistic sample of households. The information is provided by a resident who is deemed capable of representing the entire household, which is based on a structured questionnaire. The file questionnaires, dictionaries of the databases, and other data information can be accessed on the IBGE website[1]. The number of attributes and instances of the datasets are described in Table 1.

The dataset contains a wide array of variables that capture various aspects of the health and well-being of Brazilians. The database is structured into several modules where each one focus on different dimensions of health. Figure 1 shows the organization of these dimensions of health into groups.

[1]https://www.ibge.gov.br

Table 1: Summary of the variables and instances of the dataset.

Number of attributes	Number of instances
1,082	279,382

4.2 Methodology

This section outlines the methodological approach applied to the database to investigate the relationships between various health-related variables and the diagnosis of depression. The methodology was designed to ensure suitable data for subsequent formal concept analysis. Thus, the data is prepared to address dyadic analysis to extract association rules to be analysed and discussed. Figure 2 shows the method employed in this work.

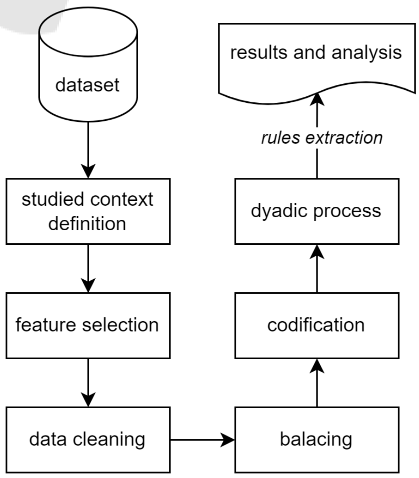

Figure 2: Adopted formal concept analysis method.

4.2.1 Studied Context Definition

In this study, we are investigating adults aged 18 to 59 years who were the same age as when they had been first diagnosed with depression. For instance, an adult who is currently 25 years old and was first diagnosed with depression at the age of 25 fits within the context of this study. Thus, following this reasoning, a person who is 25 years old and had been first diagnosed at the age of 20 does not fit the context of this study. This approach ensures that the respondent's age is as close as possible (according with the available data) to the time of their first depression diagnosis. In this way, the studied context delimited the dataset as show in Table 2:

Table 2: Summary of the studied context data.

Depression diagnosis		Total
Presence	Absence	
331	51,726	**52,057**

4.2.2 Feature Selection

Variables associated with health habits, previous chronic disease diagnostics, social support, perspective of physical and mental health, violence, anthropometry information, and demographic characteristics such as age, race, and gender were selected by the authors in order to investigate the depression disease. This approach involved systematically evaluating and selecting features based on their significance and contribution to the study objectives. Additionally, due to the dimensionality problem in FCA discussed in Section 2, the FCA method used (see Section 4.2.4), and the processing time required, using all the variables in the studied database becomes prohibitive, specially when one-hot encoding codification is employed. Thus, a smaller subset of attributes needed to be defined for the analyses.

Table 3: Summary of the feature selection.

Number of attributes	Number of selected attributes
1,082	**46**

4.2.3 Data Cleaning and Balancing

Data cleaning process was applied, handling missing values, correcting inconsistencies, and removing some instances. In general, the instances with missing values regardless to the selected attributes were removed, except for some data instances that were possible to correct their values.

An imbalance scenario is reflected in the dataset in the studied context: presence of diagnosed depressive disorder (minority) and absence of a diagnosed depressive disorder (majority). To address this issue, random undersampling technique was applied to the majority class at a 1.5 ratio (60% majority; 40% minority). This approach respects the real-world imbalance observed in the dataset while maintaining a mild to moderate level of imbalance, a well-known and recommended criterion in the literature for contexts of this nature. The balancing process was performed using the `imbalanced-learn` (Lemaître et al., 2017) library in Python. This library is widely recognized for its effectiveness in handling imbalanced datasets through various resampling techniques. Table 4 shows the result of the employed undersampling in the dataset.

Table 4: Summary of data balancing.

Minority (No.) (from→to)	Majority (No.) (from→to)	Total
331→**331**	51,726→**496**	**827**

4.2.4 Codification and Dyadic Analysis

The description of the feature engineering and codification used to prepare the data for FCA is tabularly available at *Codification Description Repository*. The FCA was performed using the `Lattice Miner tool`(Champin et al., 2004), a specialized software for formal concept analysis. This tool is intended to generate rules for further analysis and discussion.

5 RESULTS AND DISCUSSION

5.1 General Rules Analysis

Table 5 shows the rules extracted based on defined threshold for support and confidence, and size of antecedent of the rules. Note that, the smaller the antecedent size, the more generalizable the rule knowledge becomes, and its interpretation is less complex, for instance, the rule R1 is simpler to interpret than the rule R19.

The perception of good health appears to have a certain relationship with the status of depression. The presence of a perceived good health status suggests that the person does not have depression (see R6, R18, R19). Additionally, the absence of depression diagnose implies in perceived good heath state with 79.59% of confidence (see R2).

Table 5: General dyadic rules from NHS 2019.

Threshold			Rule #	Rule Description	Support (%)	Confidence (%)
Antecedent size	minimum support (%)	minimum confidence (%)				
1	15	75	R1	IF {Q132_} THEN {Q092_Y}	15.96	84.61
			R2	IF {Q092_N} THEN { N00101_}	47.64	79.59
			R3	IF {C006_M} THEN { Q092_N}	30.35	75.37
2	15	80	R4	IF {C009_W, Q092_N} THEN {N00101_}	18.74	90.11
			R5	IF {C008_A, Q092_N} THEN {N00101_}	20.19	86.97
			R6	IF {C006_M, N00101_} THEN {Q092_N}	25.03	86.97
			R7	IF {P02801_P029, Q092_N} THEN {N00101_}	24.30	86.63
			R8	IF {P035_F, Q092_N} THEN {N00101_}	22.85	86.30
			R9	IF {C006_M, Q092_N} THEN {N00101_}	25.03	82.47
			R10	IF {P02002_P02501, Q092_N} THEN {N00101_}	36.39	81.57
			R11	IF {P02801_P029, Q092_N} THEN {P02002_P02501}	22.85	81.46
			R12	IF {C008_A, Q092_N} THEN {P02002_P02501}	18.86	81.25
			R13	IF {P027_N, Q092_Y} THEN {C006_F}	19.34	81.21
			R14	IF {A02201_N, Q092_N} THEN {N00101_}	18.25	80.74
			R15	IF {P034_N, Q092_Y} THEN {C006_F}	18.98	80.51
			R16	IF {C006_M, Q092_N} THEN {P02002_P02501}	24.30	80.07
3	10	90	R17	IF {C009_W, P02801_P029, Q092_N} THEN {N00101_}	11.24	95.87
			R18	IF {C006_M, C009_NW, N00101_} THEN {Q092_N}	15.47	93.43
			R19	IF {C006_M, M01601_to_M01901 N00101_} THEN {Q092_N}	11.0	91.0

However, the individuals' perceived good health status may not be coherent with what is meant by good health status, as unhealthy habits such as frequently consuming sweets and soft drinks (R10) and frequent alcohol consumption (R7) imply in perceived good health status.

Frequent physical activity is associated with perceived good health status (see R8) as it was expected since physical activities contribute to psychical and physiologic good health (Penedo and Dahn, 2005) (Schuch et al., 2016). Due to this, an analyze of the influences on the perception of good health will be presented in a separate analysis.

When examining gender differences, it is observed that males are more commonly linked with the absence of a depression diagnosis (see R3), specially when they also have perceived of good health status (see R6, R19) and are non-white people (see R18).

Furthermore, the rules R13 and R15 suggest that people with diagnose of depression are associated with females. A separate section will be presented analysing gender and other demographic variables. Finally, the rule R1 shows that people who take medicine to sleep are associated with depression with 84.61% of confidence.

5.2 Insightful Identified Dyadic Rules

In this section, we discuss some meaningful extracted rules during the analysis, highlighting some aspects reveled, such as the associations with demographics, health conditions, and experiences of violence that were not uncovered in the general rules analysis. Table 6 presents these rules that will be discussed in the following sections.

5.2.1 Demographic Analysis

Table 6 shows gender, age groups, and race for analysing demographic aspects associated with depression. The evidence discussed that males are linked with the absence of depression is supported with 75.37% of confidence (see AR1) and 30.35% of support with a direct influence, specially males with perceived good health status (see AR2) and are non-white with 93.43% and 15.47% (AR4) compared with white ones with 78.21% and 9.55%, in the metrics of confidence and support, respectively. No relevant rules supported that age groups and race imply alone in the absence of depression diagnose, as well as gender, age groups, and race by their own do not imply either in the presence of depression status.

Table 6: Dyadic rules from NHS 2019: associations with demographics, health conditions, and experiences of violence.

Rule #	Rule description	Support (%)	Confidence (%)
	Demographic aspects: gender, age groups, and race		
AR1	IF {C006_M} THEN {Q092_N}	30.35	75.37
AR2	IF {C006_M, N00101_} THEN {Q092_N}	25.03	86.97
AR3	IF {C006_M, C008_OA} THEN {Q092_N}	14.99	78.98
AR4	IF {C006_M, C009_NW, N00101_} THEN {Q092_N}	15.47	93.43
AR5	IF {C006_M, C008_OA, N00101_} THEN {Q092_N}	11.60	88.07
AR6	IF {C006_M, C008_A, N00101_} THEN {Q092_N}	10.88	87.37
AR7	IF {C006_M, C008_YA, N00101_} THEN {Q092_N}	2.53	80.76
AR8	IF {C006_M, C009_W, N00101_} THEN {Q092_N}	9.55	78.21
	Perception of health status		
BR1	IF {N00101_} THEN {Q092_N}	47.64	73.50
BR2	IF {N00101_N} THEN {Q092_Y}	22.97	65.29
BR3	IF {C008_YA, N00101_N} THEN {Q092_Y}	3.50	80.55
BR4	IF {C009_W, N00101_N} THEN {Q092_Y}	8.10	79.76
BR5	IF {C008_A, N00101_N} THEN {Q092_Y}	8.46	73.68
BR6	IF {C006_F, N00101_N} THEN {Q092_Y}	16.80	70.91
BR7	IF {C008_YA, C009_W, N00101_N} THEN {Q092_Y}	1.57	92.85
BR8	IF {C006_F, C009_W, N00101_N} THEN {Q092_Y}	5.92	84.48
BR9	IF {C006_F, C008_YA, N00101_N} THEN {Q092_Y}	2.53	84.00
BR10	IF {C006_F, C008_A, N00101_N} THEN {Q092_Y}	5.68	79.66
	Sleep problems, eating disorders, or diagnosed conditions		
CR1	IF {Q132_} THEN {Q092_Y}	15.96	84.61
CR2	IF {Q11006_} THEN {Q092_Y}	11.72	79.50
CR3	IF {N014_} THEN {Q092_Y}	6.65	77.46
CR4	IF {N010_} THEN {Q092_Y}	7.73	72.72
CR5	IF {C009_W, Q079_} THEN {Q092_Y}	1.45	92.30
CR6	IF {C008_A, Q132_} THEN {Q092_Y}	6.28	91.22
CR7	IF {C009_NW, Q11006_} THEN {Q092_Y}	6.28	85.24
CR8	IF {C006_F, Q132_} THEN {Q092_Y}	12.93	86.99
CR9	IF {C008_A, N014_} THEN {Q092_Y}	2.66	81.48
CR10	IF {C006_F, N010_} THEN {Q092_Y}	6.04	80.64
CR11	IF {C006_F, N014_} THEN {Q092_Y}	4.83	80.00
CR12	IF {C009_W, N010_} THEN {Q092_Y}	2.41	80.00
CR13	IF {C008_A, N010_} THEN {Q092_Y}	2.9	79.99
CR14	IF {C008_OA, Q079_} THEN {Q092_Y}	3.26	75.00
CR15	IF {C009_NW, Q074_} THEN {Q092_Y}	2.05	70.83
CR16	IF {C006_F, Q074_} THEN {Q092_Y}	2.53	69.99
	Self-deprecating thoughts		
DR1	IF {N018_} THEN {Q092_Y}	2.41	83.33
DR2	IF {N017_} THEN {Q092_Y}	6.4	79.10
DR3	IF {N016_} THEN {Q092_Y}	7.49	77.49
DR4	IF {C008_YA, N016_} THEN {Q092_Y}	1.08	89.99
DR5	IF {C006_F, N018_} THEN {Q092_Y}	1.93	84.21
DR6	IF {C009_NW, N018_} THEN {Q092_Y}	1.93	84.21
DR7	IF {C008_A, N018_} THEN {Q092_Y}	1.2	83.33
DR8	IF {C008_OA, N018_} THEN {Q092_Y}	1.2	83.33
DR9	IF {C008_OA, N017_} THEN {Q092_Y}	3.5	82.85
DR10	IF {C009_W, N017_} THEN {Q092_Y}	2.17	81.81
	Violence		
ER1	IF {V02701_V02702} THEN {Q092_Y}	1.57	72.22
ER2	IF {C008_YA, V01401_to_V01403} THEN {Q092_Y}	1.2	83.33
ER3	IF {C009_NW, V02701_V02702} THEN {Q092_Y}	1.2	83.33
ER4	IF {C008_OA, V00203_} THEN {Q092_Y}	1.2	76.92
ER5	IF {C008_YA, V02801_V02802} THEN {Q092_Y}	1.93	76.19
ER6	IF {C008_YA, V00201_} THEN {Q092_Y}	2.66	75.86
ER7	IF {C006_M, V01401_to_V01403} THEN {Q092_Y}	1.45	74.99
ER8	IF {C006_F, V00203_} THEN {Q092_Y}	2.05	70.83

5.2.2 Analysis of the Perception of Health Status

Table 6 shows that perceived good health status is associated with absence of depression with 73.50% of confidence (see BR1) while the absence of perceived good health status is associated with presence of depression (see BR2) with 65.29% even though this last confidence is not as strong as the first.

Some demographic characteristics help to enhance this analysis. For instance, females (see BR6), white people (see BR4), adults (see BR5), and young adults (see BR3) without perceived good heath status are associated with presence of depression with a higher confidence.

5.2.3 Analysis of Sleep Problems, Eating Disorders, or Diagnosed Conditions

Table 6 shows that not only people that take medicine to sleep are associated with depression (see CR1), as discussed in previous analysis, but also people with sleep problems with 72.72% of confidence (see CR4), notably females (see CR8 and CR10) or adults (see CR6 and CR13), for both cases.

Concerning eating disorders, there is an association with depression, with a confidence level of 77.76% (see CR3), which also females (see CR11) and adults (see CR9) are more commonly associated with this condition.

In relation to diagnosed conditions, people with other mental illness diagnoses indicate to be associated with depression (see CR2), with 79.5% of confidence, particularly higher, with 85.24% of confidence level, in non-white people (see CR7). Additionally, asthma or asthmatic bronchitis condition was found to be slightly associated with depression in female (see CR16) or non-white people (CR15), as well as arthritis or rheumatism, however, the association is for white people (see CR5) or older adults (see CR14), with 92.3% and 75% of confidence, respectively.

5.2.4 Analysis of Self-Deprecating Thoughts

Individuals who experience self-deprecating thoughts demonstrated to be related with presence of depression in the analysis, showed in Table 6: frequently feeling depressed, down, or without hope is associated with 83.33% of confidence (see DR1); frequently feeling bad about yourself, considering yourself a failure, or feeling that you disappointed your family, with 79.10% of confidence (see DR2); having thoughts about hurting yourself in some way or thinking it would be better to be dead, with 77.49% in the same metric (see DR3).

5.2.5 Analysis of experiences of Violence

Experience of violence is related with depression in Brazilians (Harding et al., 2022). In this rules analysis, showed in Table 6, this relation is also evidenced. People who have ever experienced sexual abuse is associated with presence of depression with 72.22% of confidence (see ER1), specially non-white ones, with 83.33% in the same metric (see ER3). Additionally, recent sexual abuse is associated with depression in young adults with 76.19% of confidence (see ER5). Finally, there is exists a moderate relation between physical violence and depression in males, with 74.99% of confidence (see ER7), or in young adults, with 83.33% in the same metric (see ER2).

6 FINAL CONSIDERATIONS

In this study, we investigated depression in Brazillian population by applying Formal Concept Analysis in the NHS dataset. Several aspects were studied and some hidden knowledge was evidenced as the contribution of this work, emphasizing the potential of FCA for broader application in health contexts for identifying associations between different health factors. Thus, it is expected that numerous health domains stand to benefit from this methodology.

Due to the study's defined context, as discussed in 4.2.1, the dataset experienced a significant reduction in the number of instances, particularly within the minority class. This reflects an inherent characteristic of the dataset, as the NHS is not designed for any specific diagnosis condition. Also, ideally, for our investigation, the respondents would complete the questionnaire at the time of their diagnosis. In this study, we have made concerted efforts to approximate this timing as closely as possible, given the constraints of the database design, which ultimately reduces the data to be utilized.

For future research, we emphasize the importance of exploring instance selection and balancing approaches that more effectively represent the data for descriptive purposes. Finally, we recommend investigating and comparing the current findings with previous or subsequent editions of the NHS for comparison purposes.

ACKNOWLEDGEMENTS

The authors would like to thank the National Council for Scientific and Technological Development of Brazil (CNPq – Code: 311573/2022-3), the Coordination for the Improvement of Higher Education Personnel - Brazil (CAPES - Grant PROAP 88887.842889/2023-00 - PUC/MG, Grant PDPG 88887.708960/2022-00 - PUC/MG - Informatics and Finance Code 001), the Foundation for Research Support of Minas Gerais State (FAPEMIG – Codes: APQ-03076-18 and APQ-05058-23). The work was developed at the Pontifical Catholic University of Minas Gerais, PUC Minas, in the Applied Computational Intelligence Laboratory – LICAP.

REFERENCES

Agrawal, R. and Srikant, R. (1994). Fast algorithms for mining association rules in large databases. In *Proceedings of the 20th International Conference on Very Large Data Bases (VLDB)*, pages 487–499, Santiago, Chile. Morgan Kaufmann.

Biilah, M. A.-M., Raihan, M., Akter, T., Alvi, N., Bristy, N. J., and Rehana, H. (2022). Human depression prediction using association rule mining technique. In Khanna, A., Gupta, D., Bhattacharyya, S., Hassanien, A. E., Anand, S., and Jaiswal, A., editors, *International Conference on Innovative Computing and Communications*, pages 223–237, Singapore. Springer Singapore.

Champin, F., Garlatti, S., Napoli, A., and Kone, R. (2004). Lattice miner: A tool for concept lattice construction and exploration. In *International Conference on Formal Concept Analysis (ICFCA)*, pages 241–256. Springer.

Disease, G. ., Incidence, I., and Collaborators, P. (2018). Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the global burden of disease study 2017. *Lancet*, 392(10159):1789–1858. Epub 2018 Nov 8. Erratum in: Lancet. 2019 Jun 22;393(10190):e44. doi: 10.1016/S0140-6736(19)31047-5. PMID: 30496104; PMCID: PMC6227754.

Ganter, B., Wille, R., and Franzke, C. (1997). *Formal Concept Analysis: Mathematical Foundations*. Springer-Verlag, Berlin, Heidelberg, 1st edition.

Harding, D., Pitcairn, C. F. M., Machado, D. B., De Araujo, L. F. S. C., Millett, C., and Hone, T. (2022). Interpersonal violence and depression in brazil: A cross-sectional analysis of the 2019 national health survey. *PLOS Global Public Health*, 2(12):1–16.

Hassan, M. M., Karim, A., Mollick, S., Azam, S., Ignatious, E., and Haque, A. S. M. F. A. (2023). An apriori algorithm-based association rule analysis to detect human suicidal behaviour. *Procedia Computer Science*, 219:1279–1288. CENTERIS – International Conference on ENTERprise Information Systems / ProjMAN – International Conference on Project MANagement / HCist – International Conference on Health and Social Care Information Systems and Technologies 2022.

Hintz, A. M., Gomes-Filho, I. S., Loomer, P. M., de Sousa Pinho, P., de Santana Passos-Soares, J., Trindade, S. C., Cerqueira, E. d. M. M., Alves, C. M. C., Rios, Y. S. S., Batista, J. E. T., Figueiredo, A. C. M. G., and Cruz, S. S. d. (2023). Depression and associated factors among brazilian adults: the 2019 national healthcare population-based study. *BMC Psychiatry*, 23(1):704.

Lemaître, G., Nogueira, F., and Aridas, C. K. (2017). Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. *Journal of Machine Learning Research*, 18(17):1–5.

Lopes, C. d. S., Gomes, N. L., Junger, W. L., and Menezes, P. R. (2022). Trend in the prevalence of depressive symptoms in brazil: results from the ¡i¿brazilian national health survey¡/i¿ 2013 and 2019. *Cadernos de Saúde Pública*, 38:e00123421.

Penedo, F. J. and Dahn, J. R. (2005). Exercise and well-being: a review of mental and physical health benefits associated with physical activity. *Current opinion in psychiatry*, 18(2):189—193.

Poelmans, J., Ignatov, D. I., Kuznetsov, S. O., and Dedene, G. (2013). Formal concept analysis in knowledge processing: A survey on applications. *Expert Systems with Applications*, 40(16):6538–6560.

Roscoe, S., Khatri, M., Voshall, A., Batra, S., Kaur, S., and Deogun, J. (2022). Formal concept analysis applications in bioinformatics. *ACM Comput. Surv.*, 55(8).

Schuch, F. B., Vancampfort, D., Richards, J., Rosenbaum, S., Ward, P. B., and Stubbs, B. (2016). Exercise as a treatment for depression: A meta-analysis adjusting for publication bias. *Journal of Psychiatric Research*, 77:42–51.

Silva, M. T., Galvao, T. F., Martins, S. S., and Pereira, M. G. (2014). Prevalence of depression morbidity among brazilian adults: a systematic review and meta-analysis. *Brazilian Journal of Psychiatry*, 36(3):262–270.

Wille, R. (1982). Restructuring lattice theory: An approach based on hierarchies of concepts. In Rival, I., editor, *Ordered Sets*, pages 445–470, Dordrecht. Springer Netherlands.

Williams, S. Z., Chung, G. S., and Muennig, P. A. (2017). Undiagnosed depression: A community diagnosis. *SSM - Population Health*, 3:633–638.

Zhang, X., Zhang, L., Yu, Q., and Zhang, Z. (2022). A new association rule mining method and its application in mental health data analysis. In *2022 Global Conference on Robotics, Artificial Intelligence and Information Technology (GCRAIT)*, pages 540–543.