The Use of Formal Concept Analysis for Characterizing the Behavior of the Residents of Bangladesh Regarding the COVID-19 Pandemic

Mateus Sobreira[a], Martha Dias[b], Luiz Zarate[c] and Mark Song[d]
Universidade Católica de Minas Gerais, Computer Science Department, Minas Gerais, Brazil
{mateus.sobreira, marta.dias}@sga.pucminas.br, {zarate, song}@pucminas.br

Keywords: Formal Concept Analysis, COVID-19, Bangladesh, Behavior Analysis.

Abstract: COVID-19 reported its first case on December 31, 2019, an illness rapidly spreading, becoming a global pandemic. Due to transmission through contact with respiratory droplets from infected individuals, governments implemented various preventive measures to minimize the spread of the disease. This work utilized Formal Concept Analysis (FCA) to assess the behavior of the inhabitants of Bangladesh regarding COVID-19, based on responses to a public questionnaire conducted virtually in 64 districts from April 1 to 10, 2020. A preprocessing stage was performed on the data to fit them into the format of a formal context with objects, attributes, and the relationships among them. From the resulting general formal context, it was possible to subdivide it according to the respondent gender, enabling an analysis of the behaviors of the men and women. Based on these formal contexts, association rules containing the most predominant relationships in the database were filtered using thresholds of 40% support and 80% confidence.

1 INTRODUCTION

The acute respiratory syndrome coronavirus 2 (SARS-CoV-2) reported its first case on December 31, 2019, in Wuhan, China, and spread rapidly, taking on global proportions (Chowdhury et al., 2022). The World Health Organization designated the disease as a pandemic on January 30, 2020, and it has since been recognized as a public health emergency of international concern.

The transmission of the disease primarily occurs through direct contact with an infected individual via inhalation of respiratory droplets. Once the infection sets in, many patients may develop an asymptomatic condition, while others may present with mild, moderate, or severe symptoms, which can potentially progress to fatal outcomes.

Based on the advancements in understanding the transmission mode of the disease, various countries adopted and encouraged measures to contain the virus to reduce the proliferation of COVID-19.

Chowdhury et al. (2022) discusses in their work some of the measures implemented by the government in Bangladesh, which is the focus of this study, aimed at assisting in the prevention of COVID-19. Campaigning for mask-wearing, adherence to social distancing, and use of hand sanitizer for hand hygiene were among the recommended measures.

On March 8, 2020, Bangladesh confirmed the first case of a patient infected with SARS-CoV-2, and, according to Ahmed and Rahman (2022), the peak of cases and deaths resulting from the disease occurred in June of the same year.

On October 30, 2020, Bangladesh had confirmed 404,760 positive cases of the virus, of which 5,886 individuals did not survive the syndrome.

Based on this data, the objective of this study is to conduct a behavioral analysis of the residents of Bangladesh regarding the COVID-19 pandemic, analyzing data on preventive actions such as regular handwashing and covering the nose and mouth when sneezing or coughing.

Formal Concept Analysis (FCA) is used in this study to extract inference rules that aid in understanding these behaviors.

[a] https://orcid.org/0009-0002-8983-4935
[b] https://orcid.org/0000-0002-0520-9976
[c] https://orcid.org/0000-0002-3018-888X
[d] https://orcid.org/0000-0001-5053-5490

Sobreira, M., Dias, M., Zarate, L. and Song, M.
The Use of Formal Concept Analysis for Characterizing the Behavior of the Residents of Bangladesh Regarding the COVID-19 Pandemic.
DOI: 10.5220/0013139300003911

In *Proceedings of the 18th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2025) - Volume 2: HEALTHINF*, pages 475-482
ISBN: 978-989-758-731-3; ISSN: 2184-4305

Table 1: Formal Context Example.

	attribute1	attribute2	attribute3
Record 1	**X**		
Record 2		**X**	
Record 3	**X**	**X**	
Record 4			**X**

2 BACKGROUND

2.1 Formal Concept Analysis

The Formal Concepts Analysis (FCA) theory is an applied mathematical field whose main objective is to represent and extract knowledge (Ganter and Wille, 1999). But it is also used for representation, information handling, and data analysis, given the attention received in the Data Mining field (Carpineto and Romano (2004)).

According to Škopljanac Mačina and Blašković (2014), this method is based on lattice theory and set theory.

Developed by Rudolf Wille in the early 1980s, it employs concepts interpretation as units formed over a set of objects and a set of shared attributes. FCA comprises three fundamental elements: formal context, formal concept, and rules.

FCA is a technique based on formalizing the notion of concept and structuring concepts in a conceptual hierarchy. The hierarchization of concepts extracted from data makes FCA a valuable tool for dependency analysis.

With the increase in cases in the healthcare area and due to the large amount of data generated, the study and improvement of techniques to extract knowledge are becoming increasingly justified.

FCA also permits data analysis through associations and dependencies of attributes and objects formally described from a dataset.

2.1.1 Formal Context

Formally, a formal context is formed by a triple (G,M,I), where G is a set of objects (rows), M is a set of attributes (columns) and I is defined as the binary relationship (incidence relation) between objects and their attributes where $I \subseteq G \times M$.

Table 1 exemplifies a formal context. In this example, objects correspond to tweets, attributes are the characteristics (terms), and the relationship of incidence represents whether or not the tweet has that characteristic. An $'X'$ is present in the table if the tweet possesses the corresponding characteristic.

2.1.2 Formal Concepts

Let (G,M,I) be a formal context, $A \subseteq G$ a subset of objects, and $B \subseteq M$ a subset of attributes. Formal concepts are defined by a pair (A,B) where $A \subseteq G$ is called extension and $B \subseteq M$ is called intention.

This pair must follow the conditions where $A = B'$ and $B = A'$ (Ganter and Wille, 1999). The relation is defined by the derivation operator ($'$):

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

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

If $A \subseteq G$, then A' is a set of attributes common to the objects of A. The derivation operator ($'$) can be reapplied in A' resulting in a set of objects again (A'').

Intuitively, A'' returns the set of all objects that have in common the attributes of A'; note that $A \subseteq A''$. The operator is similarly defined for the attribute set. If $B \subseteq M$, then B' returns the set of objects that have the attributes of B in common. Thus, B'' returns the set of attributes common to all objects that have the attributes of B in common; consequently, $B \subseteq B''$.

As an example, using Table 1, objects A = $\{record2, record3\}$ result in A' = $\{attribute2, attribute4\}$ when submitted to the operator described above. So $(\{record2, record3\}, \{attribute2, attribute4\})$ is a concept. All concepts found from Table 1 are displayed in Table 2.

In Table 2 there is a concept with an empty attribute set and a concept with an empty object set. They are called *infimum* and *supremum*, respectively.

2.1.3 Association Rules and Implications

The objective of this study in utilizing Formal Concept Analysis lies in the extraction of rules that describe the relationships between the attributes of objects within the formal context. The association rules, denoted as $A \longrightarrow B$, indicate that the objects containing the attributes in set A also include the attributes in set B.

From the inference of the association rules, it is possible to calculate their support, which represents the percentage of coverage of this rule relative to the total set of objects in the formal context. Another metric derived from the association rules is their confidence, which indicates the percentage of objects that contain the attributes in set A and also possess the attributes in set B.

Support and confidence are given by Equations 1 and 2, respectively.

$$support\,(r) = \frac{|A' \cap B'|}{|G'|} \qquad (1)$$

Table 2: Existing concepts in the formal context of Table 1.

Objects	Attributes
{Record 1, Record 2, Record 3, Record 4}	{}
{Record 4}	{Attribute 3}
{Record 1, Record 3}	{Attribute 1}
{Record 2, Record 3}	{Attribute 2, Attribute 4}
{}	{Attribute 1, Attribute 2, Attribute 3, Attribute 4}

$$confidence(r) = \frac{|A' \cap B'|}{|A'|} \qquad (2)$$

When an association rule has a confidence indicator of 100%, it is said to be an implication rule.

Those inference rules extracted by Formal Concept Analysis are used only to describe the characteristics of said formal context, therefore it is not correct to generalize their aspects to other contexts, even if similar.

2.2 Lattice Miner

In seeking the analysis of the formal context and extracting their inference rules, this study utilized the Lattice Miner data mining tool, version 2.0, released in April 2017. This open-source tool is based on the Java platform and was developed by the LARIM research laboratory at the Université du Québec en Outaouais under the supervision of Professor Rokia Missaoui.

One of the primary features of the Lattice Miner tool is its ability to deliver a set of low-level operations that allow for the manipulation of input data and association rules. This tool facilitates the generation of groups, commonly referred to as formal concepts, and the inclusion of their logical implications, thereby visually representing the binary relationship between the collection of objects and the set of attributes or properties. The software offers the following features:

1. Refinement and Approximation
2. Generation of Formal Concepts
3. Association Rule Extraction
4. Data Import and Preparation
5. Visualization and Exploration

3 RELATED WORK

As mentioned, researchers use Formal Concept Analysis as a tool for knowledge representation, information handling, and data analysis. Below, we describe studies that use FCA for extracting rules from datasets.

Table 3: 5-point Likert scale.

Value	Description
1	Strongly agree
2	Agree
3	Neither agree or disagree
4	Disagree
5	Strongly disagree

In the work of Salahuddin et al. (2018), a method for image sampling is proposed to improve the results of machine learning algorithms concerning breast cancer.

This new method, termed Hyper Conceptual Sampling (HCS), utilizes concepts from FCA to ensure that the image retains all original information during data filtering. Besides, no prior knowledge of the distribution of values in the original dataset is required.

A comparison using cross-validation to validate the HCS compared its results against the Spectral-spatial processing (SSP), proposed by Liang et al. (2017).

The results demonstrated that the HCS method achieved better accuracy and F1-score than the SSP method. In one of the results, the accuracy of HCS surpassed that of SSP by 2% when employing the SVM technique.

Meanwhile, in the work of Jay et al. (2013), FCA is proposed to group care trajectories based on the sequence of patient hospitalizations in France using data from a nationwide information system named *Programme de Médicalisation des Systèmes d'Information* (PMSI).

From the formal concepts, it was possible to extract an analysis focusing on two types of breast diseases and the methods used to combat them, resulting in insights regarding average expenditures, mortality rates, and average lengths of hospitalization.

With the data from this analysis, it was possible to observe that individuals with invasive neoplasms receiving palliative care had the highest values across all indicators, with an average expenditure of €26,139, a mortality rate of 69%, and an average length of hospitalization of up to 43 days.

The authors concluded that this analysis could assist healthcare professionals and the government in resource allocation planning and efforts based on patient data rather than relying solely on visits.

Meanwhile, the work of Miranda et al. (2024) presents an analysis of the sociocultural factors and their contribution to the behavior of Chinese drivers. They used FCA to extract inference rules that articulate the most relevant relationships between the objects and their attributes within the dataset.

The dataset in question consists of responses from drivers to a self-questionnaire regarding traffic practices. The rules extracted from the used dataset identified characteristics of aggressive driving behavior related to external factors, such as friends and family, based on the ”Mind-sponge theory”.

The study concluded that Formal Concept Analysis can help characterize these issues and assist in decision-making to understand the root of the problem of dangerous driving.

In summary, the literature shows how FCA can be applied in many different contexts (Ananias et al. (2021), Alves et al. (2023)).

4 METHODOLOGY

4.1 Database

The dataset used for this study, presented in Pakpour et al. (2020), encompasses knowledge, prevention measures, psychological consequences, and suicidal tendencies related to the COVID-19 pandemic in Bangladesh.

The values presented in the dataset are responses from a publicly available questionnaire conducted virtually in 64 districts with the assistance of the Department of Public Health and Informatics of Bangladesh, and it was collected from April 1 to April 10, 2020.

The dataset is partitioned into sets of questions organized into eight main themes:

- Sociodemographic Information.
- Knowledge about COVID-19.
- Behaviors related to COVID-19.
- Quarantine and Economic Issues.
- Fear-related Questions about COVID-19.
- Insomnia-related Questions among Participants.
- Depression-related Questions among Participants.
- Suicidal Thoughts related to COVID-19.

In total, 11,000 people responded to the questionnaire, of which 10,067 individuals qualified for meeting the three essential requirements:

- Be Bengali.
- Reside in Bangladesh.
- Be over ten years of age.

4.2 Preprocessing Step

For performing the Formal Concept Analysis, it is necessary to preprocess the database to allow the formal context construction.

In the initial step of data preprocessing, we conducted a thorough analysis and explored the available attributes to identify those irrelevant to the subject of the work or that contain empty values. The primary purpose of this stage is to reduce the dimensionality of the database to facilitate the use of the FCA approach.

To obtain the implication rules, the core element of this work, it is crucial to binarize and discretize the previously selected attributes. Some of these attributes have categorical values representing the agreement level expressed by respondents on certain questions or the likelihood of performing specific actions.

The categorical values have their ranges measured by the Likert Scale (Sullivan and Artino Jr, 2013). The original dataset has categorical questions valued by the 5-point Likert scale, where respondents show their agreement or disagreement level with the statements. The values from this range are ordinal.

Table 3 shows the range from the 5-point Likert scale.

4.2.1 COVID-19 Behavior Questions

The set of attributes related to the COVID-19 Behavior Questions (BRQ) requires the binarization process of its elements for applying the FCA method. These attributes present five possible answers to describe the likelihood of the respondent engaging in the behavior in question as in the 5-point Likert Scale. The possible responses are:

- Never.
- Seldom.
- Sometimes.
- Often.
- Almost always.

The binarization process for this set of attributes entails transforming the responses given as less likely end (never and rarely) into a negative result, while the remaining (sometimes, usually, and always) into a positive answer. Figure 1 shows this characterization.

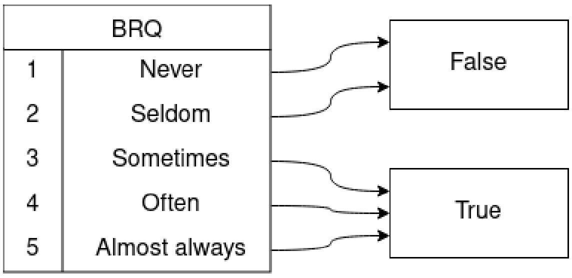

Figure 1: Preprocess steps on Behaviour Related Questions.

4.2.2 Fear of COVID-19 Scale

Another set of attributes in the database that require the binarization process are the questions related to the Fear of COVID-19 Scale (FCV-19S), which, similarly to the previously described set, also presents five possible responses to the questions.

However, these pertain to the agreement or disagreement level of the respondent regarding certain situations that could occur during the COVID-19 pandemic. The possible responses to these questions are:

- Strongly agree.
- Agree.
- Neither agree or disagree.
- Disagree.
- Strongly disagree.

In this set of attributes, the binarization process initially focused on removing the responses that contained a neutral stance on the questions posed (neither agree nor disagree). This removal facilitates the separation of the responses into two groups, where the group with a higher degree of disagreement (disagree and strongly disagree) was transformed into a negative result, while the group with a higher degree of agreement (agree and strongly agree) transformed into a positive one. Figure 2 presents this process.

4.3 Formal Context

From the stages of preprocessing the database, we obtained three formal contexts whose we study their inference rules through Formal Concept Analysis. These contexts are from the same set of 12 attributes, in which:

- 2 attributes refer to sociocultural aspects.
- 3 attributes refer to knowledge about COVID-19.

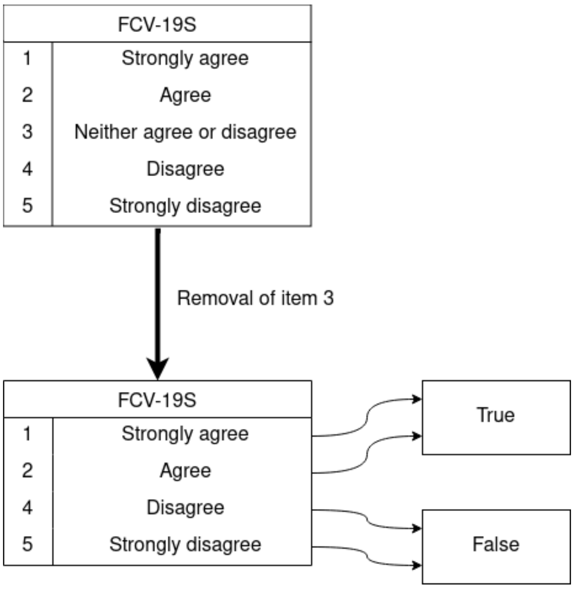

Figure 2: Preprocess steps on Fear of COVID-19 Scale.

- 4 attributes refer to behaviors adopted against COVID-19.
- 3 attributes refer to fears related to COVID-19.

The general formal context comprises all the responses obtained through the preprocessing stages, totaling 6,067 objects. The other two formal contexts are subsets of the general formal context, with the male formal context consisting of all objects identifying as male, totaling 3,351 objects.

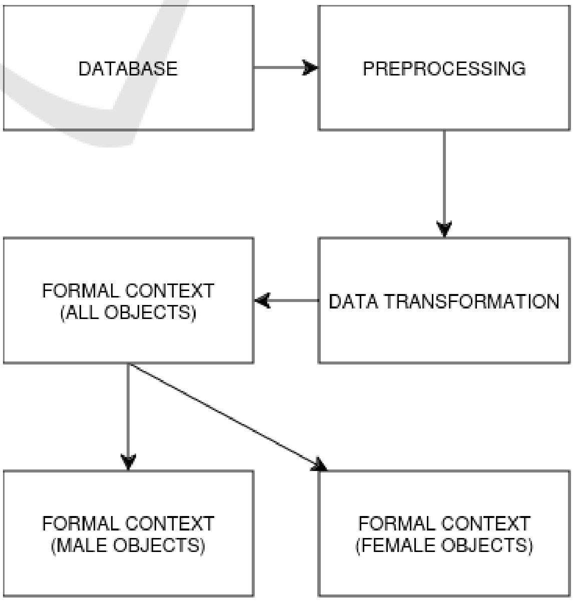

Figure 3: Preprocess steps to create the Formal Contexts.

Table 4: Part of the extracted Formal Context.

	Single	Married	I'm afraid of COVID-19	I stay in my house when not feeling well
Person 1	×		×	×
Person 2		×	×	×
Person 3	×			×
Person 4	×		×	×
Person 5	×		×	×

Table 5: Rules extracted in the Broad Formal Context.

Rule	Support	Confidence
IF I believe that COVID-19 spreads through coughs and sneezes, I believe that the mask use should be mandatory, when sneezing I cover mouth and nose, thinking about COVID-19 makes me uncomfortable, I stay in my house if I'm not feeling well and I maintain a 1 meter distance between people that are coughing or sneezing **THEN** I'll wash my hands periodically.	50%	99%
IF I believe that the mask use should be mandatory, when sneezing I cover mouth and nose, thinking about COVID-19 makes me uncomfortable, I wash my hands periodically, I am afraid of COVID-19 and I'm afraid of losing my life for COVID-19 **THEN** I'll stay in my house when I'm not feeling well.	40%	96%
IF I believe that the mask use should be mandatory, thinking about COVID-19 makes me uncomfortable, I wash my hands periodically, I maintain a 1 meter distance between people that are coughing or sneezing, I stay in my house when I'm not feeling well and I'm afraid of COVID-19 **THEN** I'll cover my nose and mouth when sneezing.	51%	98%

The female formal context comprises all objects identifying as female, totaling 2,716. Figure 3 illustrates the process for constructing the formal context, while Table 4 presents a portion of the overall formal concept.

5 RESULTS AND DISCUSSION

To analyze the most significant inference rules from the three formal contexts, we apply threshold values for support and confidence to recognize these rules.

In this work, we evaluate only rules with values exceeding 40% of support and 80% confidence.

Lattice Miner stored these rules into an XML file. We analyzed them in the form $A \longrightarrow B$, where A represents the antecedent and B represents the consequent.

In the overall scenario, represented by the formal context containing all eligible objects after the preprocessing stage of the original database, a total of 2,828 inference rules were extracted. Table 5 shows some of these rules.

From the rules presented in Table 5, it is evident that individuals who already engaged in actions against COVID-19, such as covering their nose and mouth when sneezing, maintaining social distancing of at least 1 meter from someone who is coughing or sneezing, and who feel uncomfortable or fearful about COVID-19, will regularly wash their hands. This inference rule exhibited 50% support and 99% confidence.

Another rule that can be analyzed from the data in Table 5 states that individuals who engaged in actions against COVID-19, such as regularly washing their hands and covering their nose and mouth when they are sneezing and who feel uncomfortable and afraid

Table 6: Rules extracted in the Male Formal Context.

Rule	Support	Confidence
IF I'm single and I'm afraid of COVID-19 **THEN** I'll wash my hands periodically.	52%	98%
IF I'm single, I believe that COVID-19 spreads out by coughs or sneezes, I believe that masks should be mandatory, I wash my hands periodically and I stay in my house when I'm not feeling well **THEN** I'll cover my mouth and nose when sneezing.	56%	95%

Table 7: Rules extracted in the Female Formal Context.

Rule	Support	Confidence
IF I'm single and I'm afraid of COVID-19 **THEN** I'll wash my hands periodically.	54%	98%
IF I'm single, I believe that COVID-19 spreads out by coughs or sneezes, I believe that masks should be mandatory, I wash my hands periodically and I stay in my house when I'm not feeling well **THEN** I'll cover my mouth and nose when sneezing.	52%	96%

of dying from COVID-19, will stay at home in the event of any illness. This inference rule demonstrated 40% support and 96% confidence.

Regarding the scenario that focuses exclusively on objects with the attribute of male gender, a total of 1,722 inference rules were extracted. In contrast, the scenario with only objects identifying as female gender yielded 4,054 rules. Tables 6 and 7 present some of these rules, representing objects with male and female genders, respectively.

A discrepancy in the number of inference rules extracted by the Lattice Miner is evident in the two scenarios described above.

This difference is attributed to the variation in the number of attributes in the two formal contexts since fewer objects with related attributes are needed for the rule to exceed the acceptance thresholds of support and confidence.

Upon analyzing the inference rules present in Tables 6 and 7, it is possible to observe that single males who believe that COVID-19 spreads through coughing or sneezing and engage in actions against COVID-19, such as regularly washing their hands, will cover their nose and mouth less when sneezing than females.

Support and confidence indices from these rules show this difference by comparison. For men, these indices correspond to 56% support and 95% confidence. Meanwhile, for women, 52% support and 96% confidence.

6 CONCLUSION

In this work, we applied Formal Concept Analysis to a database from Bangladesh. This dataset has attributes regarding the knowledge, preventive measures, and psychological consequences among the inhabitants of this country.

So, FCA was able to extract inference rules for studying the population's behavior regarding the COVID-19 pandemic. With this data, it was also possible to conduct a more detailed comparison of the behaviors of females and males.

From the most relevant extracted inference rules, it was possible to relate actions and psychological aspects concerning the COVID-19 pandemic with behaviors adopted by the population of Bangladesh, such as regular handwashing and the adoption of self-imposed quarantine in the event of the onset of any illness. These rules were obtained using a minimum acceptance threshold of 40% support and 80% confidence.

It is possible to identify two issues within the database used: one concerning its geographical limitation and the other regarding the imprecision of the data source.

The issue of geographical limitation arises from the fact that the data represented pertains solely to the inhabitants of Bangladesh, which restricts the generalizability of the rules found in this work.

The problem of data imprecision is related to the difficulty of characterizing more specific and individual aspects of participants due to the subjectivity and bias of the responses, as they are part of a public self-assessment questionnaire.

For future studies, I propose conducting an updated version of the public research presented in Pakpour et al. (2020), so that it will be possible to extract its inference rules and compare the most important relationships found, aiming to highlight the behavioral changes of the inhabitants of Bangladesh during the COVID-19 pandemic and in the post-pandemic period.

It would also be relevant in future works, to implement other data extraction methods, such as cluster analysis, for a comparative evaluation of their results with the inference rules obtained through Formal Concept Analysis.

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

Ahmed, A. and Rahman, M. M. (2022). Critical analysis of the possible determining factors of covid-19 epidemiological trend in bangladesh: A review. *Bangladesh Journal of Infectious Diseases*, 8(1):36–41.

Alves, A., Zarate, L., Freitas, H., and Song, M. (2023). Parallelism in the generation of concepts through the formal context object partitioning using the in-close 4 algorithm. pages 195–202.

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

Carpineto, C. and Romano, G. (2004). *Concept Data Analysis: Theory and Applications*. John Wiley & Sons Inc.

Chowdhury, N. M., Farrque, M. M. O., Hassin, C. N., and Henderson, A. (2022). Community survey on wearing mask to prevent covid-19, bangladesh, 2021. *International Journal of TROPICAL DISEASE & Health*, 43(12):9–15.

Ganter, B. and Wille, R. (1999). *Formal concept analysis : mathematical foundations*. Springer.

Jay, N., Nuemi, G., Gadreau, M., and Quantin, C. (2013). A data mining approach for grouping and analyzing trajectories of care using claim data: the example of breast cancer. *BMC Medical Informatics and decision making*, 13:1–9.

Liang, J., Zhou, J., Qian, Y., Wen, L., Bai, X., and Gao, Y. (2017). On the sampling strategy for evaluation of spectral-spatial methods in hyperspectral image classification. *IEEE*, 55(2):862–880.

Miranda, D., Zárate, L. E., and Song, M. A. (2024). Application of formal concept analysis to characterize driving behaviors and socio-cultural factors related to driving. In *BIOSTEC (2)*, pages 56–62.

Pakpour, A. H., Al Mamun, F., Hosen, I., Griffiths, M. D., and Mamun, M. A. (2020). A population-based nationwide dataset concerning the covid-19 pandemic and serious psychological consequences in bangladesh. *Data in Brief*, 33:106621.

Salahuddin, T., Haouari, F., Islam, F., Ali, R., Al-Rasbi, S., Aboueata, N., Rezk, E., and Jaoua, A. (2018). Breast cancer image classification using pattern-based hyper conceptual sampling method. *Informatics in Medicine Unlocked*, 13:176–185.

Sullivan, G. M. and Artino Jr, A. R. (2013). Analyzing and interpreting data from likert-type scales. *Journal of graduate medical education*, 5(4):541–542.

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.