Analysis of Health Indicators for Heart Disease Based on Formal Concept Analysis

Laura Xavier[a], Julio Neves[b], Luiz Zarate[c] and Mark Song[d]
Universidade Católica de Minas Gerais, Computer Science Department, Minas Gerais, Brazil
{lauraxaiver018, juliocesar.neves}@gmail.com, {zarate, song}@pucminas.br

Keywords: Formal Concept Analysis, FCA, Health Indicators, Heart Disease, Lattice Miner.

Abstract: This study addresses the global concern of cardiovascular health by analyzing key risk factors such as high blood pressure, cholesterol levels, and smoking habits, which contribute to the onset of heart disease. Using Formal Concept Analysis (FCA), a mathematical framework for uncovering relationships in complex datasets, this research examines a health dataset of over 200,000 records to identify critical behavioral and health indicators related to cardiovascular problems. Although 80 association rules were extracted, 12 were selected for detailed analysis due to their significance in both risk and protective factors. Key findings reveal strong correlations between physical inactivity, poor dietary habits, and the likelihood of heart disease, providing actionable insights for healthcare professionals and policymakers. This study aims to deepen the understanding of cardiovascular risk factors and support the development of more effective prevention measures to improve global health outcomes.

1 INTRODUCTION

In the field of computer science and data analysis, advanced analytical methods have become essential for extracting meaningful insights from complex datasets. One such method is Formal Concept Analysis (FCA), a mathematical approach rooted in set theory and logic. FCA provides a structured framework for organizing data, uncovering patterns and relationships that might remain hidden in conventional analyses.

This study applies FCA to the public "Heart Disease Health Indicators Dataset," available on Kaggle (Teboul, 2022). This dataset, which includes over 200,000 records of patient health data, offers a valuable opportunity to explore the relationships between various health indicators and the likelihood of heart disease. By mapping key dimensions such as *Body Mass Index* (BMI), physical activity, cholesterol levels, and diabetes, this analysis aims to uncover risk factors and interdependencies that contribute to heart disease.

Cardiovascular diseases (CVDs) are the leading cause of death worldwide, responsible for an estimated 17.9 million deaths annually, according to the World Health Organization (Organization, 2021). CVDs encompass a range of disorders affecting the heart and blood vessels, such as coronary heart disease and cerebrovascular disease. Many of these conditions can be prevented or mitigated by addressing behavioral risk factors, such as tobacco use, unhealthy diets, physical inactivity, and excessive alcohol consumption, all of which are captured in the dataset.

Through this analysis, we aim to provide information that can support healthcare professionals and policymakers in designing more effective preventive measures and treatment strategies for reducing the impact of cardiovascular diseases.

2 RELATED WORK

According to the paper *Formal Concept Analysis – Overview and Applications* (Škopljanac Mačina and Blašković, 2014), FCA is a method for knowledge representation, information management, and data analysis. It identifies and visualizes all concepts and their dependencies from tabular input data. The resulting structure of concepts is hierarchically organized into a concept lattice, which can be presented as a Hasse diagram. The method is based on ap-

[a] https://orcid.org/0009-0008-7751-9640
[b] https://orcid.org/0000-0002-0520-9976
[c] https://orcid.org/0000-0002-3018-888X
[d] https://orcid.org/0000-0001-5053-5490

Xavier, L., Neves, J., Zarate, L. and Song, M.
Analysis of Health Indicators for Heart Disease Based on Formal Concept Analysis.
DOI: 10.5220/0013158100003911

In *Proceedings of the 18th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2025) - Volume 2: HEALTHINF*, pages 540-547
ISBN: 978-989-758-731-3; ISSN: 2184-4305

plied lattice theory and set theory, with applications in fields such as mathematics, medicine, biology, sociology, psychology, economics, and particularly computer science.

The study highlights FCA as a valuable tool for designing exams and analyzing results. Hierarchically organized concepts can be used to determine the optimal order of exam questions, thus facilitating student comprehension and evaluation. An example is given with an electrical engineering exam, showing how this approach can provide meaningful insights into student performance and question difficulty.

In summary, the literature demonstrates how FCA can be effectively applied to derive conclusions from datasets by providing valuable insights into the relationships between attributes and elements, leading to potential improvements and problem-solving approaches (Ananias et al., 2021) (Miranda et al., 2024).

In another study, (Song et al., 2024) explored the application of FCA in a triadic approach to characterize infant mortality in different regions of Minas Gerais, Brazil. Determinant factors such as birth weight, gestation, and APGAR scores were identified. The findings revealed associations between various variables, highlighting the importance of maternal education and prenatal care consultations.

3 BACKGROUND

This section aims to review the main tools and methodologies employed in this study, including Formal Concept Analysis and the Lattice Miner software

3.1 Cardiovascular Diseases

Cardiovascular diseases (CVDs) involve a broad spectrum of disorders that affect the heart and blood vessels, and their impact on global health is well-documented. Among the most prevalent conditions are coronary artery disease (CAD), which results from the buildup of plaques in the arteries (atherosclerosis), leading to reduced blood flow to the heart, and cerebrovascular diseases, such as stroke, where the brain's blood supply is disrupted (James and Smith, 2023). Both conditions share common risk factors, including high blood pressure, high cholesterol, smoking, and poor lifestyle choices.

The link between these modifiable risk factors and the development of CVDs has been a significant focus in public health. Regular physical activity, a healthy diet, and controlling blood pressure and cholesterol are crucial strategies for preventing these diseases (Lancet, 2020). While the impact of CVDs is well-known, there remains a need for ongoing research into the specific patterns and behaviors that increase the likelihood of these conditions. In recent years, advances in data analysis have allowed for a more nuanced understanding of how these risk factors interact, particularly through the use of methods like Formal Concept Analysis (FCA).

3.2 Formal Concept Analysis

Formal Concept Analysis (FCA) is a mathematical framework originally developed by Rudolf Wille in the 1980s for the formal representation of conceptual knowledge (Wille, 1982).

The method structures data into what is known as a *formal context*, which consists of three components: a set of objects, a set of attributes, and the incidence relation that links objects to their attributes. This formal context can be represented as a triple $K = (G, M, I)$, where:

- G represents the set of objects in the dataset,
- M denotes the set of attributes,
- $I \subseteq G \times M$ is the incidence relation, indicating the association between objects and their attributes, where $(g, m) \in I$ means that object g has attribute m.

Given a subset of objects $A \subseteq G$, the corresponding set of attributes shared by all objects in A is represented as:

$$A' := \{m \in M \mid \forall g \in A : (g, m) \in I\}.$$

Similarly, for a subset of attributes $B \subseteq M$, the set of objects that possess all attributes in B is:

$$B' := \{g \in G \mid \forall m \in B : (g, m) \in I\}.$$

A *formal concept* is then defined as a pair (A, B), where:

- A (the *extent*) is the set of objects that share the attributes in B,
- B (the *intent*) is the set of attributes shared by all objects in A.

For (A, B) to be a valid formal concept, it must satisfy $A = B'$ and $B = A'$. The collection of all such concepts forms a *concept lattice*, a structure that organizes the formal concepts in a hierarchical order based on their extents and intents (Ganter and Wille, 2012). This lattice is denoted as $\beta(K)$ and provides a visual representation of the relationships among concepts.

FCA has a natural application in extracting *association rules* from the formal context. These rules take the form $A \rightarrow B$, meaning that if an object has the attributes in A, it is likely to also have the attributes in B. The quality of an association rule is evaluated through two key metrics:

- **Support** (s): This measures the proportion of objects that contain both the attributes in A and B, calculated as:

$$s = \frac{|A' \cap B'|}{|G|},$$

which reflects the frequency of the rule in the dataset.

- **Confidence** (c): This indicates how often the attributes in B appear in the objects that already have the attributes in A, defined as:

$$c = \frac{|A' \cap B'|}{|A'|}.$$

A confidence of 100% means the rule is an *implication*, implying that all objects with the attributes in A also have those in B.

Through this approach, FCA allows for the discovery of patterns that might not be evident in traditional data analysis methods. For instance, it can be used to group patients with similar health indicators or identify the relationships between risk factors and diseases, as is the case in this study. The resulting concept lattice and association rules provide an interpretable and hierarchical view of the data, which is especially valuable in fields like medicine, where understanding complex relationships is critical (Stumme et al., 2002).

3.3 Lattice Miner

Lattice Miner 2.0 is a tool developed at the LARIM laboratory at the Université du Québec en Outaouais, designed to implement Formal Concept Analysis (FCA). Through this tool, formal concepts can be extracted, visualized, and explored from binarized datasets, organizing them into a lattice structure that reveals underlying patterns and relationships among the analyzed attributes (LARIM, 2017).

In this study, Lattice Miner will be used to facilitate the discovery of patterns among health indicators, applying FCA to explore the relationships between risk factors for heart disease.

4 METHODOLOGY

This section outlines the dataset used in the analysis of heart disease health indicators, detailing the data preparation process and the application of Formal Concept Analysis (FCA) through the use of the *Lattice Miner* software.

4.1 Heart Disease and Health Indicators Dataset

The source of information used in this study is the *Heart Disease Health Indicators Dataset*, publicly available on Kaggle (Teboul, 2022).

This dataset is derived from a larger record from the Centers for Disease Control and Prevention (CDC) and contains data from the Behavioral Risk Factor Surveillance System (BRFSS) survey conducted in 2015 (Dane and Centers for Disease Control and Prevention, 2015).

The survey includes health-related risk behaviors, chronic health conditions, and the use of preventive health services, containing over 400,000 responses and 330 features.

Alex Teboul, author of the Kaggle record, performed significant preprocessing steps to clean the data and reduce its dimensionality, as detailed in the dataset's accompanying notebook (Teboul, 2022).

The process included maintaining only the most relevant attributes for predicting heart disease, such as Body Mass Index (BMI), age, gender, physical and mental health, smoking status and alcohol consumption.

The target variable is a binary attribute indicating whether the respondent has ever been diagnosed with coronary heart disease (CHD) or myocardial infarction (heart attack).

These efforts resulted in a refined dataset, which contemplate 253,680 responses and 22 health indicators (Table 1), creating the basis for the analysis in this research.

4.2 Data Processing

In this study, additional steps were taken to tailor the dataset further. Specifically, socioeconomic variables were excluded because, while they are important in broader public health analyses, they are not directly relevant to the objective of this research, which focuses on health and behavioral indicators associated with heart disease. The following attributes were removed from the dataset: *Income*, *Education*, *NoDocbcCost*, and *AnyHealthcare*.

After excluding these variables, the dataset was split into two groups:

- **Group 1:** Individuals who had experienced heart disease or a heart attack.
- **Group 2:** Individuals who had not experienced heart disease or a heart attack.

This separation allowed for a more targeted analysis of the factors that contribute to heart disease within

Table 1: Attributes and Their Information.

Column Name	**Description**	**Values**
HighBP	Has high blood pressure	0 - 1
HighChol	Has high cholesterol	0 - 1
CholCheck	Cholesterol checked within past five years	0 - 1
BMI	Body Mass Index	12 - 98
Smoker	Smoking status	0 - 1
Stroke	Has had a stroke	0 - 1
Diabetes	Diabetes diagnosis	0 - 2
PhysActivity	Does physical activity	0 - 1
Fruits	Regular fruit consumption	0 - 1
Veggies	Regular vegetable consumption	0 - 1
HvyAlcoholCon	Heavy alcohol consumption	0 - 1
AnyHealthcare	Access to healthcare	0 - 1
NoDocbcCost	Could not see doctor due to cost	0 - 1
GenHlth	General health status	1 - 5
MentHlth	Mental health status	0 - 30
PhysHlth	Physical health status	0 - 30
DiffWalk	Difficulty walking	0 - 1
Sex	Gender	0 - 1
Age	Age range	1 - 13
Education	Education level	1 - 6
Income	Income level	1 - 8

More details about the possible values of each attribute can be found in the original author's notebook on Kaggle. (Teboul, 2022)

each population.

Subsequently, a series of custom Python algorithms were applied to binarize the dataset, transforming the attributes into binary labels to facilitate analysis and make the data compatible with the *Lattice Miner* software.

In this process, each attribute was categorized into distinct health indicators (Table 2). For instance, Body Mass Index (BMI) was categorized as "Overweight" for individuals with a BMI over 24, while those within a normal range were left blank.

This threshold was selected based on the World Health Organization (WHO) guidelines, which classify a BMI between 25 and 29.9 as "Overweight" (Organization, 2000). Age was split into "Young," "MiddleAged," and "Senior" categories.

Other health-related attributes, such as fruit and vegetable consumption, mental and physical health, and diabetes status, were similarly transformed. The target variable was also binarized, with one column indicating whether an individual had a heart problem or not.

At the conclusion of this process, both groups were left with 19 attributes each, the only difference being the target variable.

In **Group 1**, the target variable *HasHeartProb* was set to true for all individuals, indicating they had heart disease or experienced a heart attack.

Conversely, in **Group 2**, the target variable *NoHeartProb* was used, signifying that these individuals had not experienced any heart problems. **Group 1** contains 23,893 instances, while **Group 2** contains 229,787 instances.

This binarization process simplified the dataset and allowed for the identification of key health behaviors and conditions that correlate with heart disease. For instance, poor mental or physical health, obesity, and diabetes were labeled as "risk factors," while individuals who showed healthy behaviors, such as regular consumption of fruits and vegetables, were also identified.

4.3 Using Lattice Miner

Finally, the binarized data was processed using *Lattice Miner* (LARIM, 2017), to apply Formal Concept Analysis. This tool was used to extract association rules and generate concept lattices, providing a structured way to identify patterns and relationships between health behaviors and the occurrence of heart disease. The resulting rules can evidence how different combinations of factors may contribute to heart disease, forming the basis for the discussions and conclusions in this study.

5 RESULTS AND ANALYSIS

After applying the method, approximately 80 rules were extracted from the dataset. One example is $CholCheck$ and $PhysActivity \Rightarrow NoHeartProb$. This rule states that individuals who check their cholesterol and engage in physical activities have no heart problems, with a support of 73% and a confidence of 100%.

From the extracted rules, we selected 12 for detailed analysis in this work, covering both groups under study.

5.1 Group 1: Individuals with Heart Problems

The analysis of Group 1, which consists of individuals with heart problems, reveals several important patterns regarding their health-related behaviors. This section will delve into the rules extracted and discuss

Table 2: New Attributes and Their Information After Data Processing.

Original Attribute	Binarized Attribute	Precessing Description
BMI	Overweight	Marked if individual has BMI over 24
Age	Young, MiddleAged Senior	Categorized by ranges 1 - 3, 5 - 9 and 10 - 13
Fruits	Fruits/Veggies	Has regular fruit and/or vegetable consumption
Veggies		
MentHlth	BadMentHlth	Marked if individual has MentHlth over 15
PhysHlth	BadPhysHlth	Marked if individual has PhysHlth over 15
GenHlth	BadGenHlth	Marked if individual has GenHlth over 3
Diabetes	PD/Diabetes	Has diabetes, pre-diabetes or borderline diabetes
HeartDiseaseorAttack	HasHeartProb NoHeartProb	Categorized by whether the patient has or not a heart problem

The remaining attributes from the original dataset were retained without modification.

their implications in detail. The rules will be referenced by their indices, as specified in Table 3.

Table 3: Association Rules for Group 1 (Individuals with Heart Problems).

	Rule	Supp.	Conf.
1	HasHeartProb → CholCheck	98%	98%
2	CholCheck → HasHeartProb	98%	100%
3	HasHeartProb → HighBP	84%	85%
4	HighBP → HasHeartProb	84%	100%
5	HasHeartProb → Male	57%	57%

5.1.1 Cholesterol Check and Heart Problems

One of the most significant rules extracted from Group 1 is the *Rule 1*. This rule indicates that 98% of the individuals with heart problems in this group regularly undergo cholesterol checks. The confidence value suggests a very strong relationship, meaning that having heart problems is highly predictive of engaging in this health behavior.

This rule aligns with what we would expect from individuals who are already diagnosed with heart conditions. Regular monitoring of cholesterol is crucial for managing cardiovascular health, as high cholesterol is a well-known risk factor that can exacerbate heart conditions. A logical conclusion is that individuals in this group likely undergo routine cholesterol checks as part of their ongoing medical care to prevent further complications, such as stroke or heart attack.

The high support and confidence of this rule underscore the importance of cholesterol monitoring in the medical management of heart disease. It reflects how medical interventions and self-care practices are intertwined in the lives of those with heart problems.

A complementary rule, *Rule 2*, was also extracted with a similar level of strength. This rule has a support of 0.98 and a confidence of 1.0, indicating that every individual who had their cholesterol checked within the past 5 years in this group has heart problems. This might initially appear surprising, as we might expect some individuals without heart problems to also check their cholesterol.

However, this result is reflective of the group being analyzed, which specifically includes individuals with heart problems. The fact that every individual who checks their cholesterol in this group has heart problems is simply a consequence of the way the groups were defined. It is important to note that this rule applies to Group 1 and does not necessarily imply that cholesterol checks are exclusively performed by individuals with heart problems in the general population. Rather, it highlights that within this specific group, cholesterol monitoring is a near-universal practice among those managing heart conditions.

5.1.2 High Blood Pressure and Heart Problems

Another key rule extracted from Group 1 is *Rule 3*, which indicates that 84% of individuals with heart problems also suffer from high blood pressure, with a high confidence of 85%. This suggests a strong association between these two attributes.

The connection between heart disease and high blood pressure is well-established in medical literature. High blood pressure, or hypertension, is a major risk factor for heart disease, and it is common for individuals with one condition to also have the other. The high support of this rule reflects the prevalence of this comorbidity in Group 1, indicating that managing blood pressure is a critical component of care for individuals with heart problems.

The confidence level further strengthens the validity of this rule, implying that it is very likely for someone with heart problems to also have high blood pressure. This finding emphasizes the need for integrated care approaches that address both conditions simultaneously.

Additionally, *Rule 4*, indicates that every individual with high blood pressure in Group 1 also has heart problems. Similar to the rule regarding cholesterol checks, this result is reflective of the group under analysis. Since this group consists of individuals with heart problems, it is not surprising that all individuals with high blood pressure also have heart issues.

The perfect confidence level suggests that, within this group, high blood pressure is not seen in isolation but rather in conjunction with heart problems.

5.2 Group 2: Individuals without Heart Problems

The analysis of Group 2, which contains individuals without heart problems, provides a complementary perspective on health behaviors. This section analysed the rules extracted from Group 2, which will also be referenced by their indices, as specified in Table 4.

5.2.1 Cholesterol Check and No Heart Problems

One of the primary rules extracted from Group 2 is *Rule 6*. This rule indicates that 95% of individuals without heart problems in this group still perform cholesterol checks, with a high confidence level of 95%.

Table 4: Association Rules for Group 2 (Individuals without Heart Problems).

	Rule	Supp.	Conf.
6	NoHeartProb → CholCheck	95%	95%
7	CholCheck → NoHeartProb	95%	100%
8	NoHeartProb → PhysActivity	76%	76%
9	PhysActivity → NoHeartProb	76%	100%
10	NoHeartProb → Overweight	70%	70%
11	Overweight → PhysActivity	52%	74%
12	NoHeartProb → Female	57%	57%

It may seem to contrast with the similar rule in Group 1, where cholesterol checks were strongly associated with the presence of heart problems. However, in Group 2, this behavior likely reflects preventive health practices. Individuals in this group may undergo cholesterol checks as a way to monitor and manage their cardiovascular risk factors, even though they have not yet developed heart problems.

The high prevalence of cholesterol checks in this group also underscores the importance of regular monitoring in maintaining good health and preventing the onset of cardiovascular diseases. It suggests that individuals without heart problems are proactive about their health, using cholesterol checks as a way to detect potential issues before they develop into more serious conditions.

The reverse rule, *Rule 7*, was also a highlight. It indicates that every individual who have their cholesterol checked in this group does not have heart problems. While similar in structure to the corresponding rule in Group 1, the context is different.

In Group 2, this rule reflects the fact that cholesterol checks are being performed by individuals as part of their preventive health measures. Since this group specifically excludes individuals with heart problems, it is expected that everyone who keeps cholesterol monitoring in this group does not have heart problems. The perfect confidence level supports this conclusion.

It is possible to conclude that regular cholesterol monitoring is not limited to those with existing health issues but is also a common practice among those without heart problems. It reinforces the notion that cholesterol checks are an important component of preventive care in maintaining cardiovascular health.

5.2.2 Physical Activity and No Heart Problems

Physical activity is another important factor. The *Rule 8*, reveals that 76% of individuals without heart problems engage in physical activity, reiterating the health benefits of exercising.

Physical activity is widely recognized as a key factor for preventing heart disease and maintaining overall health. The support for this rule indicates that the majority of individuals in Group 2 engage in regular physical activity, likely as part of their efforts to stay healthy and avoid the development of cardiovascular problems. The confidence level further suggests that physical activity is a reliable indicator of the absence of heart problems in this group.

The complementary *Rule 9*, suggests that all individuals who engage in physical activity in Group 2 do not have heart problems.

The perfect confidence level supports the idea that regular physical activity is highly effective in preventing the development of heart problems. This finding aligns with the extensive body of research that shows the protective effects of physical exercise on cardiovascular health.

5.2.3 Overweight and No Heart Problems

Another interesting rule from Group 2 is *Rule 10*, which suggests that 70% of individuals without heart problems in this group are overweight. This result may seem counterintuitive, as being overweight is often associated with a higher risk of heart disease.

However, this rule highlights that being overweight does not automatically lead to heart problems, particularly if individuals are engaging in other healthy behaviors, such as regular physical activity. In fact, another notable rule is *Rule 11*, which indicates that 52% of individuals in Group 2 who are overweight also engage in regular physical activity, and there's a 74% likelihood that someone who is overweight will also be physically active.

Interestingly, when analyzing Group 1 (individuals with heart problems), no relevant rules were found linking being overweight with physical activity. This suggests that individuals with heart problems who are overweight may be less likely to engage in regular exercise compared to those in Group 2, potentially indicating a gap in their health management strategies.

Together, these two rules suggest that while a majority of individuals without heart problems may be overweight, many of them are proactively managing their health through preventive behaviors, like physical activity, which can significantly reduce their risk of developing cardiovascular issues in the future. This highlights the nuanced relationship between weight and heart health: although excess weight is often considered a risk factor, it is not necessarily deterministic of poor cardiovascular outcomes, especially when healthy behaviors are in place.

5.3 Gender and No Heart Problems

In Group 2, the *Rule 12* reveals that 57% of individuals without heart problems are female. This suggests that women in this group are less prone to heart issues, which aligns with broader epidemiological findings indicating that women generally develop heart disease later in life compared to men. The moderate confidence (0.57) points to a noticeable association between being female and the absence of heart disease in this group. This could be due to a variety of factors, including biological differences and healthier lifestyle choices, such as higher engagement in preventive healthcare practices among women.

In contrast, in Group 1 (individuals with heart problems), we observe a similar rule but related to male individuals: *Rule 5* indicates that 57% of individuals with heart problems are male. This suggests that men in this group are more likely to suffer from heart disease or have had a heart attack. The confidence value here indicates that being male is moderately associated with the presence of heart disease, which is consistent with the previous analysis of Group 2.

The comparison of these two rules emphasizes the stark gender differences in heart health between the groups. In Group 2, being female appears to offer a protective advantage against heart disease, whereas in Group 1, being male is more strongly associated with the presence of heart issues. This highlights the need for gender-specific approaches in both prevention and treatment, as women in the general population may be benefiting from healthier behaviors, while men, particularly those with heart conditions, may require more targeted interventions to address lifestyle risks.

5.4 Conflicting Results

Upon analyzing the rules related to cholesterol checks (`CholCheck`) across both groups, a potential contradiction emerges. Analysing the rules in both groups, individuals with and without heart problems show a strong association with cholesterol checks. Specifically, we observe:

- In Group 1: `HasHeartProb` → `CholCheck` and `CholCheck` → `HasHeartProb`.
- In Group 2: `NoHeartProb` → `CholCheck` and `CholCheck` → `NoHeartProb`.

At first glance, these rules may seem to present conflicting outcomes: individuals with heart problems check their cholesterol, but so do individuals without heart problems. However, this does not represent a true contradiction when we consider the differing contexts in which cholesterol checks occur in each group. In Group 1, individuals check their cholesterol as part of a treatment strategy, with cholesterol monitoring being a key component of managing an existing heart condition. In contrast, in Group 2, cholesterol checks are undertaken as a preventive measure, aimed at maintaining heart health and preventing the onset of heart problems.

In essence, cholesterol monitoring serves dual purposes: it is both a reactive measure for those managing heart disease and a proactive measure for those aiming to avoid it.

6 CONCLUSION AND FUTURE WORK

This study utilized Formal Concept Analysis (FCA) to investigate health indicators associated with heart disease. From approximately 80 extracted rules, 12 were selected for in-depth analysis, focusing on individuals with and without heart problems. The analysis uncovered important patterns, such as the strong relationship between factors like high cholesterol and blood pressure with the incidence of heart disease.

In future work, applying this method to larger, more varied datasets would likely yield richer information, particularly when combined with alternative data processing techniques. Additionally, refining the binarization process used in this study could help capture more subtle variations in health behaviors. Exploring Triadic Concept Analysis (TCA) could also provide a richer framework by incorporating additional dimensions such as time or conditions, allowing for a more complex analysis of how health indicators interact over different contexts. This could further enhance the understanding of the multifaceted nature of cardiovascular diseases.

Finally, investigating other cardiovascular conditions or expanding the analysis to longitudinal health data might provide a deeper understanding of how health indicators evolve over time, potentially improving early detection and prevention strategies.

ACKNOWLEDGEMENTS

The authors thank the Pontifícia Universidade Católica de Minas Gerais – PUC-Minas and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior — CAPES (CAPES – Grant PROAP 88887.842889/2023-00 – PUC/MG, Grant PDPG 88887.708960/2022-00 – PUC/MG - Informática, and Finance Code 001). The present work was also carried out with the support of Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) under grant number APQ-01929-22.

REFERENCES

Ananias, K. H., Missaoui, R., Ruas, P. H., Zarate, L. E., and Song, M. A. (2021). Triadic concept approximation. *Information Sciences*, 572:126–146.

Dane, S. and Centers for Disease Control and Prevention (2015). Behavioral risk factor surveillance system.

Ganter, B. and Wille, R. (2012). *Formal Concept Analysis: Mathematical Foundations*. Springer Science & Business Media.

James, S. and Smith, A. (2023). The global burden of cardiovascular diseases and risk factors in 2023. *The Journal of Cardiology*, 89:12–34.

Lancet (2020). Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the global burden of disease study 2017. *The Lancet*, 392(10159):1736–1788.

LARIM (2017). Lattice miner. https://github.com/LarimUQO/lattice-miner.

Miranda, D., Zárate, L., and Song, M. (2024). Application of formal concept analysis to characterize driving behaviors and socio-cultural factors related to driving. pages 56–62.

Organization, W. H. (2000). *Obesity: preventing and managing the global epidemic. Report of a WHO consultation*. WHO Technical Report Series 894. World Health Organization, Geneva.

Organization, W. H. (2021). Cardiovascular diseases (cvds).

Song, M. A. J., Galvez, L. E. Z., and Santos, D. S. (2024). Applying formal concept analysis to characterize infant mortality: Aplicando análise de formal de conceitos para a caracterização da mortalidade infantil. *Concilium*, 24(3):387–407.

Stumme, G., Taouil, R., Bastide, Y., Lakhal, L., and Nourine, L. (2002). A survey of formal concept analysis. In *Formal concept analysis: foundations and applications*, pages 123–142. Springer.

Teboul, A. (2022). Heart disease health indicators dataset notebook.

Wille, R. (1982). Restructuring lattice theory: an approach based on hierarchies of concepts. In *Ordered sets*, pages 445–470. Springer.

Škopljanac Mačina, F. and Blašković, B. (2014). Formal concept analysis – overview and applications. *Procedia Engineering*, 69:1258–1267. 24th DAAAM International Symposium on Intelligent Manufacturing and Automation, 2013.