Machine Learning-based Classification of COVID-19 Preventive Behaviors Among University Staff and Students in Chiang Mai, Thailand

Thitikan Phuwitthanasap[1,2], Khanita Duangchaemkarn[3], Kitbordin Thongduang[4], Wanicha Pungchompoo[5], Waraporn Boonchieng[2]*

[1] Office of Research Administration, Chiang Mai University, Thailand
[2] Faculty of Public Health, Chiang Mai University, Thailand
[3] Department of Pharmacy Practice, University of Phayao, Thailand
[4] Faculty of Public Health, Naresuan University, Thailand
[5] Faculty of Nursing, Chiang Mai University, Thailand
Thitikan.p@cmu.ac.th, kkhanita.du@up.ac.th, kitbordint@nu.ac.th, wanicha.p@cmu.ac.th, waraporn.b@cmu.ac.th

Abstract

The COVID-19 pandemic has exposed public health management to new challenges. Most importantly, there is a need to assess risk and deploy limited resources quickly in response to emerging disease outbreaks. Previous research has shown that self-reported survey data can accurately predict infection risks. However, even this can be slow and cumbersome to implement. Herein, we explored a new approach. We collected risk assessment survey data from 1,266 members of the Chiang Mai University community. We found that students were generally at a higher risk of developing COVID-19 than faculty members. However, we were interested in knowing whether this difference was due to heterogeneous risk profiles or whether students and faculty exhibited systematic risk profiles that were largely homogeneous within each group. To assess this, we trained machine learning models to classify participants as students or faculty group members. This model achieved an accuracy of 95% in a test data set, confirming its ability to categorize group membership and suggesting that risk profiles must have been relatively homogenous within each group. This result suggests that public health decision makers can confidently make decisions to deploy resources to help large societal groups based on survey data collected from relatively small samples.

Keywords: Predictive analysis, COVID-19 risk, Machine learning, Coronavirus, Hygiene

1 Introduction

The COVID-19 pandemic has presented new challenges for the effective management of public health. In order to prevent the spread of disease, there are many basic preventative measures that can be taken (e.g., regular hand washing and wearing a face mask in public) that have been proven effective. There are also a number of risk behaviors (e.g., travel, indoor socializing) that have been found to increase disease transmission. Screening potential patients for these risk behaviors has led to more accurate predictions of positivity for Sars-Cov-2, the virus which causes COVID-19 [1-4].

The ability to categorize individuals into low or high-risk groups based on easily-collected survey data has made it possible to use limited healthcare resources in an efficient manner. For example, PCR testing can be targeted among members of high-risk groups to find asymptomatic cases even when testing resources are limited [1]. Risk assessment can also help assess compliance with public health policies and guidelines [4].

However, even relatively simple survey data can be difficult to collect from large populations in a timely manner. When disease outbreaks strike, it is critical that resources be directed effectively and quickly without the need to wait for large data sets to be collected. The application of information systems and modern data analytics for public health data management can play a vital role in informing fast and accurate decision making [5].

One question would be whether individuals within societal groups have consistent risk profiles. If this is the case, then it would allow public health workers to assign risk levels to large segments of the population based on relatively small samples from different groups. For example, do students have a uniform risk profile? If they do, then public health measures concerning students could be confidently decided without the need for exhaustive population sampling.

To answer this question, we have developed a COVID-19 related risk assessment method of surveying a large number of people in the Chiang Mai University (CMU) community. The survey assessed individual adherence to COVID-19 prevention measures as well as their engagement in COVID-19 risk behaviors. In this study, we have proposed a machine learning-based classification model using the collected data to promptly classify the risk profile for individuals.

*Corresponding Author: Waraporn Boonchieng; Email: waraporn.b@cmu.ac.th
DOI: https://doi.org/10.70003/160792642025122607007

2 Materials and Methods

2.1 Questionnaire Building

The behavior-related questionnaire for COVID-19 exposed risk was constructed based on the basic knowledge pertaining to personal hygienic control as is shown in Table 1. The results for the designed questionnaires were scored based on Y/N type answers. Risk levels were categorized into three levels based on the total scores: 0-13 as 'low risk', 14-27 as 'moderate risk', and 28-41 as 'high risk'. The questionnaires were evaluated using the content validity index and Cronbach's alpha coefficient, which yielded results of 0.96, and 0.83, respectively.

Table 1. List of questionnaire items used for data gathering and machine-learning model building

Personal hygienic related risks	
1)	Contact with respiratory droplets within the past 14 days
2)	Have participated in community activities or gatherings within the past 14 days
3)	Have been in enclosed spaces, e.g., shopping malls, cinemas, classrooms, or auditoriums, within the past 14 days
4)	Wearing a face mask at all times
5)	Mask use while participating in activities and in community or public spaces, e.g., markets
6)	Mask use in enclosed space, e.g., shopping malls, cinemas, classrooms, or auditoriums
7)	Mask use while in the proximity of patients under investigation
8)	Sharing items capable of disease transmission, i.e., microphones or public items, within the past 14 days
9)	Handwashing or cleaning before eating or drinking
10)	Handwashing or cleaning after eating or drinking
11)	Handwashing or cleaning before using the toilet
12)	Handwashing or cleaning after using the toilet
13)	Hand cleaning with alcohol (gel or spray) after using the public toilet
14)	Hand cleaning with alcohol (gel or spray) after touching public goods
15)	Touching eyes, mouth, and nose with one's hands
16)	Using separate plates and glasses while having a meal with others
17)	Consuming undercooked or raw flesh
18)	Visiting live animal markets
19)	Touching dead animals without wearing gloves and handwashing afterwards
20)	Sharing personal items, i.e., handkerchiefs, water glasses, or towels
21)	Maintaining a warm body temperature
22)	Exercising regularly
23)	Sleeping > 6 hours a night
History behavior related to COVID-19 exposed risk	
24)	Travel or transit to high-risk countries within the past 14 days
25)	Living among people who have traveled abroad
26)	Having used domestic transportation services, such as airplanes, trains, public buses, public vans, the sky-train (BTS), or the metro (MRT) within the past 14 days
27)	Working in the field of health sciences or serving as medical personnel having contact with patients
28)	Teaching, studying, or researching with foreigners
29)	Working with or participating in activities with groups of people
30)	Contact with health science workers, medical personnel, tour guides, merchandisers, foreigners, or groups of people
31)	Sneezing or having experienced a fever, runny nose, or difficulty breathing within the past 14 days

2.2 Data Collection

The cross-sectional survey was distributed among the population of Chiang Mai University (CMU). The participants were divided into two groups and defined as: faculty staff members (N=606) and students (N=650). Random convenience sampling was used to collect as much sample data as possible. All methods were approved by the CMU Institutional Review Board (IRB number ET021/2020 and ET022/2020). Each record retrieved from the database was put through a de-identification step before proceeding to the data analysis step.

2.3 Data Analysis

Students and Faculty were compared on an item-by-item basis using chi-squared and independent sample t-tests as appropriate. Data were analyzed using IBM SPSS Statistics Software [6].

In addition, a multivariate neural network model was developed using Rapid Miner [7]. Rapid Miner is a data science software platform that provides an integrated environment for data preparation and predictive analysis. All survey items were submitted as input variables to Rapid Miner. Student vs. faculty group membership was considered the target output variable.

To do this, we constructed a neural network model with six hidden nodes comprised of a single hidden layer. Each of the 33 survey items shown in Table 3 and Table 4 were linked to all six of the hidden nodes by a weight value, resulting in 198 weights. There were no connections between nodes. The resulting six node values were aggregated to yield a single prediction of student or faculty via an unweighted mean value that passed through a logistic regression function. The 198 weights, plus the two parameters of the logistic regression, were fit using Rapid Miner Studio.

To assess the model's output, we determined the correlation between each of the survey items individually and the model's predictions. Further, we checked for model stability using 10-fold validation. In each fold, 10% of the data were held out and the model's fit was recalculated.

Furthermore, 20% of the data were not included in the process of fitting the model. The fitted model was challenged with this test data set. The primary metric of model performance was accuracy at classifying test data points as either students or faculty members.

3 Results

3.1 Data Exploratory

The participants were comprised of 1,266 individuals (650 students; 616 faculty members). Demographic data are reported in Table 2. Unsurprisingly, faculty tended to be older (mean age = 38.91) than students (mean age = 22.8). Groups were then similarly divided in terms of sex and field of study.

Table 2. Comparison of demographic data between students and faculty members

Participant demographics	Student (n=650)		faculty (n=616)	
	Number	%	Number	%
Sex				
Female	453	69.7	415	67.4
Male	197	30.3	201	32.6
Age				
< 20 years old	230	35.4	1	0.2
21-30 years old	370	56.9	156	25.3
31-40 years old	43	6.6	214	34.7
41-50 years old	6	0.9	151	24.5
51-60 years old	1	0.2	88	14.3
> 60 years old	0	0.0	6	1.0
Field of study				
Health sciences	99	15.2	127	20.6
Non-health sciences	551	84.8	489	79.4

We classified COVID-19 prevention and risk behaviors into two broad categories: history and behaviors. History items addressed things like travel and activities engaged in within the recent past. Behavior items addressed daily habits like mask wearing and hand washing.

We found that there were many differences between students and faculty members in terms of history. For example, faculty members were more likely to have traveled to high-risk countries within the past 14 days, but students were more likely to have been on public transportation. Faculty members were also more likely to work with patients, but students were more likely to take part in activities with groups. Finally, students were more likely to have exhibited symptoms of respiratory illness within the last two weeks, as is shown in Table 3.

Table 3. Comparison of history behavior between students and faculty members as related to COVID-19 exposed risk factors

	History behavior related to COVID-19 exposed risk	Students (n=650)		Faculty (n=616)		P-value
		Number	%	Number	%	
1)	Travel or transit to high-risk countries within the past 14 days	6	0.9	15	2.4	0.035
2)	Living among people who have traveled abroad	9	1.4	15	2.4	0.171
3)	Having used domestic transportation services, such as airplanes, trains, public buses, public vans, the sky-train (BTS), or the metro (MRT) within the past 14 days	145	22.3	94	15.3	0.001
4)	Working in the field of health sciences as medical personnel or having contact with patients	99	15.2	127	20.6	0.012
5)	Teaching, studying, or researching with foreigners	59	9.1	63	10.2	0.488
6)	Working or participating in activities with a group of people	363	55.8	305	49.5	0.024
7)	Contact with health science workers, medical personnel, tour guides, merchandisers, foreigners, or groups of people	223	34.3	231	37.5	0.237
8)	Sneezing or having experienced a fever, runny nose, or difficulty breathing within the past 14 days	245	37.7	106	17.2	< 0.001

There were also many differences between students and faculty members in terms of their daily behaviors. For example, students engaged in more sharing of personal items, but faculty were more likely to touch their faces. Perhaps this was because faculty also reported a lower prevalence of facemask wearing. Faculty did report more hand washing than students, got regular exercise more often, and slept more, as is shown in Table 4.

We also found differences in the overall risk profiles of students and faculty members, as shown in Table 5. While personal hygiene risks were similar between the groups, students reported significantly higher exposure-related risks than faculty members.

Table 4. Personal hygienic related risk data gathered from the survey distributed amongst students and faculty members

Personal hygienic related risks	Students (n=650)		Faculty staffs (n=616)		P-value
	Number	%	Number	%	
1) Contact with respiratory droplets within the past 14 days	114	17.5	72	11.7	0.003
2) Having participated in community activities or having gathered in public places within the past 14 days	334	51.4	308	50.0	0.622
3) Having been in enclosed spaces, e.g., shopping malls, cinemas, classrooms, or auditoriums, within the past 14 days	498	76.6	416	67.5	< 0.001
4) Wearing face masks at all times	434	66.8	367	59.6	0.008
5) Mask use while participating in community activities or when visiting public spaces, e.g., markets	598	92.0	566	91.9	0.939
6) Mask use in enclosed spaces, e.g., shopping malls, cinemas, classrooms, or auditoriums	565	86.9	551	89.4	0.165
7) Mask use while in the close proximity of patients under investigation	627	96.5	590	95.8	0.529
8) Sharing items capable of disease transmission, i.e., microphones or other public items within the past 14 days	195	30.0	222	36.0	0.022
9) Handwashing or cleaning before eating or drinking	572	88.0	573	93.0	0.002
10) Handwashing or cleaning after eating or drinking	505	77.7	548	89.0	< 0.001
11) Handwashing or cleaning before using the toilet	261	40.2	307	49.8	0.001
12) Handwashing or cleaning after using the toilet	629	96.8	606	98.4	0.064
13) Hand cleaning with alcohol (gel or spray) after using the public toilet	567	87.2	528	85.7	0.430
14) Hand cleaning with alcohol (gel or spray) after touching public goods	519	79.8	494	80.2	0.877
15) Touching one's eyes, mouth, and nose with hands	120	18.5	214	34.7	< 0.001
16) Using separate plates and glasses while having meals with others	493	75.8	485	78.7	0.221
17) Consuming undercooked or raw flesh	64	9.8	112	18.2	< 0.001
18) Visiting live animal markets	32	4.9	38	6.2	0.332
19) Touching dead animals without wearing gloves and handwashing afterwards	12	1.8	25	4.1	0.020
20) Sharing personal items, i.e., handkerchiefs, water glasses, or towels	82	12.6	59	9.6	0.086
21) Maintaining a warm body temperature	587	90.3	576	93.5	0.037
22) Exercising regularly	132	20.3	218	35.4	< 0.001
23) Sleeping > 6 hours	480	73.8	510	82.8	< 0.001

Table 5. Average risk scores between students and faculty staff members

Risk score	Students (n=650)		Faculty staff (n=616)		P-value
	Mean	Standard deviation	Mean	Standard deviation	
Personal hygienic related risks	2.38	2.002	2.28	2.176	0.429
History behavior related to COVID-19 exposed risk	7.60	3.085	6.92	3.174	< 0.001
All cause risks	9.96	4.037	9.21	4.186	0.001

3.2 Model Development

The health data set consisted of 1,266 participants, each of which contained 33 features. Feature values were taken from 33 questionnaires. The participant data are categorized and divided according to the score levels obtained from answering the questionnaire as follows: 1,050 participants belonged to the 'Low' class; 215 participants belonged to the 'Medium' class; 1 example was in the 'High' class. The dataset was split into two parts for training and testing using a 70/30 ratio basis.

The neural network structure was set up. In the input layer, the features are retrieved from 33 items in the questionnaire. Accordingly, there were two hidden layers. The first layer contained 64 hidden neurons, while the second contained 32 hidden neurons. This neural network construct used the Activation Function Sigmoid and used 20%-dropout for both hidden layers. In the output layer, the SoftMax function was used to establish the probabilities of each feature. The three classes were 'Low', 'Medium', and 'High', where the predicted class label was the class with the highest probability, as is shown in Figure 1.

With regard to the artificial neural network, we compared the prediction performance with four other standard machine learning models including: K-nearest neighbor (K-NN), Decision tree, Gaussian Naïve Bays, and Support Vector Machine. The prediction results are shown in Table 6 below and are presented as a confusion matrix table.

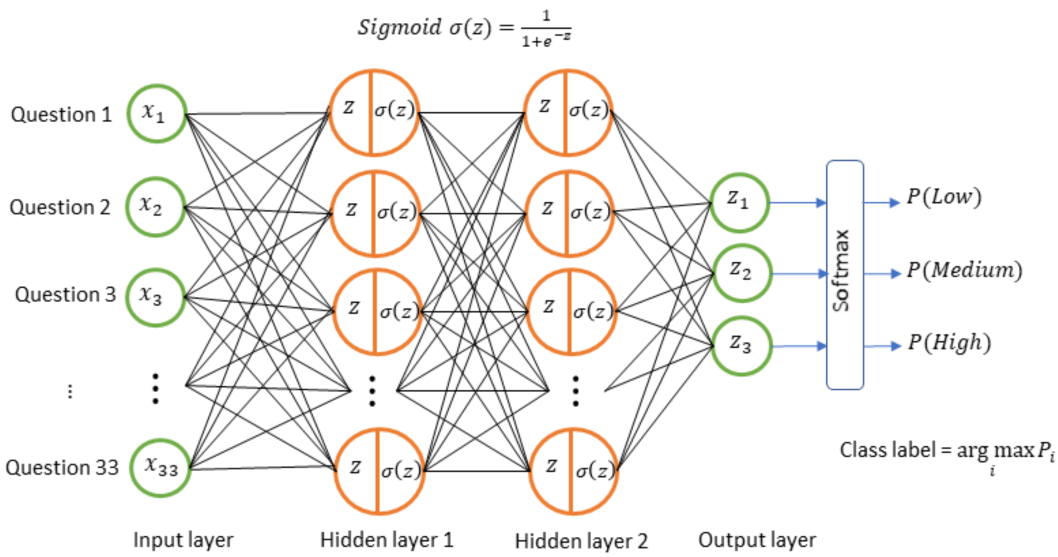

Figure 1. Neural network structure diagram

Table 6. Confusion matrix from each machine learning model

		Low	Medium	High
K-NN	Low	323	1	0
	Medium	35	21	0
	High	0	1	0
Decision Tree	Low	296	28	0
	Medium	30	26	0
	High	0	0	1
Gaussian Naïve Bays	Low	296	28	0
	Medium	17	39	0
	High	0	0	1
Support Vector Machine	Low	322	2	0
	Medium	14	42	0
	High	0	1	1
Artificial Neural Network	Low	317	7	0
	Medium	12	44	0
	High	0	0	1

Table 7. Model performances for each prediction class categorized by a machine learning model

Model	Prediction class	Sensitivity		Specificity		Accuracy
		True positive rate	False negative rate	False positive rate	True negative rate	
K-Nearest Neighbors	Low	1.00	0.00	0.61	0.39	0.91
	Medium	0.38	0.63	0.01	0.99	0.90
	High	0.00	1.00	0.00	1.00	1.00
Decision Tree	Low	0.91	0.09	0.53	0.47	0.85
	Medium	0.46	0.54	0.09	0.91	0.85
	High	1.00	0.00	0.00	1.00	1.00
Gaussian Naive Bayes	Low	0.91	0.09	0.30	0.70	0.88
	Medium	0.70	0.30	0.09	0.91	0.88
	High	1.00	0.00	0.00	1.00	1.00
Support Vector Machine	Low	0.99	0.01	0.24	0.76	0.96
	Medium	0.75	0.25	0.01	0.99	0.96
	High	0.50	0.50	0.00	1.00	1.00
Artificial Neural Network	Low	0.98	0.02	0.21	0.79	0.95
	Medium	0.79	0.21	0.02	0.98	0.95
	High	1.00	0.00	0.00	1.00	1.00

As is shown in Table 7, the prediction performance of the support vector machine and the artificial neural network outperformed other machine learning models. However, the artificial neural network model exhibited better performance for predicting a ‘high risk’ level as well as for giving a lower false positive rate when compared with the support vector machine.

4 Discussion

Students exhibited a higher overall risk score than faculty members. This was primarily driven by risk behaviors rather than histories. Specifically, students were more likely to wash their hands less, exercise less, and get less sleep. By contrast, faculty members also exhibited some risk behaviors that were not present in students. For example, faculty members were more likely to interact with patients in a medical setting.

The variabilities for risk histories and behaviors were complex and difficult to capture. Although students were associated with higher overall risks than faculty members, there was no clear driver for this effect. Nevertheless, it has been shown previously that combined index scores derived from survey data, such as those presented here, can accurately predict the likelihood of a COVID-19 diagnosis [1]. Thus, we predict that the CMU student population can be expected to have a higher risk of COVID-19 diagnosis than the faculty population. Indeed, since the period of the data collection reported on here, Thailand has experienced an unfortunate and fast-moving COVID-19 outbreak. CMU has carefully monitored the situation in its internal population. However, the statistics reported by the university have not been categorized according to students and faculty members. This result largely echoes the findings of previous research work demonstrating the viability of risk assessment via self-report surveys [3-4, 8-11].

When an outbreak strikes, it is often not feasible to survey entire populations. For example, in the present study, surveying 3.5% of the CMU population required two passes. Surveying the entire population fast enough to meet public health decision making needs would not have been feasible. To address this issue, we wanted to know if it was possible to predict risk behaviors based on societal group membership. To do this, we developed a neural network model using a widely available and user-friendly GUI-based application, RapidMiner. We found that the model was able to predict student/faculty group membership in a test data set with 96% accuracy, which is well above the chance rate of 50%. Thus, public health decision makers can make decisions about outbreak mitigation measures and resource assignment using survey data collected from limited samples of different societal groups. This can be done with confidence that the risk profiles in even relatively small samples (3.5% in our case) would be able to be generalized for the overall population. Leveraging this finding could dramatically speed up decision making in the critical early stages of an outbreak.

For example, it has been shown that targeted information campaigns can improve COVID-19 prevention measures [8]. Making these initiatives more targeted could be highly beneficial. It has also been found that some societal groups do not have access to certain prevention measures. Identifying these groups and helping them address risks within their environment could slow the spread of disease.

Interestingly, mask wearing was the most heavily weighted input variable in our model. This is an indication that mask wearing was the most common difference between students and faculty. Complicating the interpretation of this variable, faculty members wore masks less often than students. Although students exhibited a higher overall degree of risk and a higher number of COVID-19 infections, based solely on mask wearing, faculty were at a greater risk. This demonstrates that risk indices, which aggregate profiles of risk behaviors and histories, may be more effective than singular risk factors at predicting COVID-19 infection. Thus, even though mask wearing has been found to be one of the most important mitigation measures that an individual can take, it is not necessarily monolithic [12].

The main limitation of this study was that we were not able to test participants for COVID-19 to directly assess empirical risks. The main reason for this limitation was that at the time of data collection, there were very few COVID-19 cases in Chiang Mai. Future studies that combine risk assessment with an empirical measurement of COVID-19 test positivity should test whether risk profiling, as explored here, can help to effectively identify at risk groups and individuals.

Acknowledgments

Conceptualization, W.B. and W.P.; Data curation, T.P. and K.T.; Formal analysis, W.B.; Funding acquisition, W.B.; Investigation, K.T.; Methodology, T.P.; Project administration, W.B.; Resources, Software, W.B. and K.D.; Supervision, W.B.; Validation, W.B.; Writing – original draft, T.P. and K.T. All authors have read and approved of the final version of this manuscript and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work will be appropriately investigated and resolved.

This research was partially supported by Chiang Mai University, Fundamental Fund 2022, Chiang Mai University under grant number FF65/059 and NSRF via the Program Management Unit for Human Resources & Institutional Development, Research and Innovation [Grant Number B05F640183].

This work was supported by the Center of Excellence in Community Health Informatics, Faculty of Science, Chiang Mai University and the Faculty of Public Health, Chiang Mai University, and supported by the CMU Proactive Researcher Program, Chiang Mai University (Grant Number: 729/2567), Chiang Mai, Thailand. The authors extend their gratitude to all participants for their valuable contributions to this study.

References

[1] Y. Liu, Z. Wang, J. Ren, Y. Tian, M. Zhou, T. Zhou, K. Ye, Y. Zhao, Y. Qiu, J. Li, A Covid-19 Risk Assessment Decision Support System for General Practitioners: Design and Development Study, *Journal of Medical Internet Research*, Vol. 22, No. 6, Article No. e19717, June, 2020. https://doi.org/10.2196/19786

[2] K. Intawong, D. Olson, S. Chariyalertsak, Application Technology to Fight the COVID-19 Pandemic: Lessons Learned in Thailand, *Biochemical and Biophysical Research Communications*, Vol. 538, pp. 231-237, January, 2021. https://doi.org/10.1016/j.bbrc.2021.01.093

[3] I. Kim, J. Lee, J. Lee, E. Shin, C. Chu, S. K. Lee, KCDC Risk Assessments on the Initial Phase of the COVID-19 Outbreak in Korea, *Osong Public Health and Research Perspectives*, Vol. 11, No. 2, pp. 67-73, April, 2020. http://doi.org/10.24171/j.phrp.2020.11.2.02

[4] R. Chatterjee, S. Bajwa, D. Dwivedi, R. Kanji, M. Ahammed, R. Shaw, Covid-19 Risk Assessment Tool: Dual Application of Risk Communication and Risk Governance, *Progress in Disaster Science*, Vol. 7, Article No. 100109, October, 2020. https://doi.org/10.1016/j.pdisas.2020.100109

[5] E. Dong, H. Du, L. Gardner, An Interactive Web-Based Dashboard to Track COVID-19 in Real Time, *The Lancet Infectious Diseases*, Vol. 20, No. 5, pp. 533-534, May, 2020. https://doi.org/10.1016/S1473-3099(20)30120-1

[6] IBM Corporation, *IBM SPSS Statistics for Windows, Version 24.0*, IBM Corporation, Armonk, NY, November, 2016.

[7] RapidMiner, *RapidMiner Studio for Academics*, RapidMiner, Boston, MA, November, 2020.

[8] P. Srichan, T. Apidechkul, R. Tamornpark, F. Yeemard, S. Khunthason, S. Kitchanapaiboon, P. Wongnuch, A. Wongphaet, P. Upala, Knowledge, Attitudes and Preparedness to Respond to COVID-19 among the Border Population of Northern Thailand in the Early Period of the Pandemic: A Cross-Sectional Study, *WHO South-East Asia Journal of Public Health*, Vol. 9, No. 2, pp. 118-125, September, 2020. https://journals.lww.com/wsep/fulltext/2020/09020/knowledge,_attitudes_and_preparedness_to_respond.7.aspx

[9] P. Doung-ngern, R. Suphanchaimat, A. Panjangampatthana, C. Janekrongtham, D. Ruampoom, N. Daochaeng, N. Eungkanit, N. Pisitpayat, N. Srisong, O. Yasopa, P. Plernprom, P. Promduangsi, P. Kumphon, P. Suangtho, P. Watakulsin, S. Chaiya, S. Kripattanapong, T. Chantian, E. Bloss, C. Namwat, D. Limmathurotsakul, Case-Control Study of Use of Personal Protective Measures and Risk for SARS-CoV-2 Infection, Thailand, *Emerging Infectious Diseases*, Vol. 26, No. 11, pp. 2607-2616, November, 2020. https://doi.org/10.3201/eid2611.203003

[10] T. Nania, F. Dellafiore, R. Caruso, S. Barello, Risk and Protective Factors for Psychological Distress among Italian University Students during the COVID-19 Pandemic: The Beneficial Role of Health Engagement, *International Journal of Social Psychiatry*, Vol. 67, No. 1, pp. 102-103, February, 2021. https://doi.org/10.1177/0020764020945729

[11] D. V. Nguyen, G. H. Pham, D. N. Nguyen, Impact of the COVID-19 Pandemic on Perceptions and Behaviors of University Students in Vietnam, *Data in Brief*, Vol. 31, Article No. 105880, August, 2020. https://doi.org/10.1016/j.dib.2020.105880

[12] W. Lyu, G. L. Wehby, Community Use of Face Masks and COVID-19: Evidence from a Natural Experiment of State Mandates in the US, *Health Affairs*, Vol. 39, No. 8, pp. 1419-1425, August, 2020. https://doi.org/10.1377/hlthaff.2020.00818

Biographies

Thitikan Phuwitthanasap is a proactive researcher at the Office of Research Administration, Chiang Mai University, Chiang Mai, Thailand. She received her B.Sc. in Public Health from the School of Medicine, University of Phayao, Thailand, in 2016, and her M.P.H. from Chiang Mai University, Thailand, in 2021. Her research interests focus on community health promotion.

Khanita Duangchaemkarn is a lecturer at the University of Phayao with a Ph.D. in Biomedical Engineering. She leads the CATERx research group, applying artificial intelligence and biomedical data engineering to pharmaceutical sciences. Her work improves medication safety and contributes to national regulations for medical AI.

Kitbordin Thongduang is a lecturer at Naresuan University. He holds a Master's degree in Public Health and Bachelor's degrees in Community Health and Physical Therapy. His research focuses on NCD and healthcare systems, with expanding interests in health economics and health informatics.

Wanicha Pungchompoo holds a Doctoral degree (DClinP.) from the University of Southampton, School of Health Sciences, UK. She was a post-doctoral trainee at university of Michigan, School of Nursing, USA. Her area of interests are NCD, Chronic Kidney Disease, Aging, Palliative Care, Mixed Method research, Healthy University, Telehealth and Web-application.

Waraporn Boonchieng is the Dean of the Faculty of Public Health at Chiang Mai University. She earned her Doctor of Public Health (Dr.PH.) from Mahidol University, Thailand. Her expertise encompasses community health promotion and application of public health informatics.