Journal Healthcare Treatment Development
Vol: 03, No. 01, Jan-June 2023
ISSN: 2799-1148, DOI: 10.55529/jhtd.33.17.26

Research Paper

Analysis of COVID-19 database for defining the most important symptoms

Mohanad A. Salih[1*], Murtadha M. Hamad[2], Wesam M. Jasim[3]
[1*,2,3]Department of Computer Science, College of Computer Science and Information Technology, University of Anbar, Iraq.

Article Info

Article History:
Received: 01 February 2023
Revised: 05 April 2023
Accepted: 14 April 2023
Published: 30 May 2023

Keywords:
COVID-19 Data
Symptoms
Analysis
Diagnostic
Clustering
Data Mining

ABSTRACT

Background: In Wuhan, China, COVID-19 was found at the end of 2019. The virus quickly spread to every country in the world. In order to offer enough treatment at the appropriate time, implementing detection strategies for patient status (Negative, Positive) is one of the top issues for managers and decision-makers in modern hospitals of all sorts. Enhancing the standard of medical treatment may help to stop a COVID-19 pandemic. A summary of the health care provided to comparable individuals may be obtained by grouping patients with similar characteristics and symptoms. In most cases, algorithms are employed in medical machine learning.
Objective: This article creates an electronic questionnaire. It covers every nation in the world and is promoted online utilizing Twitter and other networking sites.
Results: In the course of getting ready to prepare COVID-19 data, the raw COVID-19 data including non-applicable occurrences is handled after the feedback interpretation and analysis process is complete, after the questionnaires were distributed, and after all the findings have been collated.
Conclusions: The patient state is predicted using these traits and symptoms, either positively or negatively.

Corresponding Author:
Mohanad A. Salih
Department of Computer Science, College of Computer Science and Information Technology, University of Anbar, Iraq.
Email: mohanad.a.alhadethy@gmail.com

1. INTRODUCTION

COVID-19 is a hurricane to the majority of the nations in the world. The globe is working together to combat this pandemic. Sneezing, coughing, chatting, and touch all contribute to the transmission of COVID-19. Therefore, it poses a risk for COVID-19 to spread quickly. A nation's capacity to fight the epidemic could be enhanced by higher-quality healthcare. The situation is made worse by the fact that COVID-19 may live for up to 72 hours in environments that seem normal [1].

The best way to increase the COVID-19 survival rate is early discovery, and the work currently being done can benefit from the use of medical data mining technologies. Medical machine learning is frequently used in the medical field. If patterns of patient data are acquired, it is generally accepted that medical machine learning may address the urgent issues related to COVID-19 [2]. Clinicians are assisted in the development of trustworthy diagnostic methods and processes by medical machine learning. Both medical supervised and unsupervised learning are frequently employed in the fields of data science and medical machine learning to address a variety of real-world issues [3]. Clustering patients with comparable COVID-19 symptoms aids in gaining insight into the standard of care given to a range of individuals.

[4] Unsupervised machine learning techniques are being used for clustering. Unsupervised medical machine learning algorithms frequently distinguish between data created using vision and unlabeled data contained in the data collection. [5], [6] This work provides a real dataset based on patient data to assist researchers in predicting cases of COVID-19 possibilities (negative, positive) using unsupervised medical machine learning techniques.

2. RELATED WORK

Only two recent studies include the COVID-19 dataset, which was developed to forecast and examine COVID-19 instances. G. Manimannan, & Co., A. Pompaavai, 2019, Up till March 24, 2020, using secondary sources of data from the Indian Health and Family Welfare Organization, it is attempted in this study to determine the Indian states and Union Territories that have been impacted by COVID-19 [7].The categorization of COVID-19 cases and fatalities in Southeast Asia is discussed in this essay. The data was then organized into several clusters using Rapid Miner tools. Data from the WHO for 2020 fatalities in April, national statistics, and COVID-19 confirmed cases have all been utilized [8].

3. METHODOLOGY

In this research a questionnaire electronic is created. It includes all countries of the world and promotes it via the Internet using social networking sites and Twitter. According to Table 2, the symptoms and traits that were questioned for the training dataset were restricted to personal characteristics and several laboratory test variables. The patient state is predicted using these traits and symptoms, either positively or negatively. In order to get comprehensive patterns about them, the questionnaire was filled out by 3000 patients from various countries. Iraq's Al-Anbar University website hosted the database. You may get it online by clicking this link: https://github.com/mohanadalhadethy/covid-19-database.

Once all of the data was acquired and the questionnaires were delivered, the return analysis and interpretation procedure was complete. The method used to prepare COVID-19 data involves processing the raw COVID-19 data, which includes events that aren't relevant. The pre-cleaning and data processing approaches were used to clean the data set. This was brought on by contradictions and mistakes that needed to be fixed. Data from the COVID-19 was imported into Excel or a csv file for analysis and alterations. During the pre-processing and cleaning of the data, Outlier and missing values are removed from the dataset. Data generalization is therefore considered to be one of the methods for minimizing data.

Table 1. Lists the Patients' Characteristics and Symptoms Questionnaire

Question NO.	Description	Question NO.	Description

Q1	the patient's age	Q12	Does the Patient Experience Fatigue? (Yes, No)
Q2	Patient's Sexuality	Q13	Does the Patient Have Sputum? (Yes, No)
Q3	Administration of the Sick	Q14	Does the Patient Experience Chest Pain? (Yes, No)
Q4	Nationality of the Patient	Q15	Does the Patient Have Diarrhea? (Yes, No)
Q5	Body temperature of the patient (High, Medium, Low)	Q16	Does the Patient Have Hypertension? (Yes, No)
Q6	The patient may be coughing. (Yes, No)	Q17	Does the Patient Have Expectoration? (Yes, No)
Q7	Is the patient experiencing a headache? (Yes, No)	Q18	Does the Patient Experience Myalgia? (Yes, No)
Q8	Does the Patient Experience Loss of Smell? (Yes, No)	Q19	Does the Patient Experience Gasping? (Yes, No)
Q9	Does the Patient Experience a Loss of Taste? (Yes, No)	Q20	Does the Patient Experience Drowsiness? (Yes, No)
Q10	Does the patient have any recurring illnesses? (Yes, No)	Q21	Anorexia: Does the Patient Have It? (Yes, No)
Q11	Does the Patient Suffer from Vertigo? (Yes, No)	Q22	Patient COVID-19 Status (Positive, Negative) Target Class

4. RESULTS AND DISCUSSION

The results of the questionnaire showed that a number of features are very similar to the symptoms of seasonal flu. In addition, the symptoms differ from one person to another (for example, the temperature of many people with covid-19 did not feel a rise in temperature, unlike other patients who had a high temperature) as well. The results of the questionnaire proved that only 4 out of 22 features included in the questionnaire were the main symptoms by which the patient could be diagnosed, is the result positive or negative, namely (headache by 51%, sense of smell by 52%, fatigue by 76%, loss of appetite by a percentage 56%,). The most susceptible ages are those between 25-35 years, due to the frequent mixing of these ages with others through work. The questionnaire also proved that the death rate from covid-19 disease was 9%, especially for those who suffer from diseases Chronic (heart, kidney, diabetes). As shown in Figure 1.

Table 2. Agricultural Field Dataset Containing Crop, Soil, Spraying, and Pest Variables

Sr. No.	Date	Farm Name	Crop Type	Soil Type	Irrigation Material	Irrigation Method	Spray Type	Nozzles /Plot	Rodents in each field	Stops /Plot	Eggs /Plot
1	2021/07/03 14:25	Abu Ghreeb Farm	Wheat	Clay	Without material	Without spray	Without	7	3	2	1
2	2021/07/03 12:54	Abu Ghreeb Farm	Wheat	Clay	Insecticide	Manual	Without	3	1	1	2
3	2021/07/03 12:53	Abu Suhayb Farm	Barley	Sandy	Insecticide	Without spray	Without	2	2	2	3

4	2021/07/03 12:37	Abu Ghreeb Farm	Barley	Clay	Without material	Manual	Without	5	2	1	4
5	2021/07/03 13:42	Al-Hamdaniya Farm	Barley	Sandy	Without material	Without spray	Without	7	1	1	5
6	2021/07/03 21:26		Barley	Clay	Without material	Without spray	Without	7	3	3	6
7	2021/07/03 2:27	Abu Ghreeb Farm	Barley	Clay	Insecticide	Surface spray	Chemical	6	2	1	7
8	2021/07/03 2:48	Abu Ghreeb Farm	Barley	Sandy	Without material	Without spray	Without	6	2	2	8
9	2021/07/03 2:48	Abu Ghreeb Farm	Barley	Clay	Insecticide	Surface spray	Chemical	6	2	1	9
10	2021/07/03 13:04	Abu Ghreeb Farm	Barley	Clay	Insecticide	Surface spray	Chemical	6	1	2	10
11	2021/07/03 13:04	Abu Ghreeb Farm	Barley	Clay	Insecticide	Surface spray	Chemical	6	2	2	11
12	2021/07/03 13:43	Abu Ghreeb Farm	Barley	Clay	Insecticide	Surface spray	Chemical	7	3	3	12
13	2021/07/03 13:51	Al-Hamdaniya Farm	Barley	Sandy	Insecticide	Surface spray	Chemical	6	1	1	13
14	2021/07/03 13:51	Al-Hamdaniya Farm	Barley	Sandy	Insecticide	Surface spray	Chemical	4	2	2	14
15	2021/07/03 4:09	Al-Hamdaniya Farm	Barley	Sandy	Insecticide	Surface spray	Chemical	6	3	3	15
16	2021/07/03 4.48	Al-Hamdaniya Farm	Barley	Sandy	Insecticide	Surface spray	Chemical	6	2	2	16
17	2021/07/03 5:01	Al-Hamdaniya Farm	Barley	Clay	Insecticide	Surface spray	Chemical	6	3	3	17
18	2021/07/03 5:28	Al-Hamdaniya Farm	Barley	Clay	Insecticide	Surface spray	Chemical	7	2	2	18
19	2021/07/03 5:15	Al-Hamdaniya Farm	Barley	Sandy	Insecticide	Surface spray	Chemical	6	2	2	19
20	2021/07/03 16:42	Al-Hamdaniya Farm	Barley	Clay	Insecticide	Surface spray	Chemical	7	4	4	20

21	2021/07/03 0:14	Abu Ghreeb Farm	Barley	Sandy	Insecticide	Surface spray	Chemical	6	4	2	21
22	2021/07/03 6:30	Abu Ghreeb Farm	Barley	Sandy	Insecticide	Surface spray	Chemical	7	5	5	22
23	2021/07/03 6:44	Al-Hamdaniya Farm	Barley	Clay	Insecticide	Surface spray	Chemical	6	2	3	23

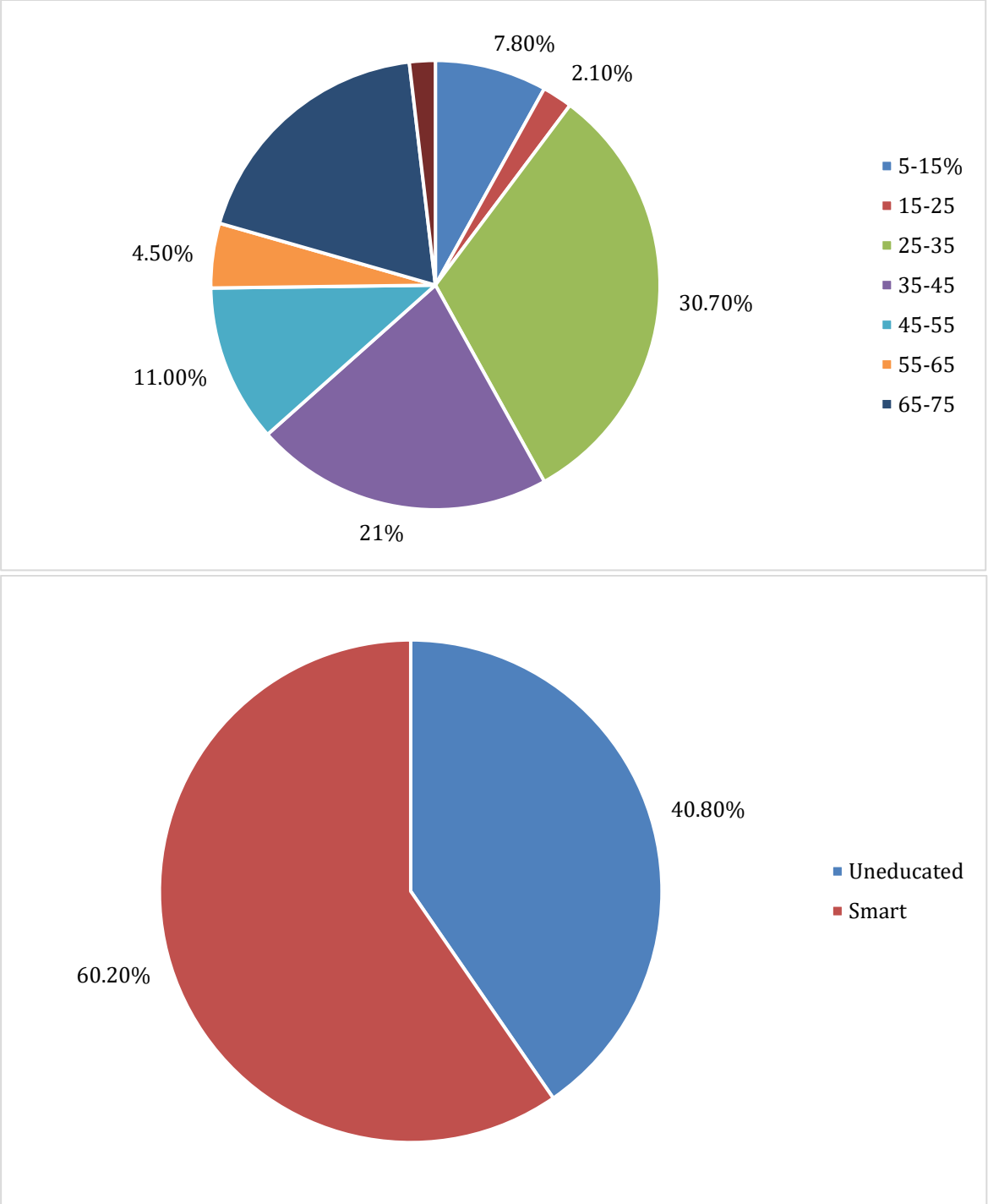


Figure 1. Distribution of Respondents by Age and Gender

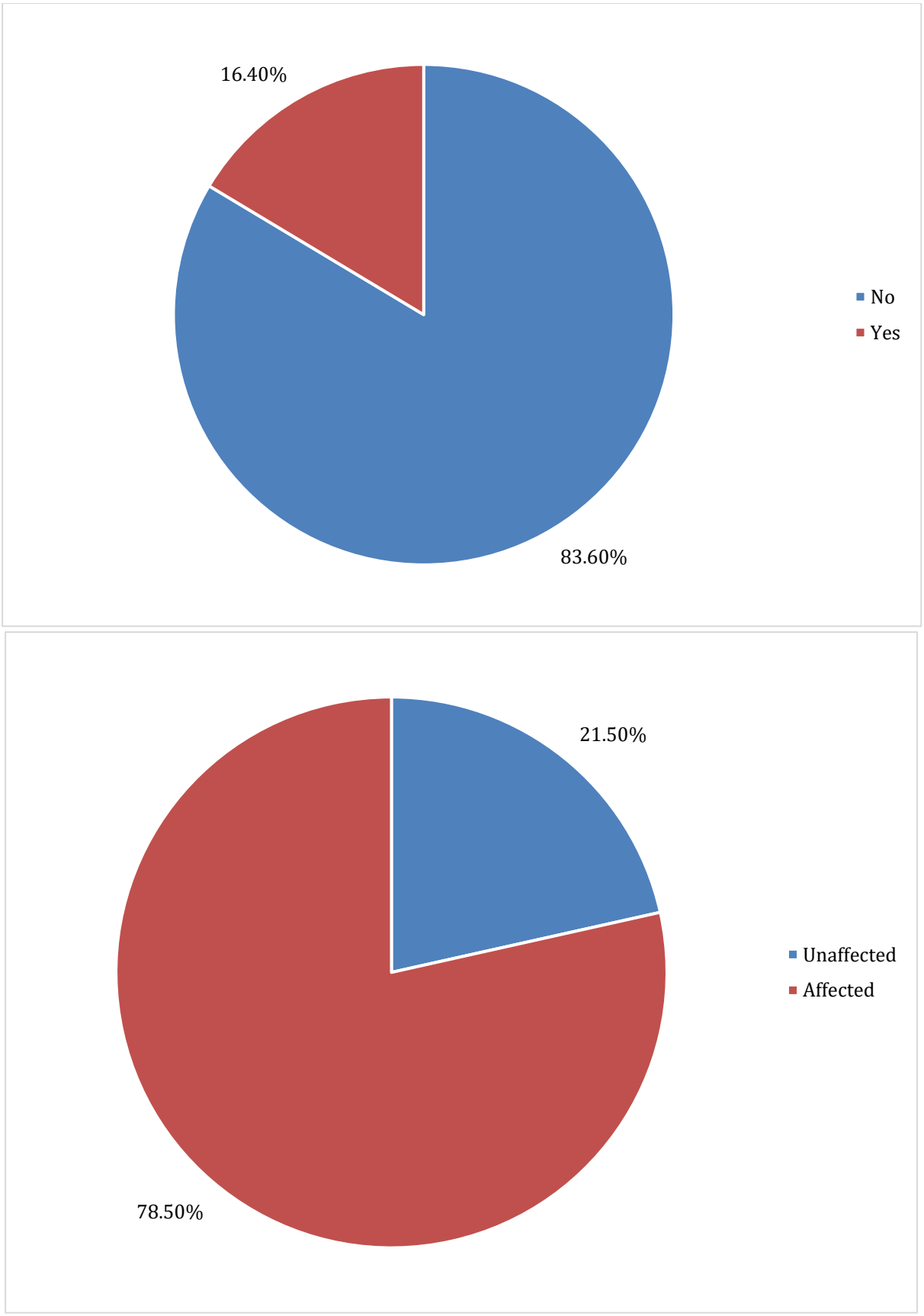


Figure 2. Survey Results on Eye Swelling and Condition

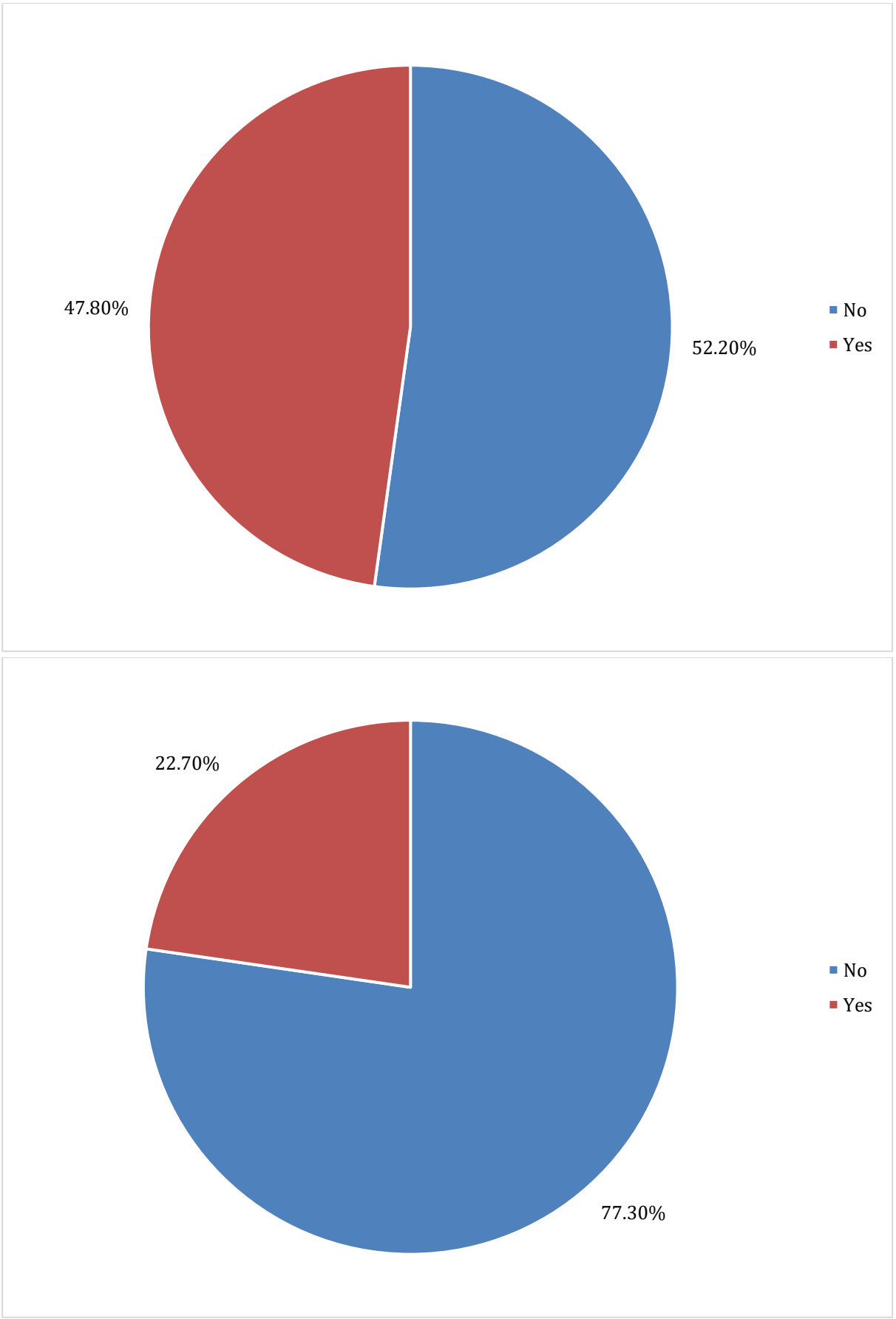


Figure 3. Distribution of Patients with Sore Throat and Fever Symptom

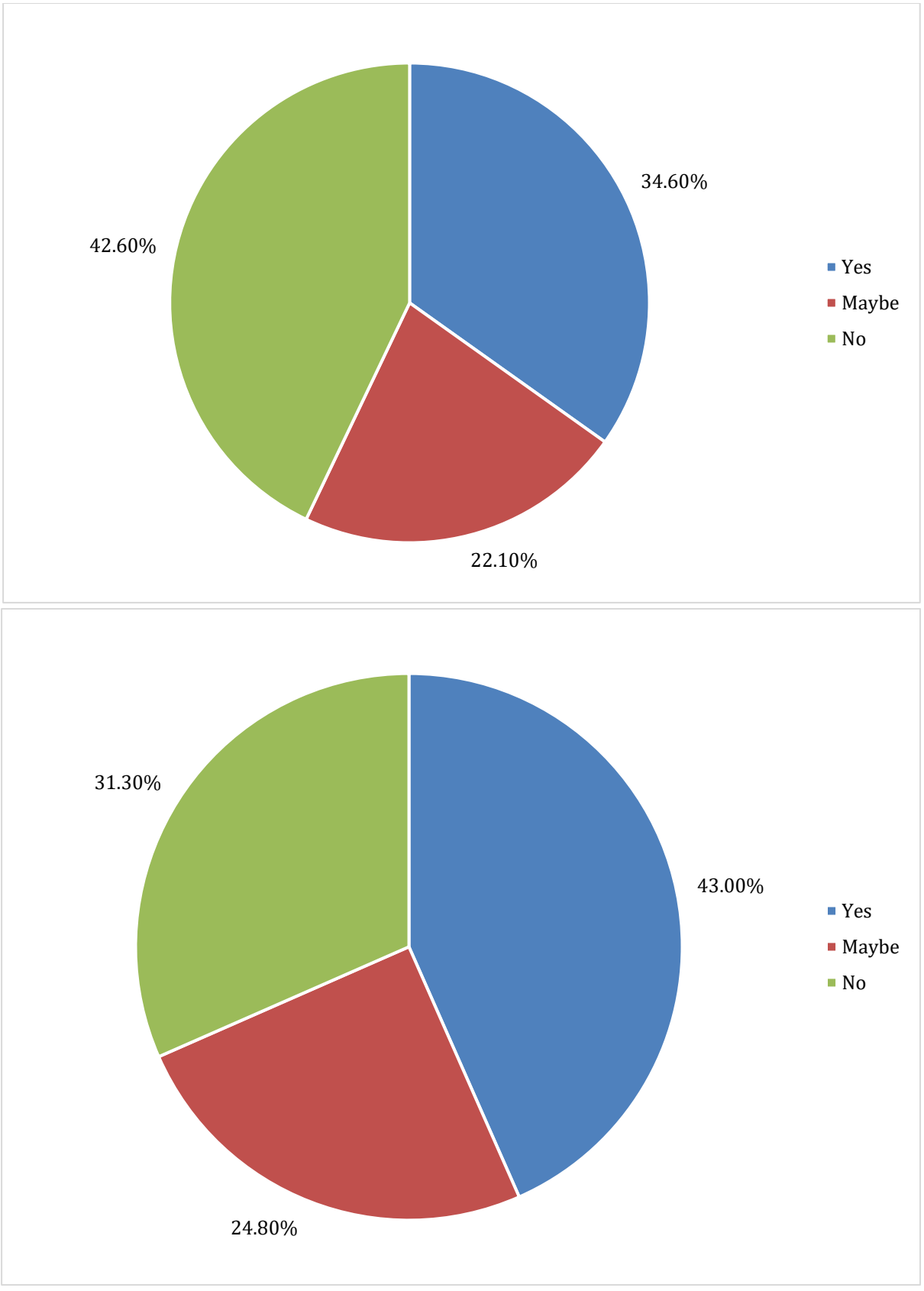


Figure 4. Distribution of Patients with Shortness of Breath and Runny Nose Symptom

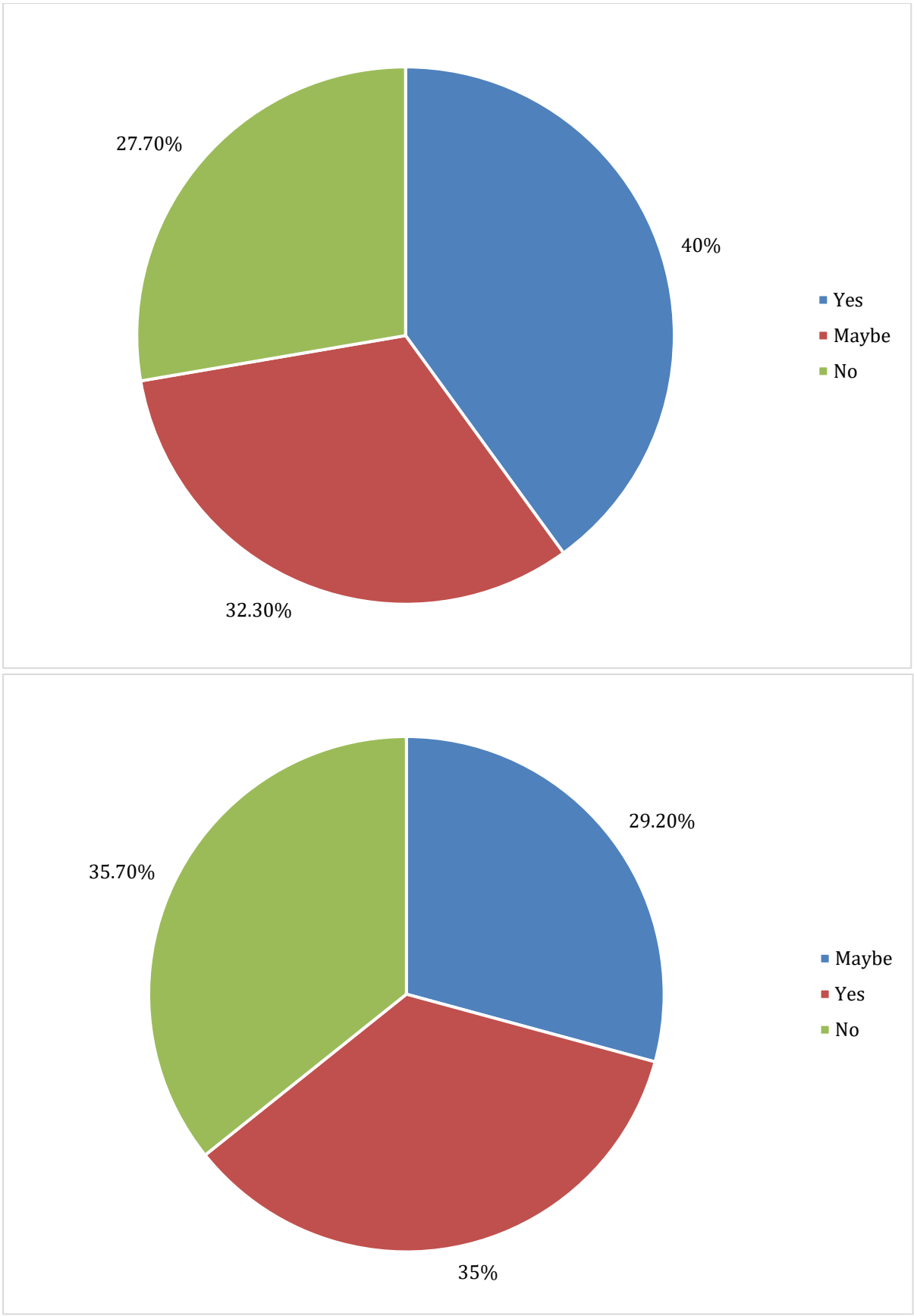


Figure 5. Distribution of Patients with Chills and Common Cold Symptoms

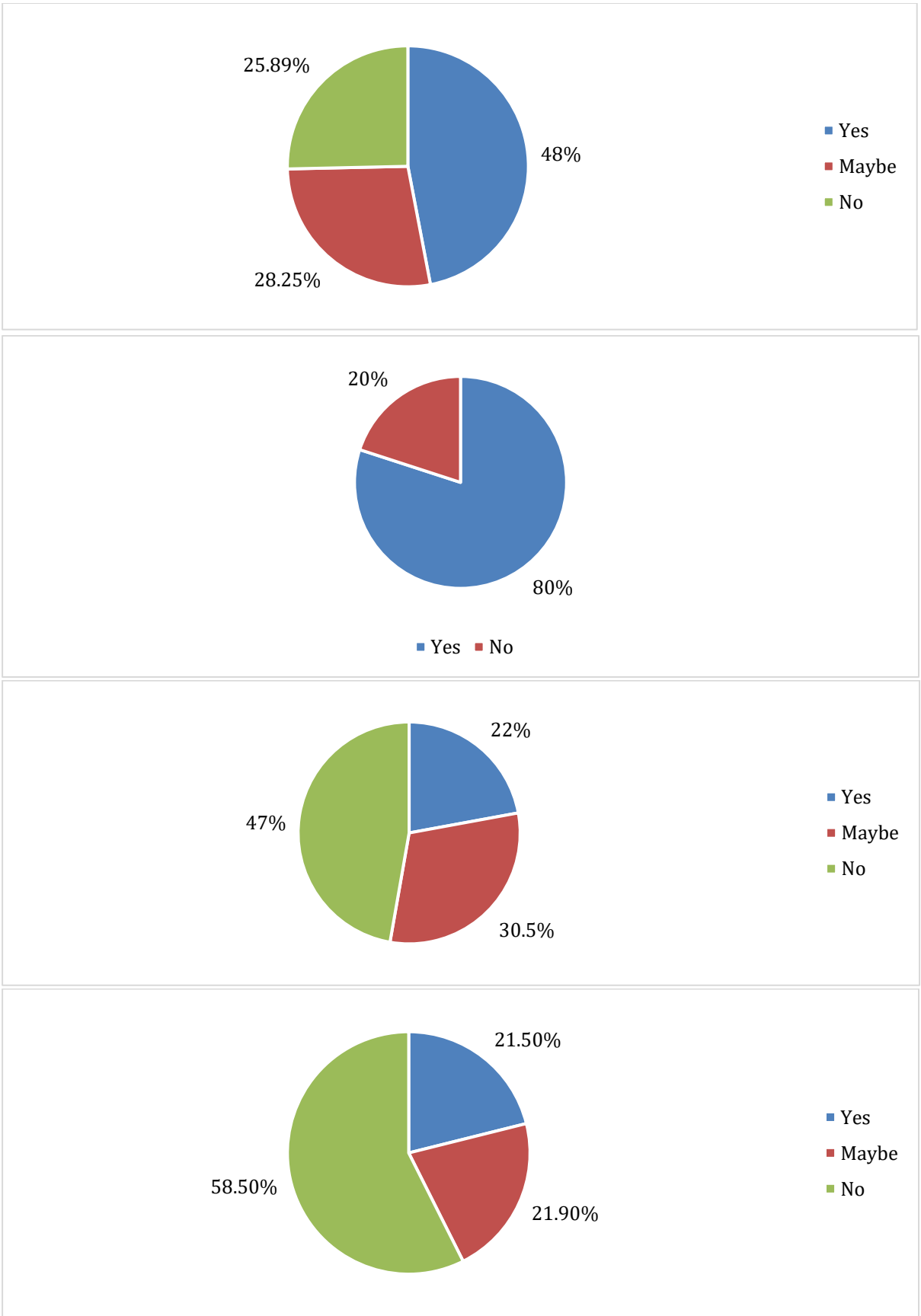


Figure 6. Distribution of Patients with Diarrhea, Hypertension, Dyspnea, and Loss of Appetite Symptom

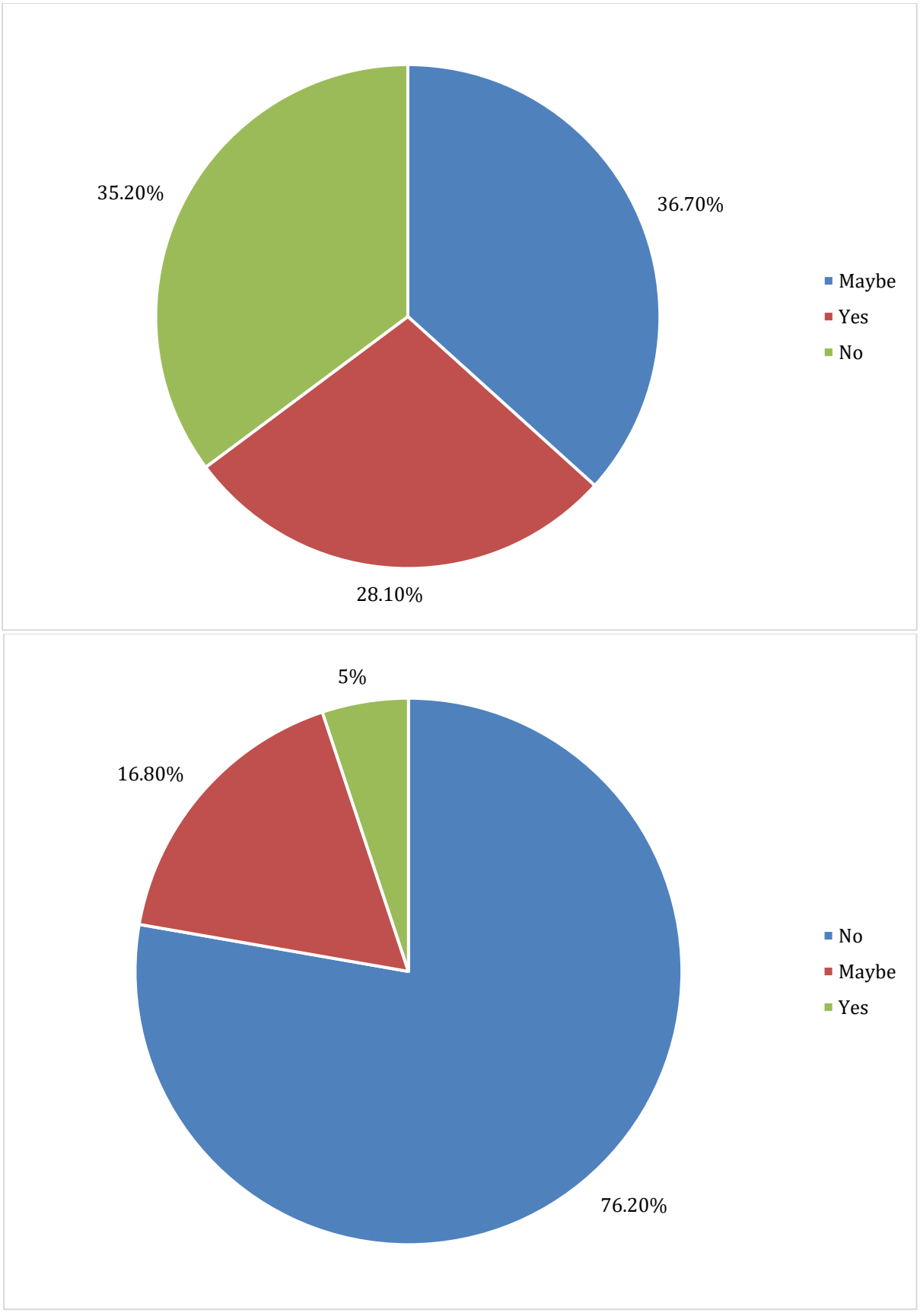


Figure 7. Distribution of Patients with Dizziness and Fatigue Symptoms

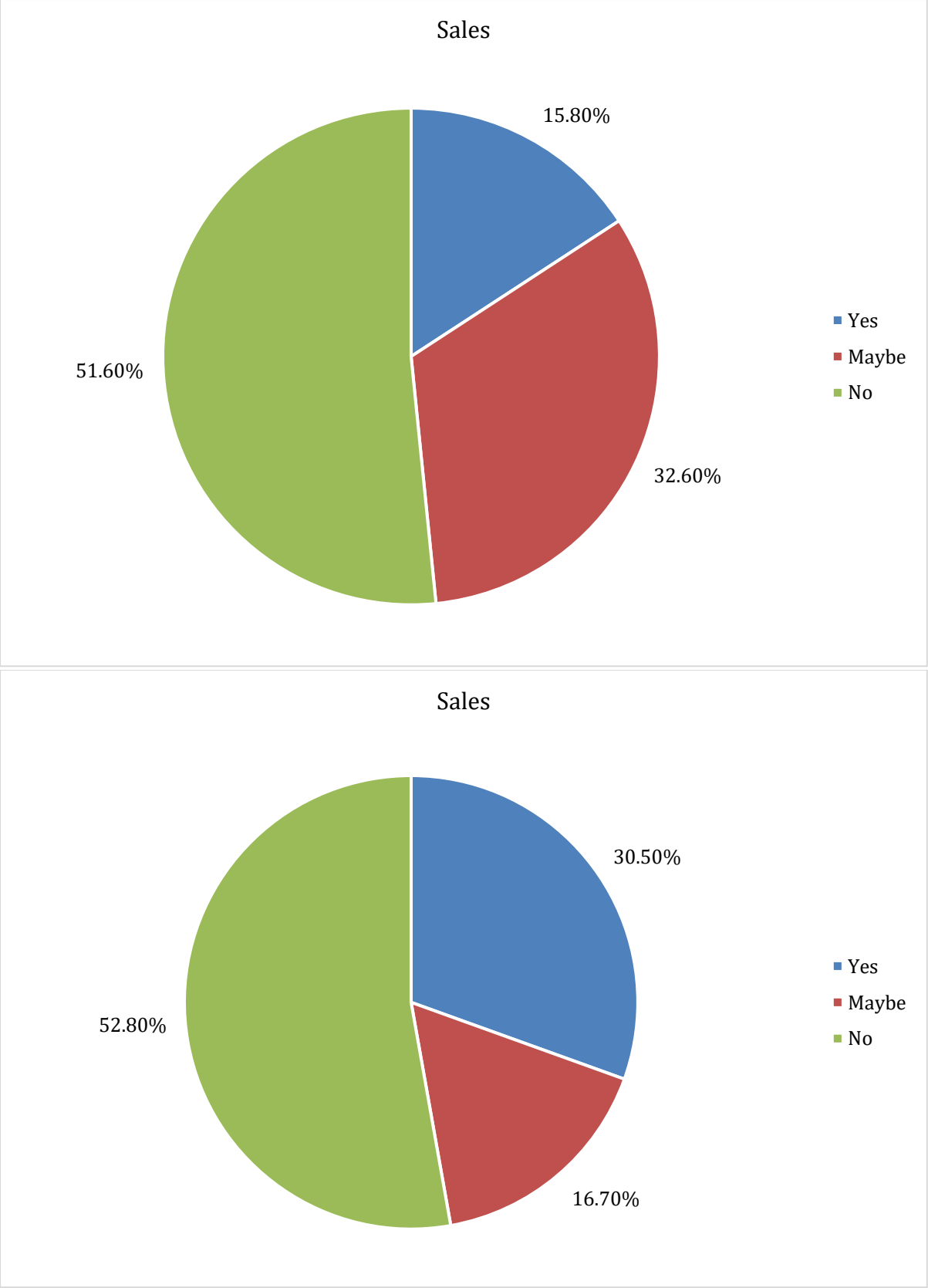


Figure 8. Distribution of Survey Responses

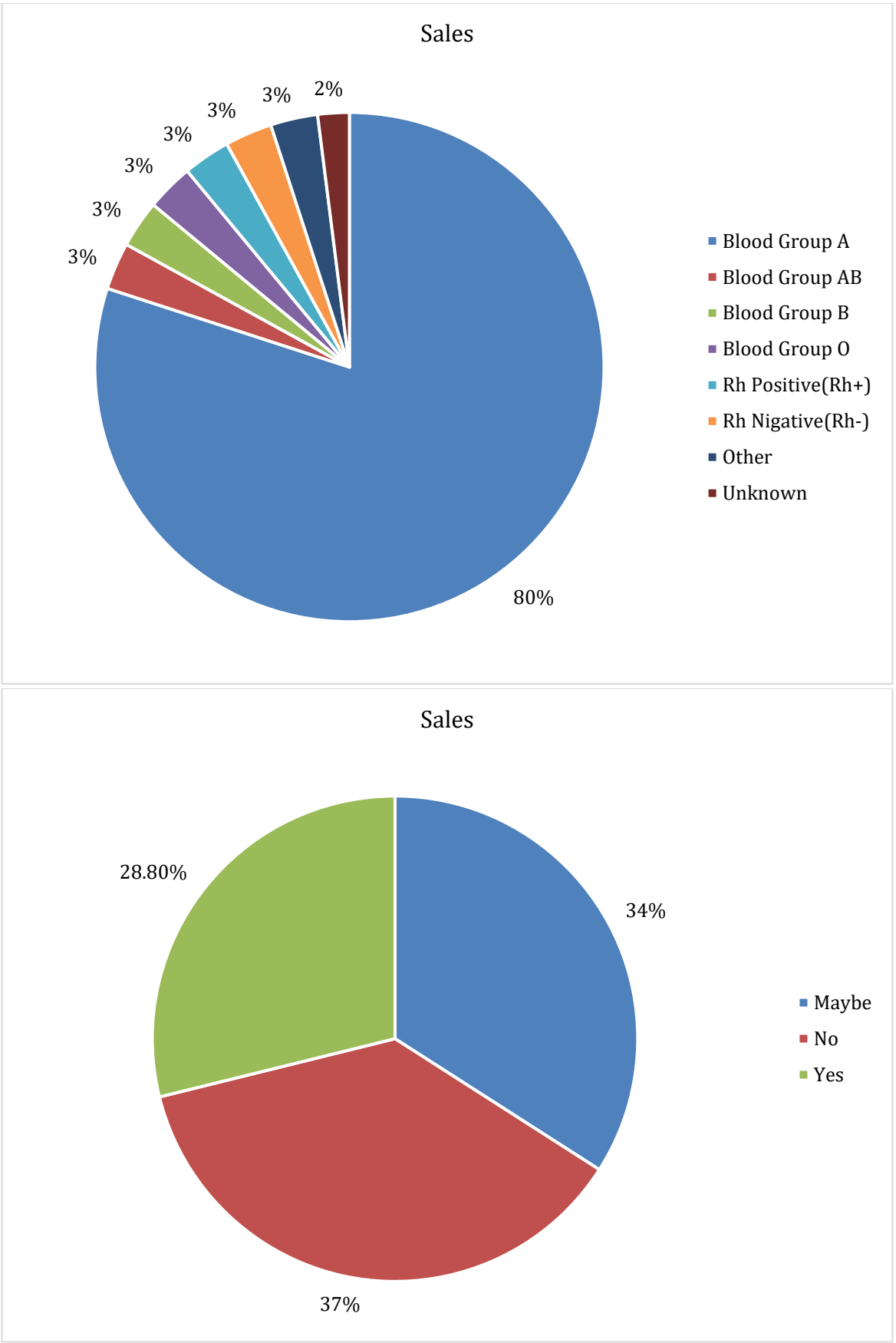
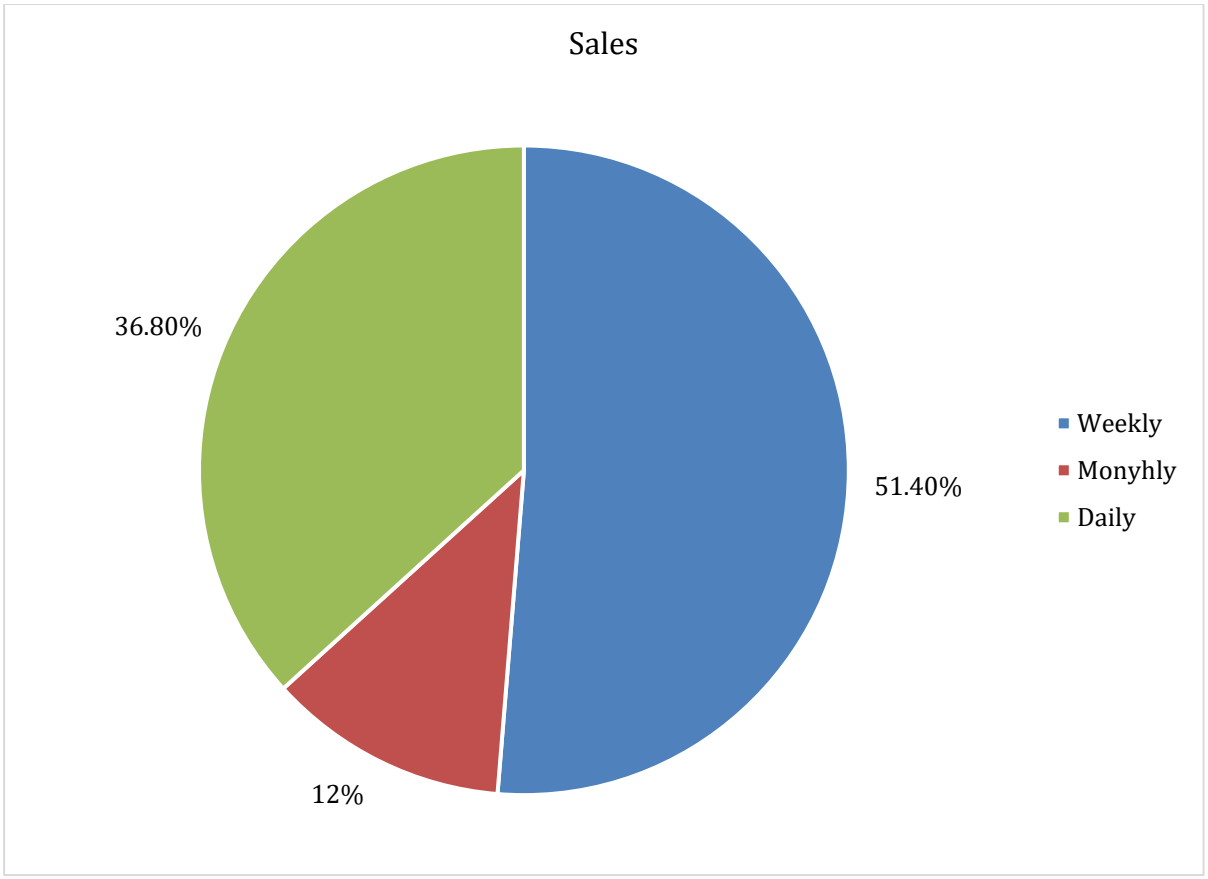


Figure 9. Demographic Characteristics and Patient Satisfaction of Survey Respondents



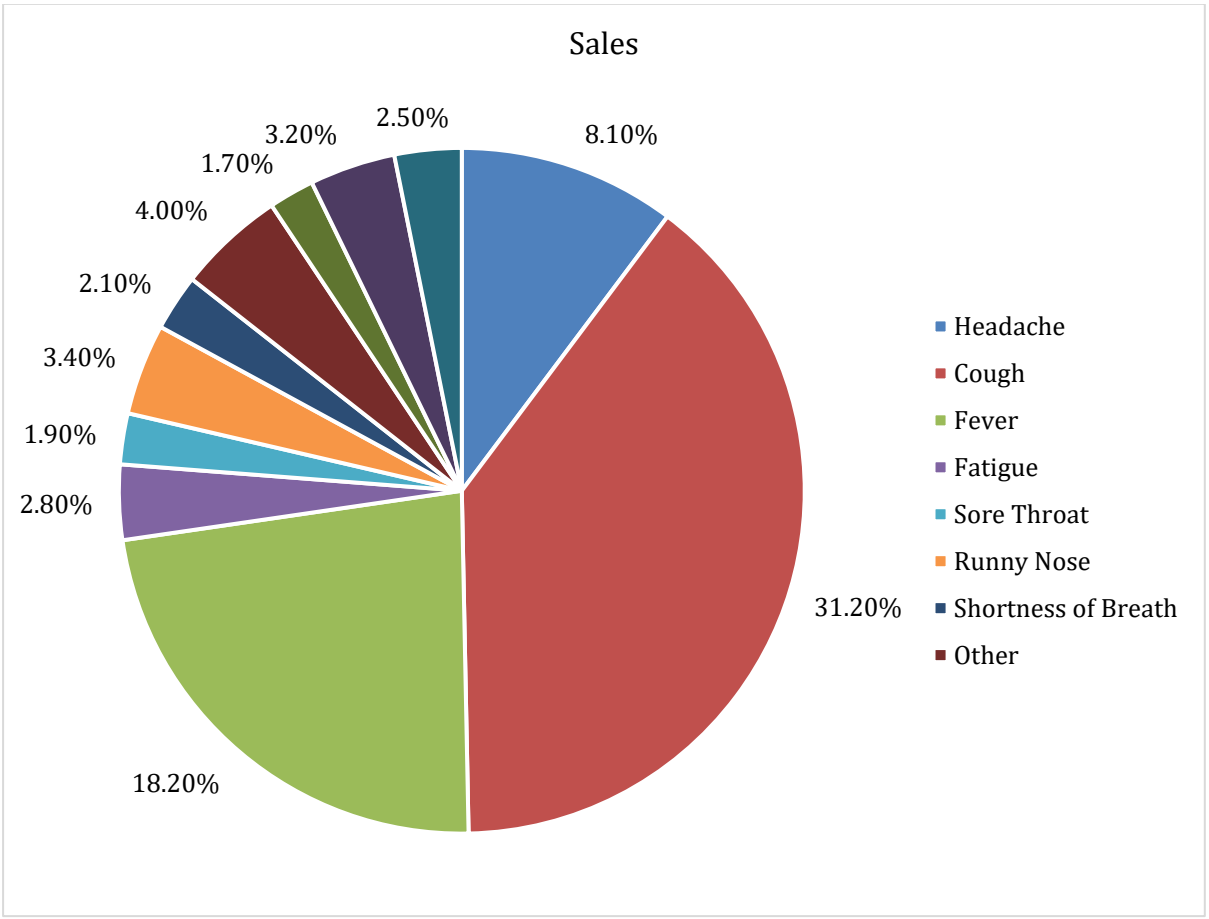


Figure 10. Patient Satisfaction Levels and Healthcare Department Utilization

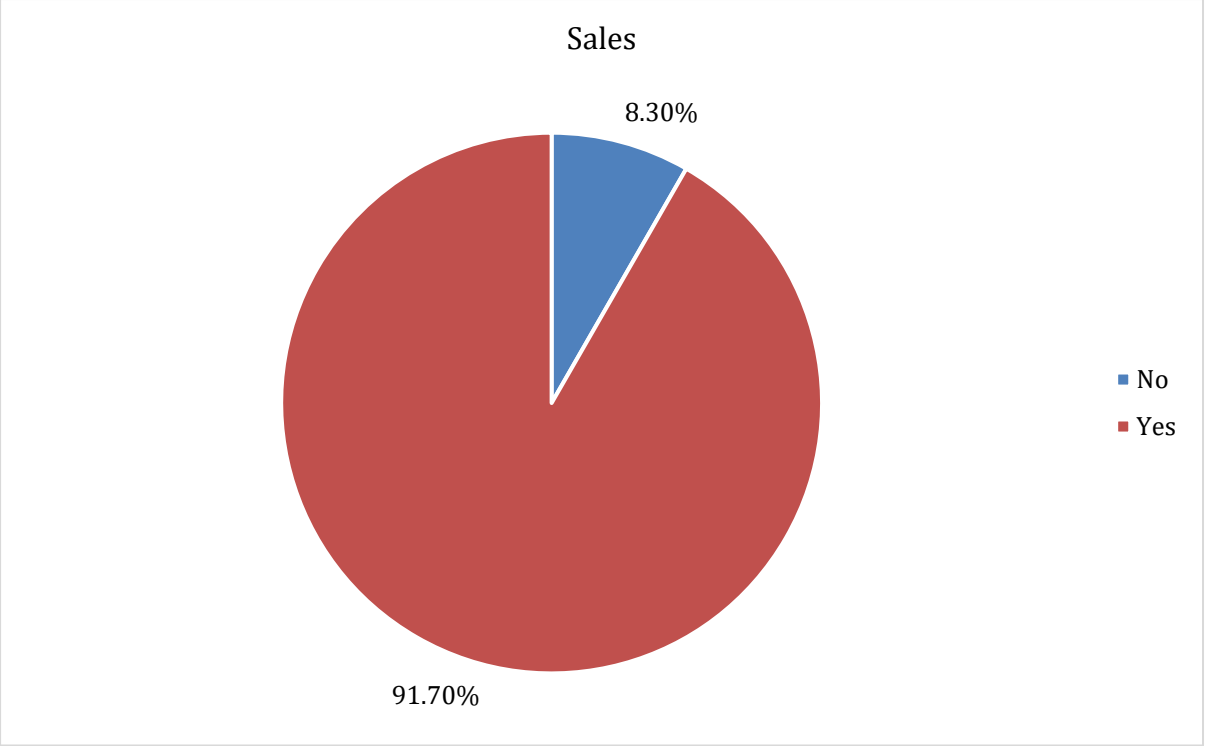


Figure 11. Statistical Percentages of Symptoms for Covid-19 Disease

5. CONCLUSION

In this paper, a questionnaire electronic is created. It includes all countries of the world and promotes it via the Internet using social networking sites and Twitter. Once the feedback interpretation and analysis process has been finished once the questionnaires have been sent out and all the results have been compiled, the raw COVID-19 data containing non-applicable occurrences is processed as part of the COVID-19 data processing method. The results of the questionnaire proved that only 4 out of 22 features included in the questionnaire were the main symptoms by which the patient could be diagnosed, is the result positive or negative, namely (headache by 51%, sense of smell by 52%, fatigue by 76%, loss of appetite by a percentage 56%,). The patient state is predicted using these traits and symptoms, either positively or negatively.

Acknowledgments

The authors have no specific acknowledgments to make for this research.

Funding Information

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Author Contributions Statement

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Mohanad A. Salih	✓	✓	✓	✓	✓	✓		✓		✓	✓	✓	✓	✓
Murtadha M. Hamad	✓		✓		✓	✓	✓		✓	✓	✓	✓		✓
Wesam M. Jasim	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration

Va : **Va**lidation
Fo : **Fo**rmal analysis
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Fu : **Fu**nding acquisition

Conflict of Interest Statement

The authors declare that there are no conflicts of interest regarding the publication of this research.

Informed Consent

All participants were informed about the purpose of the study, and their voluntary consent was obtained prior to data collection.

Ethical Approval

Not Applicable.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] J. F. S. Soekiman, T. D. Putranto, D. Susilo, and E. M. A. Garcia, 'Economic Sector during the COVID-19 Pandemic: Indonesian Instagram Users Behaviour', Webology, vol. 18, no. 1, 2021. doi.org/10.14704/WEB/V18I1/WEB18081

[2] S. Gupta, M. Kaur, S. Lakra, and Y. Dixit, 'A Comparative Theoretical and Empirical Analysis of Machine Learning Algorithms', Webology, no. 1, 2020. doi.org/10.14704/WEB/V17I1/WEB17011

[3] Polignano, M., Suriano, V., Lops, P., de Gemmis, M., & Semeraro, G. (2020). A study of Machine Learning models for Clinical Coding of Medical Reports at CodiEsp 2020. In Working Notes of Conference and Labs of the Evaluation (CLEF) Forum. CEUR Workshop Proceedings.

[4] M. Kato et al., 'Clustering patients by depression symptoms to predict venlafaxine ER antidepressant efficacy: Individual patient data analysis', Journal of Psychiatric Research, vol. 129, pp. 160-167, 2020. doi.org/10.1016/j.jpsychires.2020.06.011

[5] M. Alloghani, D. Al-Jumeily, J. Mustafina, A. Hussain, and A. J. Aljaaf, 'A systematic review on supervised and unsupervised machine learning algorithms for data science. Supervised and Unsupervised Learning for Data Science', pp. 3-21, 2020. doi.org/10.1007/978-3-030-22475-2_1

[6] Korach, Z. T., Cato, K. D., Collins, S. A., Kang, M. J., Knaplund, C., Dykes, P. C., & Zhou, L. (2019). Unsupervised machine learning of topics documented by nurses about hospitalized patients prior to a rapid-response event, 952-963. doi.org/10.1055/s-0039-3401814

[7] A. Poompaavai and G. Manimannan, 'Clustering study of Indian states and union territories affected by coronavirus (COVID-19) using k-means algorithm', International Journal of Data Mining And Emerging Technologies, vol. 9, no. 2, pp. 43-51, 2019. doi.org/10.5958/2249-3220.2019.00006.5

[8] J. Hutagalung, N. L. W. S. R. Ginantra, G. W. Bhawika, W. G. S. Parwita, A. Wanto, and P. D. Panjaitan, 'COVID-19 Cases and Deaths in Southeast Asia Clustering using K-Mean Algorithm', in Journal of Physics: Conference Series, vol. 1783, IOP Publishing, 2021. doi.org/10.1088/1742-6596/1783/1/012027

How to Cite: Mohanad A. Salih, Murtadha M. Hamad, Wesam M. Jasim. (2023). Analysis of COVID-19 database for defining the most important symptoms. Journal Healthcare Treatment Development (JHTD), 3(1), 103–119. https://doi.org/10.55529/jhtd.33.17.26

BIOGRAPHYS OF AUTHORS

	**Mohanad A. Salih,** He got his B.Sc. and then M.Sc. degrees in Computer Science from the University of Anbar, Iraq. Right now, he is a faculty member in the Department of Computer Science, College of Computer Science and Information Technology, University of Anbar. His research interests sit around artificial intelligence, machine learning, and data mining, not just one. He has contributed to multiple national and international research publications and he also keeps showing up at academic conferences. He acts as a corresponding author and also a research supervisor for postgraduate students at the University of Anbar. Email: mohanad.a.alhadethy@gmail.com
	**Murtadha M. Hamad,** his B.Sc. in Computer Science from the University of Anbar Iraq, and later completed his M.Sc. in Computer Science from the same place. He is currently working as a faculty member in the Department of Computer Science, College of Computer Science and Information Technology, University of Anbar. His main attention is on software engineering, intelligent systems and computer networks, in general. He has taken part in different collaborative research efforts and has published papers in peer reviewed journals, and at international conferences too, within computing and information technology**.** Email: dr.mortadha61@gmail.com
	**Wesam M. Jasim,** He received his B.Sc and M.Sc. degrees in Computer Science from the University of Anbar, Iraq. He is sort of affiliated with the Department of Computer Science, College of Computer Science and Information Technology, University of Anbar, yes. His research interests range from computer vision, image processing, and deep learning applications. He has actively contributed to research in intelligent computing systems, and also co-authored some publications in reputable journals and conference proceedings. He seems to be committed to advancing computing education, and applied research inside the academic environment of the University of Anbar. Email: co.wesam.jasim@uoanbar.edu.iq