



Development of COVID-19 Health-Risk Assessment and Self-Evaluation (CHaSe): a health screening system for university students and staff during the movement control order (MCO)

Azimah Ismail¹ · Mohd Salami Ibrahim² · Samhani Ismail² · Azwa Abdul Aziz³ · Harmy Mohamed Yusoff² · Mokhairi Mokhtar³ · Hafizan Juahir¹

Received: 11 June 2021 / Revised: 26 January 2022 / Accepted: 27 January 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Austria, part of Springer Nature 2022

Abstract

COVID-19 has triggered a global health crisis. Death from severe respiratory failure and symptoms, including fever, dry cough, sore throat, anosmia, and gastrointestinal disturbances, has been attributed to the disease. Development of screening and diagnosis methods prove to be challenging due to shared clinical features between COVID-19 and other pathologies, such as Middle Eastern respiratory syndrome, severe acute respiratory syndrome, and common colds. This study aims to develop a comprehensive one-stop online public health screening system based on clinical and epidemiological criteria. The immediate target populations are the university students and staff of University Sultan Zainal Abidin and the civil servants of the Malaysian Ministry of Science, Technology, and Innovation. Forty-nine (49) clinical and epidemiological factors associated with COVID-19 were identified and prioritized based on their prevalence via rigorous review of the literature and vetting sessions. A pilot study of 200 volunteers was conducted to assess the extent of risk mitigation of COVID-19 infection among the university students and civil servants using the prototyped model. Consequently, twelve (12) clinical parameters were identified and validated by the medical experts as essential variables for COVID-19 risk-screening. The updated model was then revalidated via real mass-screening of 5000 resulting in the final adopted CHaSe system. Principal component analysis (PCA) was used to confirm the weightage of risk level toward COVID-19 to procure the optimal accuracy, reliability, and efficiency of this system. Twelve (12) factor loadings accountable for 58.287% of the clinical symptoms and clinical history variables with forty-nine (49) parameters of COVID-19 were identified through PCA. The variables of the clinical and epidemiological aspects identified are the C6 (History of joining high-risk gathering (where confirmed cases had been recorded)), CH11 [History of contact with confirmed cases (close contact)], CH13 [Duration of exposure with confirmed cases (minutes)] with substantial positive factors of 0.7053, 0.706 and 0.5086, respectively. The contribution toward high-risk infection of COVID-19 was firmly attributable to the variables CH14 [Last contact with confirmed cases (days)], CH13 [Duration of exposure with confirmed cases (minutes)], and S1 (Age). The revalidated PCA for 5000 respondents also yielded twelve significant PCs with a cumulative variance of 58.288%. Importantly, the medical experts have revalidated the CHaSe system for accuracy of all clinical aspects (clinical symptoms and clinical history) and epidemiological links to COVID-19 infection. After revalidating the model for 5000 respondents, the PC variance for PC1, PC2, PC3, and PC4 was 27.36%, 11.79%, 10.347%, and 8.785%, respectively, with the cumulative explanation of 58.288% in data variability. The level of risks detected using the CHaSe system toward COVID-19 provides optimal accuracy, reliability, and efficiency to conduct mass-screening of students and government servants for COVID-19 infection.

Keywords Pandemics · Risk assessment · Coronavirus · Multivariate analysis

1 Introduction

Since the first case was reported more than a year ago, the coronavirus disease 2019 (COVID-19) has significantly changed the world. Common symptoms attributable to COVID-19 include dry cough, fever, sore throat, diarrhea,

✉ Hafizan Juahir
hafizanjuahir@unisza.edu.my

Extended author information available on the last page of the article

vomiting, headache, and altered smell sensation. More severe complications include respiratory failure with invasive airway support and death (WHO & China 2020). The broad range of symptoms and complications appears to contribute to the easy spread of the disease. Many infected individuals with mild to moderate symptoms had unknowingly spread the disease, increasing its transmissibility. At the time of writing, the pandemic is responsible for more than 2.2 million deaths while the total number of global cases surge pass 102 million (Jhu 2020). COVID-19 has thus brought a record scale of impact on the economy, industries, health, and social well-being of societies around the world.

Consequently, countries worldwide have implemented mass population testing and extensive contact tracing to track, isolate, and treat infected individuals at the earliest time possible. The critical need for mass testing to identify at-risk and infected individuals early is evident by China and South Korea's successful containment strategies. In both countries, mass population diagnostic testing enabled early diagnosis, identification, and isolation infected individuals. This achievement significantly reduced the number of new cases than was initially predicted (WHO & China 2020; Balilla 2020). Unfortunately, mass diagnostic testing requires heavy use of resources, limiting this option to countries with access to rich supplies and human resources. Therefore, a more practical alternative population-based screening with acceptable validity will complement the existing measures.

In response, Universiti Sultan Zainal Abidin (UniSZA) developed the COVID-19 Health-Risk Assessment and Self-Evaluation (CHaSe) System. This system, accessible via the internet browser and mobile app application, aimed to provide an intelligent and comprehensive analysis of at-risk individuals at the earliest stage possible. The development of CHaSe was primarily motivated by the critical need to conduct a thorough daily risk-assessment screening among all staff and students of the university as well as civil servants under the ministry of science, technology, and innovation (MOSTI) and various other government and non-governmental agencies. During the CHaSe development, Malaysia implemented a movement control order (MCO) policy that warranted work-from-home for all working class and student groups. Upon lifting the MCO policy, many travelers from areas with high-density cases of COVID-19 may return to workplaces, such as university campuses and MOSTI headquarter, to resume study and work duty. Consequently, the magnitude of risk of outbreaks associated with the influx of returning students and staff mandates more practical, robust, vigilant, and reliable population-based screening measures.

CHaSe fills this gap by serving as a platform for intelligent and comprehensive analysis of COVID-19 screening, promotes preventive measures against COVID-19, and recommends advice on action according to individual

risk categories. Featuring daily risk-screening assessment, CHaSe extends its role as a comprehensive surveillance tool where individuals with an elevated risk of contracting COVID-19 may declare their status and be contactable by the surveillance team while preserving the ethical medical practice autonomy and confidentiality.

Hence, CHaSe enhances the conventional active case detection and contract tracing practice, typically initiated by the health personnel. Through CHaSe, the community can access a valid population-based screening and declare these risk assessments voluntarily to public health personnel. This access significantly increases contact tracing and active case detection outreach, which otherwise would have been practically impossible to attain. This study outlines the development of the risk-prediction model employed within the CHaSe system to achieve this goal.

2 Methodology

The urgent development of the CHaSe system began by having brainstorming sessions among UniSZA experts from the Faculty of Medicine, Faculty of Informatics and Computer and the Deputy of Research and Innovation Vice Chancellor. Due to the significant morbidity and mortality associated with COVID-19, the development of the CHaSe system adopted the Six Sigma approach of DMAIC (Define, Measure, Analyze, Improve, and Control) (Juahir et al. 2017) in its design.

Six Sigma is a data- and statistical-driven approach for continuous quality improvement via eliminating defects or problems in the manufacturing and development process. Hundreds of businesses and conglomerates worldwide, such as McDonald's, Texas Instruments, and many Japanese companies, have adopted the Six Sigma approach. Six Sigma's emphasis on quality translates to increased productivity speed with reduced defects at a lower utilization of resources and cost (Juahir et al. 2017; Cox et al. 2016). The systematic approach assures goal attainment and continuous improvement given the relatively rapid development in knowledge and policies associated with this novel pathology.

3 Development of CHaSe via Six Sigma approach of DMAIC

There are five major phases to DMAIC six-sigma in the development of CHaSe (Juahir et al. 2017; Aris et al. 2010).

a. Define:

Lead by the Vice Chancellor of UniSZA, this phase concentrates on identifying the gaps to address. The ultimate goal is to determine the solutions to ensure health

and safety of all students, staff, and general members of the public within the locality of UniSZA and MOSTI. The thorough discussion identified the need for a population-based screening among students and staff that is practical, valid, robust, and timely for decision-making. Eleven medical specialists from the faculty of medicine, UniSZA, with diverse medical and surgical expertise, participated in a comprehensive literature review to identify established risks associated with COVID-19 infection. We also elicit assistance from an information specialist from the library to perform a robust searching strategy with sensitive search strings from all relevant databases. Two independent medical experts screened and selected articles associated with COVID-19. All papers and reports were then uploaded into Microsoft Teams (Microsoft Inc. 2018) for real-time online collaboration during the next phase.

b. Measure:

The measuring phase involved substantiating the prevalence, frequency, and relevance of variables associated with COVID-19 based on evidence from the literature and medical experts' consensus. Extensive COVID-19 literature was divided and delegated among all individual experts. As a critical-to-quality (CTQ) measure, a standardized data extraction tool was employed within Microsoft Team. The use of cloud-based data extraction enabled a secured synchronization of collaboration and co-authoring (Microsoft Inc. 2018), which brought together a total of 49 variables extracted from the literature by individual medical experts into a single master data collection (WHO & China 2020; Jhu 2020; Balilla 2020; Center for Disease Control and Prevention 2021; Ministry of Health Malaysia 2020; Islam et al. 2020; Siddiqi et al. 2021; Kluster Tawar 2020; Noor Hisham Abdulah 2020; McMichael et al. 2020; MOH 2020a, b). Based on all data collected in the database, the associated reported prevalence and relevance were thoroughly scrutinized, vetted, and revalidated by medical experts on their significance and relevance. The revalidated variables were then used for simulation model calculation, where each variable was given a dedicated weightage/scoring to reflect their significance toward risk prediction of COVID-19. This model's outcome are three distinct risk categories, namely, Low, Medium, and High.

c. Analyze:

This phase encapsulated the complex process of data preprocessing, data analyses, data analytics, and data validation to generate a risk-prediction model that meets or exceeds the minimum standard of public health practice. We simulated a pilot study by distributing a questionnaire consisting of the validated variables to 200 healthy volunteers. Deliberate weightage determination was assigned for selected parameters. These include

close contact with patients of COVID-19, overseas travel abroad within the last 14 days, cough symptom, fever symptom, breathing difficulty, and critical illnesses.

d. Improve:

To further improve the modeling and CHaSe system itself, we conduct the first trial of mass-screening among 5000 university students during the nation-wide movement control order (MCO). Respondents under the high-risk category were contacted by medical experts and advised to visit the nearest local health center for further assessment.

The real mass-screening provides empirical evidence to facilitate the modification of processes as CHaSe is continuously monitored and improvised. The iterative process of recalculating each medical parameter's weightage/scoring was employed using the principal component analysis (PCA). After the new weightages are recalculated, the improved model statistically categorized respondents into three groups; (a) Low Risk, (b) Moderate Risk, and (c) High Risk. This improvement phase of the DMAIC Six Sigma for the CHaSe system acts to evaluate the modification of processes to obtain a more meaningful data interpretation.

e. Control

As a system-driven methodology, the quality of recommendation from the CHaSe system conforms to the standard operating procedure (SOP) and guidelines by the health ministry.

- (a) Low Risk: students and civil servants should record the daily Self-Assessment Module from CHaSe until further notice. They should also be practicing the recommended personal preventive measures against COVID-19 infection.
- (b) Medium Risk: students and civil servants in this category should limit contact with the community and contact UniSZA's contact tracing team for a telephone-based clinical assessment and recommendation.
- (c) High Risk: UniSZA's contract tracing team initiates telephone contact at the earliest time practical under the active-case-detection strategy. Respondents in this category must limit contact with the community and urgently visit the nearest health center with COVID-19 diagnostic capacity for further assessment.

4 Development health screening model within CHaSe

The health screening model within the CHaSe system was developed with the consideration of the following aspects:

4.1 Principal component analysis (PCA)

Principal component analysis (PCA) is that variance in common from the original dataset-based multivariate techniques in revealing the strong factors or factor loadings of a dataset via the underlying latent vectors of principal components (PCs) (Ismail et al. 2018; Fahmi 2011). These vectors transform the original dataset into new uncorrelated variables producing the best linear fitting with an orthogonal basis. The new axes lie along the orthogonal lines in the directions of maximum variance (Shrestha and Kazama 2007). The equation below defines the derived outcome (Cloutier et al. 2008);

$$\text{PCs } z_{ij} = a_{i1}x_{1j} + a_{i2}x_{2j} + \dots + a_{im}x_{mj} \quad (1)$$

In this equation, z denotes the component score and a represents the component loading. Meanwhile, x is the measured value of the variable, i is the component number, j is the sample number, and m is the total number of variables. At the same time, it maximizes the variance and defines the association between the PCs, namely the clinical aspects and epidemiological link to COVID-19 (Epid-Link). Recycled data were important in this study. The index values from the factor score of PCA in the first place consisted of varying scales (internally inconsistent values of negatives) that the algorithm was not easy to make the comparison or assumption with it (Helena et al. 2000).

Generally these PCs are unstable and must be subjected to rotation known as the varimax rotation generating the varimax factors (VFs) (Kamaruddin et al. 2015). Varimax rotation is important to regulate the overfitting and complexity of the components by making enhancing the positive factor loadings and negative factor loadings within each principal component. The varimax rotation produces a more transparent quantitative measure to correlate each variable with one PC or almost one factor only (Juahir et al. 2017; Ismail et al. 2018; Kamaruddin et al. 2015; Dominick et al. 2012; Fazilah et al. 2017). The varimax factors with an eigenvalue significantly greater than one (1), form the critical criterion for further analysis (Juahir et al. 2017). In this study, the strong factor loading of VFs equal to or greater than 0.4 is considered a strong factor loading. The value ranging from 0.2 to 0.39 is deemed moderate, while 0.10 to 0.199 is regarded as weak factor loading (Liu et al. 2003). The negative value of factor loadings represents the variables' insignificant loading in either the clinical aspects or Epid-Link.

In this study, PCA was applied to all initial and revaluated variables in the questionnaire distributed among 200 respondents during the measurement phase and 5000 respondents in the improvement phase. Specifically, the PCA re-constructs factors (such as clinical symptoms of fever and overseas travel) from the large dataset into their

respective underlying latent variables, the clinical aspects and the Epid-Link. Subsequently, the factor scores of PCA were used as a single index for every contributing factor. The weighted score was used to determine the risk level of the respondents toward the COVID-19 infection based on an algorithm expressed by the equation below:

$$\text{PRI} = \sum_i^n F_i w_i \quad (2)$$

where n denotes the number of factors, F_i represents the score of factor i and w_i is the percentage of the variance factor that factor i explains.

The index values from the factor scores of PCA consisted of varying scales (internally inconsistent values of negatives) which results in algorithmic difficulty in making comparison and developing assumptions. To achieve consistency, the scales were normalized via the rescaling method; the index values were rescaled to z values (to ensure variance for each variable would equal to unity) using the expression below:

$$\text{Rescaling}(1\text{to}100) = a + [(xi - A) \times (b - a)/(B - A)] \quad (3)$$

where a is equal to 1, xi is the actual observation, A and B are the lowest and highest factor score, and b is the constant value of 100. As the rescaling involved the data normalization, univariate clustering was selected to categorize the risks, namely: Low Risk, Moderate Risk, and High Risk. The lowest value of risk indicates that the respondents have a low risk of COVID-19 infection. Likewise, a high value suggests that respondents have a high risk of contracting COVID-19, thus requiring an urgent further clinical assessment and management.

4.2 Discriminant analysis (DA)

Discriminant analysis is a technique for developing the independent variables' discriminant functions. These functions significantly distinguish these independent variables among themselves (predictors variables) according to the dependent variables (categorical variables). The result is a classification method that produces perfect separation or very minimal misclassification rate (Juahir et al. 2017; Ismail et al. 2018; Gazzaz et al. 2012). By applying a linear combination of the predictor variables (Lambrakis et al. 2004), achieved by maximizing the correlation of variance–covariance between classes while minimizing the variance–covariance correlation within the same class, the discriminant factors (DFs) reveal the derived categorical-variables-based classification (Juahir et al. 2017; Ismail et al. 2018; Gazzaz et al. 2012). The DFs for each cluster follow the equation:

$$f(G_i) = ki + \sum_{j=1}^n w_{ij}p_{ij} \tag{4}$$

where i denotes the number of groups (G), ki represents the constant inherent of each group, n is the number of parameters used to classify a set of data into a given group and w_j is the weight coefficient assigned by DF analysis (DFA) to a given parameter (p_j) (Kannel et al. 2007).

The DFs to evaluate the categorical variables of clinical aspects and the Epid-Link were constructed via the DA technique based on two distinct modes, viz. standard and forward stepwise. In the initial phase of the pilot study, the construction of DFs was for 200 respondents with 49 variables. Later, the variables were revalidated that subsequently reduced to 12 variables to serve 5000 respondents. Therefore, the data input to discriminant analysis constituted 200 and 5000 samples of respondents. In the DA stepwise forward mode, variables were added step by step, starting from the most significant variables until no changes were observed (Fazillah et al. 2017).

5 Results and discussion

5.1 Define phase

The first phase of the Six Sigma DMAIC’s critical-to-quality (CTQ) approach identified several root causes of outbreaks and potential interventions to COVID-19 infection in Malaysia. The Fishbone Diagram in Fig. 1 below illustrates the dynamic relationship between crucial elements from the known COVID-19 clusters in Malaysia with the underlying mechanism of transmission.

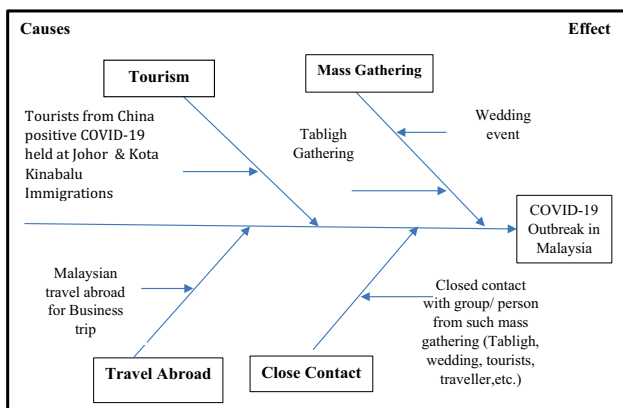


Fig. 1 A fishbone diagram illustrating the cause and effect of COVID-19 outbreak in Malaysia

5.2 Measure phase

The initial extensive literature review identified 49 parameters associated with COVID-19, as shown in Table 1 below.

5.3 Analyze phase

5.3.1 Pilot study using forty-nine (49) clinical aspects and epidemiological variables linked to COVID-19 for 200 respondents

5.3.1.1 Determination of weightage scoring using the principal component analysis (PCA) All the initial forty-nine (49) parameters of COVID-19 were scrutinized before confirming the weightage score for categorizing the index using the PCA. PCs with the eigenvalues of more than 1.0 were considered significant. The PCA output revealed twelve significant PCs extracted from the variables with eigenvalues greater than 1 (Table 2). The variance explanations of the PCs were 23.84%, 11.794%, 10.347%, and 8.785% for PC1, PC2, PC3, and PC4, respectively, with a total cumulative explanation of 58.287% in data variability. Thus, the PCA's twelve factor loadings are accountable for most clinical symptoms and clinical history variations. Therefore, predicting the risk of contracting COVID-19 infection based on clinical symptoms and Epid-Link is scientifically determined by these twelve factors (Niranjnamurthy et al. 2020; Cao et al. 2020) (Fig. 2).

Referring to PC1 in Table 2, C6 (History of joining high-risk gathering where confirmed cases had been recorded), CH11 [History of contact with confirmed cases (close contact)], and CH13 [Duration of exposure with confirmed cases (minutes)] display strong positive factors of 0.705, 0.706 and 0.509, respectively. These values suggest they are significant variables to predict the risk of contracting COVID-19. Nevertheless, the variables of CS11 [Duration of muscle pain/ fatigue (days)] and CS15 (Headache) with the factor loadings of 0.444 and 0.463 are also considered significant.

The PC2 denotes underlying risks associated with age (S1) and underlying history of hypertension (Com2) at the weighted value of 0.8478 and 0.5880, respectively. In the United States of America (US), every 8 out of 10 deaths are from adults aged 65 or older (Center for Disease Control and Prevention 2021). Similarly, those aged 60 or above account for the majority of deaths from COVID-19 in Malaysia (Ministry of Health Malaysia 2020). The relatively lower cardiovascular reserve may explain the high burden of disease and complications caused by COVID-19 among this age group. Furthermore, the older population is more likely to have limited mobility issues and possibly take residence in aged-care centers. These factors explain the increased risk of contracting COVID-19 due to shared

Table 1 Forty-Nine (49) parameters or variables used for pilot study based on literature reviews

No	Section	Var	Description
1	Sociodemographic	S1	Age (Years)
2		S2	Gender
3		S3	Race
4		S4	Place of living (State)
5		S5	Type of residency
6		S6	Locality
7		S7	Occupation
8		S8	At risk occupation
9	Clinical history	CH1	Use of mask
10		CH2	Use of sanitizer
11		CH3	Practice of Social Distancing (More than 1 m apart)
12		CH4	History of involving with mass gathering of more than 50 people since last two weeks
13		CH5	History of family members/co-workers involved with mass gathering of more than 50 people since last two weeks
14		CH6	History of joining high-risk gathering (where confirmed cases had been recorded)
15		CH7	History of Visit to any Clinic or Hospital
16		CH8	International Travel History
17		CH9	***Dynamic Variable—Travel to and from high-density state
18		CH10	***Dynamic Variable—Travel to and from high-density countries
19		CH11	History of contact with confirmed cases (Close contact)
20		CH12	History of contact with close contacts (Casual contacts)
21		CH13	Duration of exposure with confirmed cases (in minutes)
22		CH14	Last contact with confirmed cases (in days)
23	Clinical symptom	CS1	Fever
24		CS2	Duration of fever (Days)
25		CS3	Dry Cough
26		CS4	Duration of dry cough (Days)
27		CS5	Productive cough
28		CS6	Runny nose
29		CS7	Sore throats
30		CS8	difficulty breathing
31		CS9	Duration of difficulty breathing (Days)
32		CS10	Muscle pain/fatigue
33		CS11	Duration of muscle pain/fatigue (Days)
34		CS12	Vomiting
35		CS13	Diarrhea
36		CS14	Conjunctival congestion
37		CS15	Headache
38		CS16	Abdominal Pain
39	Comorbidity	Com1	Diabetes
40		Com2	Hypertension
41		Com3	Ischemic Heart Disease
42		Com4	Malignancy
43		Com5	Heart Failure
44		Com6	Respiratory disease (e.g., asthma, COPD, bronchiolitis etc.)
45		Com7	Smokers
46		Com8	Renal Disease
47		Com9	Low Immunity (e.g., HIV, on long-term steroid, on chemotherapy, on immunospressant etc.)
48		Com10	Stroke/TIA
49		Com11	Chronic Liver Disease

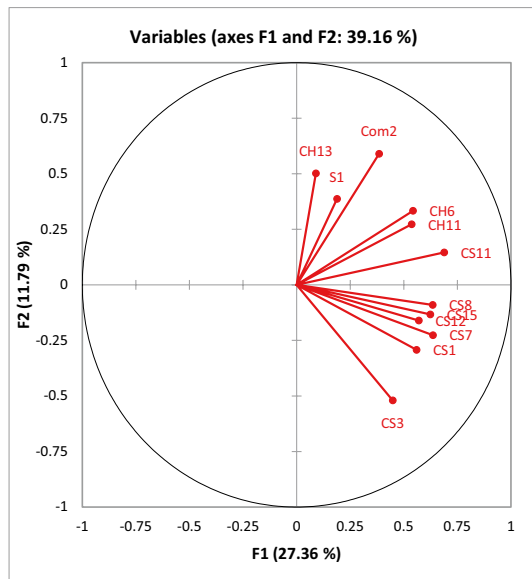


Fig. 2 Variance explanation of the strong factor loadings of 27.36% for PC1 and 11.794% for PC2 after varimax rotation of 49 parameters for 200 respondents

Table 2 Eigenvalues from the principal component analysis illustrating variance, cumulative variance, and factor loading of clinical aspects (clinical symptoms and clinical history) and epidemiological link of COVID-19 (Epid-Link) after varimax rotation of 49 parameters for 200 respondents

Variables	D1	D2	D3	D4
S1	-0.0708	0.8478	0.0048	0.0792
CH6	0.7053	-0.0134	0.2743	-0.0641
CH11	0.7067	-0.0200	0.0481	0.1781
CH13	0.5086	0.1669	-0.4748	0.0440
CS1	0.2254	-0.2029	0.6456	0.1917
CS3	-0.0850	0.0203	0.0985	0.8513
CS7	0.2782	0.0693	0.1802	0.6824
CS8	0.0910	0.3343	0.7064	0.1951
CS11	0.4444	0.3305	0.2784	0.3701
CS12	0.1517	0.0804	0.6718	0.1469
CS15	0.4632	-0.1151	0.2089	0.4913
Com2	0.4183	0.5580	0.2133	-0.2155

The factor loadings from 0.4 and above are considered strong (highlighted in bold)

environment and close physical contact. In contrast, hypertension, which also steadily increases with age, is commonly treated with angiotensin-converting enzyme inhibitor (ACE-i) and angiotensin-receptor blocker (ARB). Both ACE-i and ARB are associated with COVID-19 through upregulation of angiotensin-converting enzyme 2 receptor (ACE2) (Islam et al. 2020). ACE2 is a crucial factor for the mechanism of COVID-19 infection into the human body (Islam et al.

2020). On the other hand, the extent of COVID-19 infection has manifested beyond localized symptoms. The third factor (PC3) of PCA illustrates fever (CS1), difficulty breathing (CS8), and vomiting (CS12) as significant variables with the loading values of 0.6456, 0.7064, and 0.6718, respectively. These three clinical symptoms may represent bodily response toward an infection that has turned systemic. There is emerging evidence that COVID-19 leads to multi-organ dysfunction through endothelial injury, a mechanism conventionally associated with a vascular pathology (Siddiqi et al. 2021).

Nonetheless, it is possible COVID-19 may only cause localized or predominantly respiratory symptoms. The PC4 summarizes sore throats (CS7) and difficulty breathing (CS8) as strong factors, with loading values of 0.8513 and 0.6824, respectively. Lung tissue damage caused by COVID-19 is evident by the autopsy of a 50-year-old patient from the initial case in Wuhan in 2019, which showed bilateral diffuse alveolar damage with cellular fibromyxoid exudates (WHO & China 2020). Difficulty breathing, however, is noticeably a significant factor for both PC3 and PC4. Clinically, both systemic and localized pathologies can cause breathing difficulty. Thus, it is conceivable that COVID-19, which may initially cause difficulty breathing primarily due to the assault to the respiratory system, may later cause the same clinical symptoms when it has gone systemic. The mixed significant factor loading of difficulty breathing to PC3 and PC4 thus reflects the dynamic and complexity of COVID-19 clinical presentations attributable to this symptom.

5.4 Validation of risk index for pilot study based on discriminant analysis (DA)

Discriminant Analysis (DA) provides a further statistical estimation on the index variations of clinical aspects and Epid-Link (Table 3). Based on PCA's risk indices, the DA evaluates the most significant discriminant functions of variables from the clinical aspects (clinical symptoms and clinical history) or Epid-Link. The risk indices were treated as dependent variables, and the clinical aspects (clinical symptoms and clinical history) and Epid-Link variables were treated as independent variables. The discriminant functions (DFs) and classification matrices (CMs) obtained from the DA standard stepwise mode and DA forward stepwise mode are depicted in Fig. 3a and b.

The standard DA mode constructs DFs that contained all clinical aspects (clinical symptoms and clinical history) and epidemiological link of COVID-19 variables. The high-risk indices were 92.31% discriminated correctly (Table 3) with different significant clinical aspects and epidemiological variables. Table 3 displays the discriminant analyses that 100% correctly assign categories of Low Risk and Moderate Risk. Three variables appear to be strong discriminants

Table 3 Confusion Matrix of DA (Standard Step Wise and Forward Step Wise) validating the percentage correction of the risk indices

Index scale	Predicted Index Scale			Total	% correct
	High risk	Low risk	Moderate risk		
DA standard step wise					
High risk	60	4	1	65	92.31%
Low risk	0	114	0	114	100.00%
Moderate risk	0	0	21	21	100.00%
Total	60	118	22	200	97.50%
DA forward step wise					
High risk	39	15	6	60	65.00%
Low risk	2	116	0	118	98.31%
Moderate risk	1	0	21	22	95.45%
Total	42	131	27	200	88.00%

toward the high-risk category. They are the last contact with confirmed cases in days (CH14), duration of exposure with confirmed cases in minutes (CH13), and age (S1).

In standard mode, the 200 respondents act as discriminant variables, which the clustering matrix of DA correctly assigns categories at 97.50% accuracy. In contrast, the DA forward stepwise mode worked on Low and Moderate Risks' matrices with nine (9) discriminant variables, responsible for 98.31% and 95.45% correctness of the CMs. Through this study deliberation, the clinical aspects and epidemiological variables of CS11 [Duration of muscle pain/ fatigue (days)], CS9 [Duration of difficulty breathing (days)], and S1 (Age) are attributable to Moderate Risk of COVID-19. However, CM produced 88.00% correct assignments in the entire stepwise forward mode that used 200 discriminant parameters with different clinical aspects and epidemiological links to COVID-19. The matrix of High Risk, however,

yielded the lowest at 65% percentage correctness of CMs. These results are notably dissimilar to those obtained from the DA standard mode.

The Wilks' Lambda test for standard mode and forward modes revealed a Lambda value of 0.025 ($p < 0.0001$), and 0.181 $p < 0.0001$, respectively. The DA null hypothesis, H_0 , states that the mean vectors were equal among the three classes. The alternative hypothesis, H_a states that at least one of the mean vectors is different from the others. As the computed p value is lower than the significance level alpha (0.05), this study rejects the null hypothesis, H_0 , and accepts the alternative hypothesis, H_a . The risk of rejecting H_0 is true when the result is lower than 0.05. Thus, the discriminant analysis suggests that the included variables were sufficient to discriminate the three groups and accounted for most of the index scale variations of COVID-19 infections. These variables were the last contact with confirmed cases (days)(CH14), duration of exposure with confirmed cases (minutes) (CH13), age (years) (S1), duration of muscle pain/ fatigue (days) (CS11), and duration of difficulty breathing (days) (CS9).

5.5 Improvement

5.5.1 Validation of CHaSe system using live data of 5000 respondents

A real mass-screening using the CHaSe system was conducted among 5000 university staff and students from the pilot study analysis phase. The modeling employed in the CHaSe system comprises fifteen (15) questions as outlined in Table 4 below:

After consideration was made based on these 5000 respondents' actual responses, the system's modeling was revalidated. The association of the clinical and Epid-Link parameters was reduced to eleven (11) variables. To

Fig. 3 a DA Standard step mode. b DA Forward step mode of COVID-19 for 200 respondents

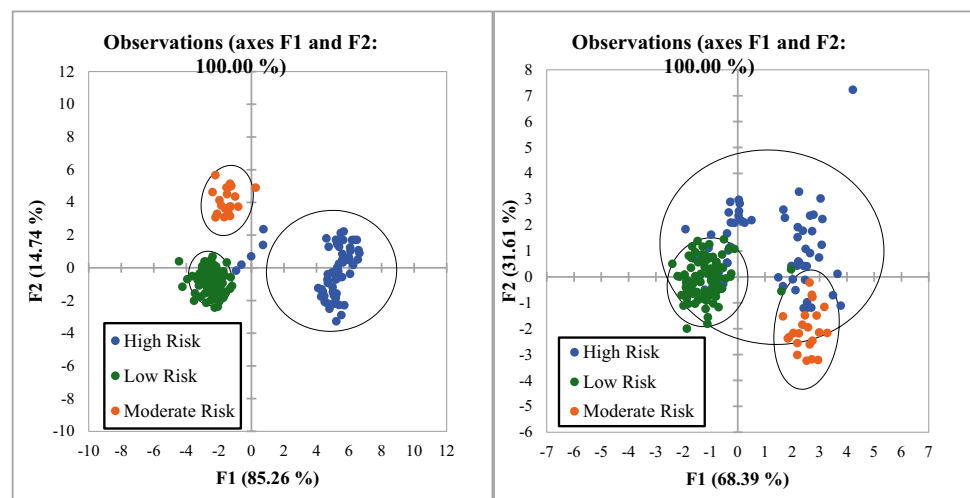


Table 4 List of Questionnaire distributed to University students and staff

No.	Category	Variables
1	Clinical history	History of joining high-risk gathering (where confirmed cases had been recorded)
2	Clinical history	History of contact with confirmed cases (Close contact)
3	Clinical history	Duration of exposure with confirmed cases (in minutes)
4	Clinical symptoms	Fever
5	Clinical symptoms	Dry cough
6	Clinical symptoms	Sore throats
7	Clinical symptoms	Difficulty breathing
8	Clinical symptoms	Muscle pain/fatigue
9	Clinical symptoms	Vomiting
10	Clinical symptoms	Headache
11	Co-morbidities	Diabetes
12	Co-Morbidities	Hypertension
13	Co-morbidities	Ischemic Heart Disease
14	Co-morbidities	Heart failure
15	Co-morbidities	Respiratory disease (e.g., asthma, COPD, bronchiolitis etc.)
16	Co-morbidities	Stroke/TIA

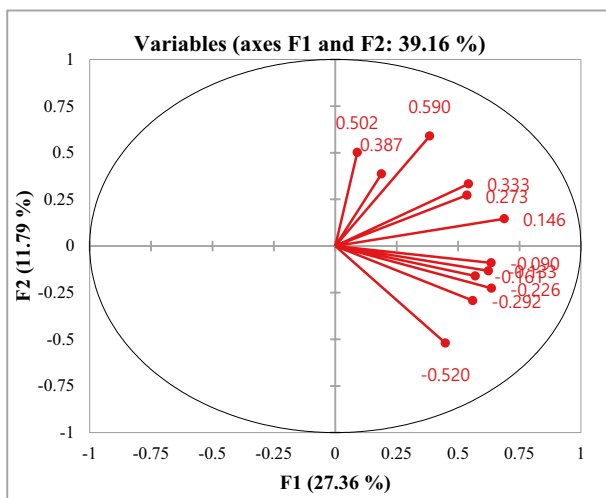


Fig. 4 Variance explanation of the strong factor loadings of 27.36% for PC1 and 11.79% for PC2 after varimax rotation of 11 parameters for 5000 respondents

re-confirm the risk level of COVID-19 (Low Risk, Moderate Risk and High Risk), we conducted a thorough iterative process. These include interactive machine learning algorithm (the weightage scoring), simulation of the modeling, and predictive modeling integration into the CHaSe System with live raw data of 5000 respondents using the matrices index of PCA (Fig. 4). The medical experts revalidated all the initial eleven parameters of COVID-19 before re-confirming the weightage score for categorizing the index using the PCA based on 5000 respondents. Interestingly, the revalidated PCA for 5000 respondents has also yielded twelve significant PCs with an eigenvalue greater than 1 (Table 3;

Fig. 3) for a total of the explained cumulative variance of 58.288%. Importantly, the CHaSe system’s modeling underwent repeated revalidation by medical experts for knowledge accuracy of all the clinical aspects and the Epid-Link (clinical symptoms and clinical history). After repeated revalidation by medical experts, the PCs’ variance explanations are 27.36% for PC1, 11.79% for PC2, 10.347% for PC3, 8.785% for PC4, with a cumulative explanation of 58.288% in data variability. Figure 4 indicated twelve factor loadings in the PCA accountable for 58.288% (Fig. 3) of the clinical symptoms and clinical history variations. Therefore, these results demonstrate the initial clinical aspects and Epid-Link variables yielded from PCA of the pilot study were validated by further analyses of the real mass-screening among 5000 respondents (Niranjanamurthy et al. 2020; Cao et al. 2020) (Figs. 5 and 6; Tables 5 and 6).

After model revalidation, the CHaSe System of the COVID-19 screening was accessible by students and the civil servants via the internet browser or mobile application for self-monitoring (Fig. 7a). As part of the organizational policy to prevent COVID-19 outbreaks, it was made mandatory for the civil servants and students to undergo daily screening through the CHaSe system throughout the national Movement Control Order (MCO). The CHaSe system administration will monitor the updated data, and the outbreak details were illustrated as shown in Fig. 7b.

The administration of the CHaSe system prompted the mapping of risk trends covering the entire states of Malaysia. Figure 8 portrays the COVID-19 data outbreaks that comprise Risk Trend over Time, Risk Level of Malaysia as a whole, and Risk by States. This mobile application of COVID-19 is the first online monitoring developed in

Fig. 5 The development of Chase System with the integration of raw data (clinical aspects, data analytics, and modeling

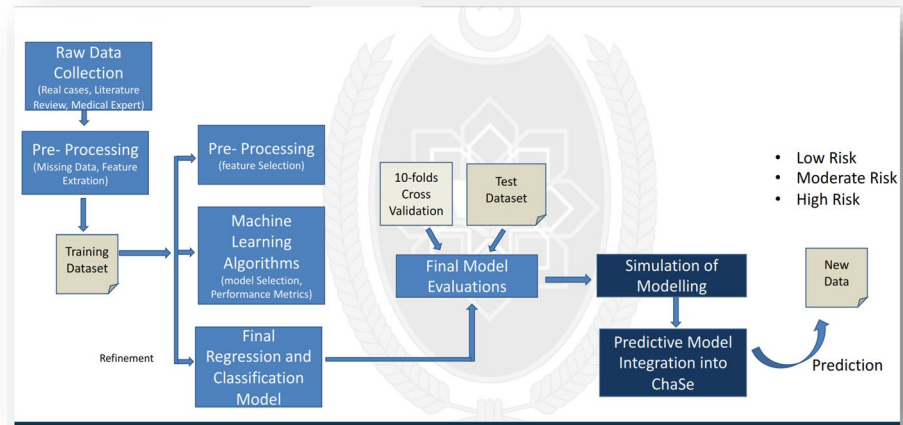


Fig. 6 **a** DA standard stepwise mode of the pilot study for 200 respondents, and **b** DA stepwise forward mode of the pilot study for 200 respondents

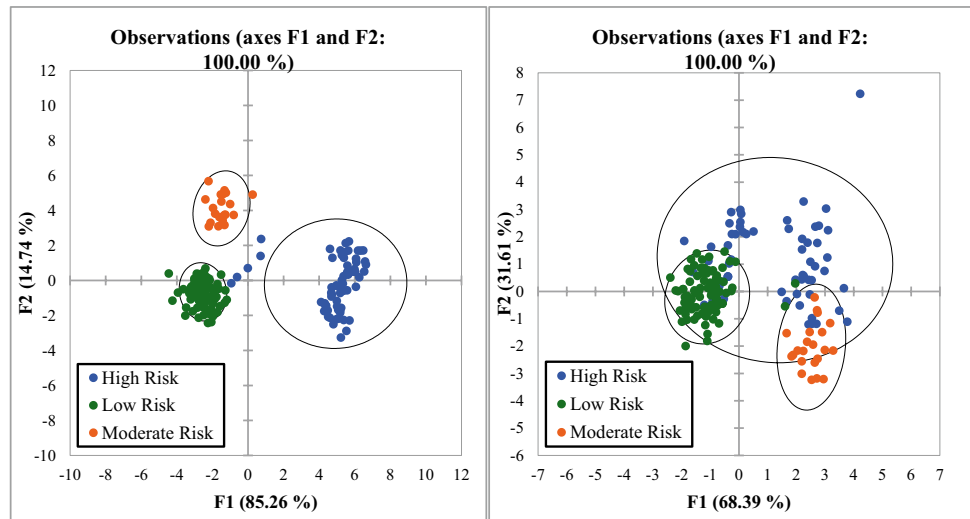


Table 5 Eigenvalues from the principal component analysis illustrating variance, cumulative variance, and factor loading of clinical aspects (clinical symptoms and clinical history) and Epid-Link after varimax rotation of 11 parameters for 5000 respondents

Variables	D1	D2	D3	D4
S1	-0.071	0.848	0.005	0.079
CH6	0.705	-0.013	0.274	-0.064
CH11	0.707	-0.020	0.048	0.178
CH13	0.509	0.167	-0.475	0.044
CS1	0.225	-0.203	0.646	0.192
CS3	-0.085	0.020	0.099	0.851
CS7	0.278	0.069	0.180	0.682
CS8	0.091	0.334	0.706	0.195
CS11	0.444	0.330	0.278	0.370
CS12	0.152	0.080	0.672	0.147
CS15	0.463	-0.115	0.209	0.491
Com2	0.418	0.558	0.213	-0.215

Malaysia based on the integration of computer science, data analytics (algorithms on the weightage score for index category) and medical factors based on actual respondents.

5.6 Control

The Six Sigma approach adopted in this study serves as a problem-solving tool to address the critical need to conduct validated mass-screening when resources are limited (Juahir et al. 2017). As discussed in the define phase, this study identifies the crucial gap in achieving a valid mass screening to mitigate the risk of cluster outbreak upon the return of staff and students post MCO. This objective is a form of control measure in evaluating if the CHaSe system meets its purpose. Through the CHaSe system, this study completed a mass-screening of 5000 university students and staff. Robust active case detection and contact tracing ensure that respondents that have a high risk of contracting

Table 6 Weightage Scoring of 200 respondents (pilot study) and 5000 respondents (actual data) of the development of CHaSe System

No	Section	Var	Description	Scoring				
				Model-49 V	Model-11 V	Model-9 V	Model-11VR	Model-9VR
1	Sociodemographic	S1	Age (Years)					
2		S2	Gender					
3		S3	Race					
4		S4	Place of living (State)					
5		S5	Type of residency					
6		S6	Locality					
7		S7	Occupation					
8		S8	At risk occupation					
9	Clinical history	CH1	Use of mask	0.8088161				
10		CH2	Use of sanitizer	0.8327972				
11		CH3	Practice of social distancing (More than 1 m apart)	0.750417				
12		CH4	History of involving with mass gathering of more than 50 people since last two weeks	0.5660143				
13		CH5	History of family members/co-workers involved with mass gathering of more than 50 people since last two weeks	0.7675449				
14		CH6	History of joining high-risk gathering (where confirmed cases had been recorded)	0.4602594	9.788	11.213	0.69670113	0.6113535
15		CH7	History of Visit to any Clinic or Hospital	0.5890887				
16		CH8	International Travel History	0.5732973				
17		CH9	***Dynamic Variable—Travel to and from high-density state					
18		CH10	***Dynamic Variable—Travel to and from high-density countries	0.7130434				
19		CH11	History of contact with confirmed cases (Close contact)	0.703928	10.06		0.83649972	
20		CH12	History of contact with close contacts (Casual contacts)	0.6710848				
21		CH13	Duration of exposure with confirmed cases (in minutes)	0.7170683	6.827		0.70039132	
22		CH14	Last contact with confirmed cases (in days)	0.7552337				

Table 6 (continued)

No	Section	Var	Description	Scoring				
				Model-49 V	Model-11 V	Model-9 V	Model-11VR	Model-9VR
23	Clinical symptom	CS1	Fever	0.8420126	8.702	11.669	0.53846818	0.6375996
24		CS2	Duration of fever (Days)	0.8713695				
25		CS3	Dry Cough	0.824329	4.991	6.708	0.46676926	0.5974462
26		CS4	Duration of dry cough (Days)	0.8714863				
27		CS5	Productive cough	0.3378407				
28		CS6	Runny nose	0.7992252				
29		CS7	Sore throats	0.5241973	9.122	11.862	0.59975037	0.5819468
30		CS8	Difficulty breathing	0.4738243	9.906	14.113	0.58112963	0.6294568
31		CS9	Duration of difficulty breathing (Days)	0.6155505				
32		CS10	Muscle pain/fatigue	0.4471942	8.146	11.732	0.58942573	0.5798006
33		CS11	Duration of muscle pain/fatigue (Days)	0.3949501				
34		CS12	Vomiting	0.7388543				
35		CS13	Diarrhea	0.6960631				
36		CS14	Conjunctival congestion	0.4245814				
37		CS15	Headache	0.3963389	9.057	11.446	0.75070043	0.4475942
38		CS16	Abdominal pain	0.6446945				
39	Comorbidity	Com1	Diabetes	-0.5719882				
40		Com2	Hypertension	0.4040926	7.807	10.443	0.59859511	0.6292796
41		Com3	Ischemic heart disease	0.7606792				
42		Com4	Malignancy	0.7835047				
43		Com5	Heart failure	0.6845714				
44		Com6	Respiratory disease (e.g., asthma, COPD, bronchiolitis, etc.)	0.6232216	7.691	10.388	0.62548257	0.5671423
45		Com7	Smokers	0.4397769				
46		Com8	Renal disease	0.5377361				
47		Com9	Low Immunity (e.g., HIV, on long-term steroid, on chemotherapy, on immunospressant, etc.)	0.7867499				
48		Com10	Stroke/TIA	0.6253828				
49		Com11	Chronic liver disease	0.6650013				

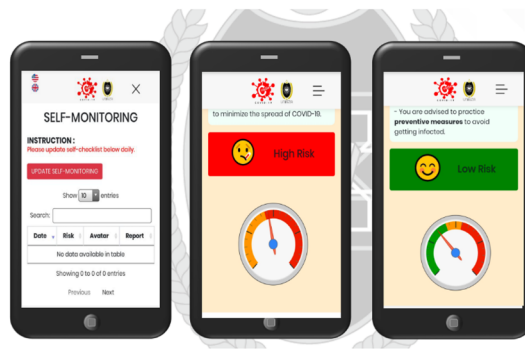
Model-49V represents the modeling of initial forty-nine (49) variables based on literature review, whereas Model-11V and Model-9V represent eleven and nine variables based on actual 5000 respondents. Both Model-11VR and Model-9VR were also based on 5000 respondents, but the front liners medical expert revalidated the model to procure the optimal accuracy, reliability, and efficiency of the CHaSe system.

COVID-19 be identified. Without the advanced application of modern internet of things (IoT) technology and prediction model based on data analytics (Aris et al. 2010), the traditional practice warrants enormous human resources to initiate one-to-one contact for personalized risk assessment. China shared this experience on how they managed to contain COVID-19 when it first emerged in Wuhan in November 2019 (WHO & China 2020). The experience from that exercise was described as “*painstaking*” (WHO & China 2020) and practically not achievable with our resources, given the pandemic's current scale of impact. Hence, by facilitating early detection for many respondents within the university's current capacity, the CHaSe system meets and exceeds

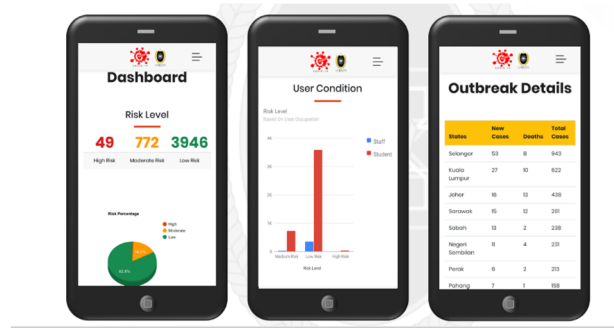
public health practice standards. Therefore, the CHaSe system demonstrates the significant scale of clinical outcomes attained by harvesting the benefits of the mobile technology.

However, the screening model based on self-declaration of clinical history and symptoms has its limitations. There were reports on asymptomatic infection of COVID-19 (WHO & China 2020; Kluster Tawar 2020; Noor Hisham Abdulah 2020). Unfortunately, reports on the prevalence of asymptomatic infected individuals are conflicting (WHO & China 2020; Noor Hisham Abdulah 2020; McMichael et al. 2020). Additionally, there is a need to assign caution on interpreting ‘asymptomatic infection’ from cross-sectional data since patients may not develop symptoms during the

Fig. 7 a A self-monitoring CHaSe System of COVID-19 for health closed-surveillance of individuals. **b** A COVID-19 closed-surveillance monitored by the administration embedded in CHaSe System



a A self-monitoring CHaSe System of COVID 19 for health closed-surveillance of individuals



b A COVID 19 closed-surveillance monitored by the administration embedded in CHaSe System

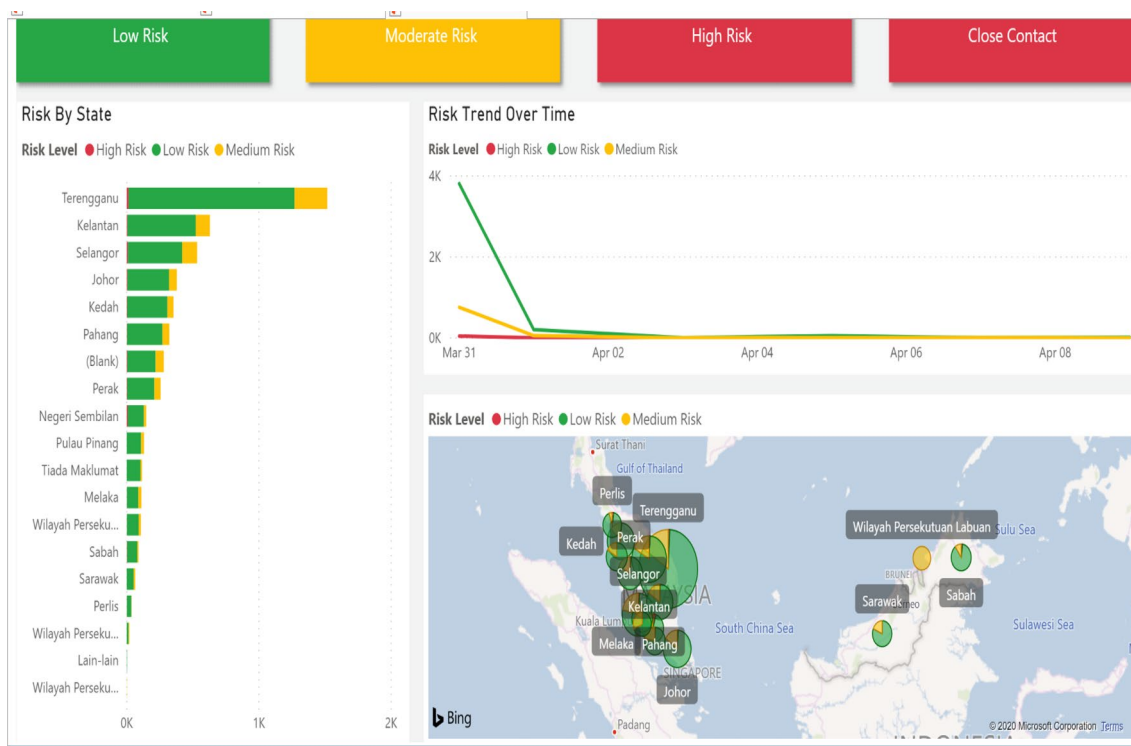


Fig. 8 Visual data of COVID-19 outbreaks, comprising of risk trend over time, risk level in entire Malaysia and risk by states

diagnostic test but may consequently develop them as the disease progresses. Thus, asymptomatic infection is defined as lack of symptoms since exposure until complete recovery. These challenges motivate us to advocate daily screening with validated models via the CHaSe system for our students and staff to increase screening sensitivity further.

Therefore, the technology-enhanced screening via the CHaSe system does not just make possible a one-time COVID-19 screening. The use of mobile technology also enables regular risk assessment as a surveillance strategy for a large population. The ability to conduct risk assessments and further personalized clinical evaluations through daily follow-up phone contact for high-risk groups serves as a robust and comprehensive screening and surveillance strategy to reduce the risk of cluster outbreaks. Nonetheless, asymptomatic individuals with no known Epid-Link were also reported among those diagnosed with COVID-19 in Malaysia (Kluster Tawar 2020; Noor Hisham Abdulah 2020; MOH 2020a, b). Hence, the CHaSe system is not replacing the existing public health measures to control the pandemic. Instead, CHaSe serves to complement and enhance the current strategy for screening and surveillance for COVID-19. On the system side, CHaSe has been developed into two main applications: web-based system and mobile apps (android and IOS). The technologies employed include MySQL as the Database Management System (DBMS) and Laravel Framework (PHP) for the web-based programs. The rules extracted from experiments using various machine learning techniques (PCA for dimensional reduction, features selection) as well as classification and regression methods were translated into PHP programming. Pilot testing conducted to verify the results and the rules generated from the computed models were observed to be consistent with the result by obtained from the medical experts.

6 Conclusion

COVID-19 outbreak has rapidly escalated to become a pandemic within months of its emergence. The easy-spread and scale of the disease's impact demand medical experts and policy-makers to establish a valid screening model that can be applied effectively and efficiently to the mass population. Comprehensive conventional active case detection and contact tracing are effective in containing the outbreak. Nonetheless, these measures require large supplies and resources which are not always immediately accessible to government and private organizations. In Malaysia, students and staff of higher educational institutions in Malaysia will return to campus at the end of the movement restrictions policy to invigorate the economy and educational activities. Thus, there is an urgent need for educational institutions to conduct

mass screening and surveillance as part of the organization's initiative to mitigate the risk of COVID-19 cluster outbreaks.

We have developed a robust risk-prediction modeling within the CHaSe system by means of a collaboration between medical, data and computing experts to address this need. DMAIC Six Sigma approach was applied as a systematic and critical-to-quality (CTQ) approach to achieving this goal. In this study, we report the vigorous iterative process of how the risk-prediction modeling was comprehensively developed and then revalidated by front liner medical experts to procure optimal accuracy, reliability, and efficiency.

Combined with the mobile technology, the results demonstrate the feasibility of mass screening 5000 students and staff within the capacity of our university's resources. Furthermore, the analysis provides risk trend over time, risk level of the entire country and risk by states. Thus, the modeling system within CHaSe serves as a valuable approach to complement and enhance the existing public health measures against COVID-19.

Acknowledgements The authors wish to thank the Office of the Prof. Dato' Dr. Hassan Basri, Vice Chancellor of UniSZA, Office of the Vice Chancellor, East Coast Environmental Research Institute (ESERI), UniSZA, and other CHaSe development members: Assoc. Prof. Dr. Aniza Abd Aziz, Dr. Nor Iza Abdul Rahman, Prof. Dr. Nyi Nyi Naing @ Syed Hatim Noor, Dr. Nor Kamaruzaman Esa, Dr. Ras Azira Ramli, Dr. Syadiah Nor Wan Shamsuddin, Dr. Elissa Nadia Madi, Dr. Suhailan Safei, Ahmad Aiman Shallahuddin, Mohammad Hariz Hasni, Muhammad Kamal Arif Razali, Mohd Masrul Ibrahim, Afiq Nasri Othman, Assoc. Prof. Dr. Mohd Ariff Sharifudin and Dr. Rabiatal Adawiyah Abdul Rohim. This study was funded by the Ministry of Science, Technology and Innovations (MOSTI), Ministry of Primary Industries and Commodities (MPIC) and UniSZA Internal Fund with project numbers of UniSZA/202/PPL/MTC (008), R0270 & R0271; and was approved by Universiti Sultan Zainal Abidin Human Research Ethics Committee (UHREC) (UniSZA/UHREC/2020/173).

References

- Aris AZ, Abdullah M, Praveena S, Yusoff M, Juahir H (2010) Extenuation of saline solutes in shallow aquifer of a small tropical island: a case study of Manukan Island. *North Borneo Environ* 3:84–92. <https://doi.org/10.14456/ea.2010.45>
- Balilla J (2020) Assessment of COVID-19 mass testing: the case of South Korea. *SSRN Electron J*. <https://doi.org/10.2139/ssrn.3556346>
- Cao B et al (2020) A trial of lopinavir-ritonavir in adults hospitalized with severe Covid-19. *N Engl J Med* 382:1787–1799. <https://doi.org/10.1056/NEJMoa2001282>
- Center for Disease Control and Prevention (2021) COVID-19—Older adults. <https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/older-adults.html>
- Cloutier V, Lefebvre R, Therrien R, Savard MM (2008) Multivariate statistical analysis of geochemical data as indicative of the hydrogeochemical evolution of groundwater in a sedimentary rock aquifer system. *J Hydrol* 353:294–313. <https://doi.org/10.1016/j.jhydrol.2008.02.015>
- Cox I, Gaudard M, Ramsey P (2016) *Visual Six Sigma: making data analysis lean*. John Wiley & Sons Inc, Hoboken

- Dominick D, Juahir H, Latif MT, Zain S, Aris AZ (2012) Spatial assessment of air quality patterns in Malaysia using multivariate analysis. *Atmos Environ* 60:172–181. <https://doi.org/10.1016/j.atmosenv.2012.06.021>
- Fahmi M et al. (2011) River water quality modeling using combined principle component analysis (PCA) and multiple linear regressions (MLR): a case study at Klang River, Malaysia. *World Appl Sci J* 14
- Fazillah A et al (2017) Evaluation of socioeconomic status on drug addicted person. *Int J Adv Sci Eng Inf Technol* 7(6):2122. <https://doi.org/10.18517/ijaseit.7.6.1009>
- Gazzaz NM, Yusoff MK, Ramli MF, Aris AZ, Juahir H (2012) Characterization of spatial patterns in river water quality using chemometric pattern recognition techniques. *Mar Pollut Bull* 64:688–698. <https://doi.org/10.1016/j.marpolbul.2012.01.032>
- Helena B et al (2000) Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by principal component analysis. *Water Res* 34:807–816. [https://doi.org/10.1016/S0043-1354\(99\)00225-0](https://doi.org/10.1016/S0043-1354(99)00225-0)
- Islam N, Khunti K, Chowell G (2020) RAAS Inhibitors and Risk of Covid-19. *N Engl J Med* 383:1990–1994. <https://doi.org/10.1056/NEJMc2030446>
- Ismail S et al (2018) Attentional process during listening to quantitative quranic verses (Fatihah Chapter) associated with memory, speech and emotion. *Asian J Med Biomed* 2:36–43
- Juahir H et al (2017) Improving oil classification quality from oil spill fingerprint beyond six sigma approach. *Mar Pollut Bull*. <https://doi.org/10.1016/j.marpolbul.2017.04.032>
- Kamaruddin A et al (2015) Spatial characterization and identification sources of pollution using multivariate analysis at Terengganu River Basin. *Malaysia Jurnal Teknologi* 77:269–273. <https://doi.org/10.11113/jt.v77.4054>
- Kannel PR, Lee S, Lee YS, Kanel SR, Khan SP (2007) Application of water quality indices and dissolved oxygen as indicators for river water classification and urban impact assessment. *Environ Monit Assess* 132:93–110. <https://doi.org/10.1007/s10661-006-9505-1>
- Lambrakis N, Antonakos A, Panagopoulos G (2004) The use of multicomponent statistical analysis in hydrogeological environmental research. *Water Res* 38:1862–1872. <https://doi.org/10.1016/j.watres.2004.01.009>
- Liu C-W, Lin K-H, Kuo Y-M (2003) Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan. *Sci Total Environ* 313:77–89. [https://doi.org/10.1016/S0048-9697\(02\)00683-6](https://doi.org/10.1016/S0048-9697(02)00683-6)
- McMichael TM et al (2020) Epidemiology of Covid-19 in a long-term care facility in king county. Washington. <https://doi.org/10.1056/NEJMoa2005412>
- Microsoft Inc. (2018) Overview of security and compliance in Microsoft Teams. <https://docs.microsoft.com/en-us/microsoftteams/security-compliance-overview>
- Ministry of Health Malaysia (2020) Update on COVID-19 status and statistics. <<https://www.bharian.com.my/berita/nasional/2020/05/684741/covid-19-kematian-warga-emas-berumur-61-hingga-70-tahun-paling-tinggi>
- MOH (2020) Kluster Tawar: Kluster aktif terbesar dan strain D614G. <http://covid-19.moh.gov.my/sorotan/kluster-tawar-kluster-aktif-terbesar-dan-strain-d614g>
- MOH (2020a) Official Facebook of The Ministry of Health Malaysia—Kluster Tawar. <https://www.facebook.com/kementeriankesihatanmalaysia/posts/10157339198421237>
- MOH (2020b) Official Facebook of The Ministry of Health Malaysia—Kluster PUI Sivagangga. <https://www.facebook.com/kementeriankesihatanmalaysia/posts/10157286878531237>
- Niranjanamurthy M, Amulya MP, Dayananda P, Pradeep HG (2020) Coronavirus—COVID-19 before and after solution through web application and app. *Int J Adv Sci Technol* 29:27–41
- Noor Hisham Abdullah (2020) Alhamdulillah, pada hari ini 12 April 2020. https://m.facebook.com/story.php?story_fbid=3201111949912671&id=578835322140360&__tn__=R
- SCCE JHU (2020) Coronavirus Resource Centre, <https://coronavirus.jhu.edu/map.html>
- Shrestha S, Kazama F (2007) Assessment of surface water quality using multivariate statistical techniques: a case study of the Fuji river basin. *Jpn Environ Model Softw* 22:464–475. <https://doi.org/10.1016/j.envsoft.2006.02.001>
- Siddiqi HK, Libby P, Ridker PM (2021) COVID-19—a vascular disease. *Trends Cardiovasc Med* 31:1–5. <https://doi.org/10.1016/j.tcm.2020.10.005>
- WHO & China (2020) Report of the WHO-China joint mission on coronavirus disease 2019 (COVID-19). World Health Organization, Geneva

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Azimah Ismail¹ · Mohd Salami Ibrahim² · Samhani Ismail² · Azwa Abdul Aziz³ · Harmy Mohamed Yusoff² · Mokhairi Mokhtar³ · Hafizan Juahir¹

¹ East Coast Environmental Research Institute (ESERI), Universiti Sultan Zainal Abidin (UniSZA), Gong Badak Campus, 21300 Kuala Nerus, Terengganu, Malaysia

² Faculty of Medicine, UniSZA, Universiti Sultan Zainal Abidin (Medical Campus), Jalan Sultan Mahmud, 20400 Kuala Terengganu, Terengganu, Malaysia

³ Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin (UniSZA), Gong Badak Campus, 21300 Kuala Nerus, Terengganu, Malaysia