Proceedings of Science and Mathematics
Faculty of Science, Universiti Teknologi Malaysia

Vol. 10, 2022, page 71 - 79

Logistic Regression on Covid-19 Patient with Their Symptoms

Farah Nur Amira Aziz, Haliza Abd Rahman*
Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia
*Corresponding author: halizarahman@utm.my

Abstract

This research aims to analyze Covid-19 stages in relation with the symptoms related to the disease. Coronavirus 2019 (COVID-19) is the greatest communicable disease to have hit Malaysia and with many patients having higher stages of Covid-19 especially those with comorbidities. Aside from that, even those who survive from the disease have after effects such as breathing difficulty and tiredness. Some common symptoms of the disease are fever, dry cough, and shortness of breath; in mild cases, people may merely experience a headache or chest pain. In the most severe cases, infected people have trouble breathing and may eventually develop organ failure. Infection with severe acute respiratory syndrome coronavirus appears to put older patients with comorbidities at the greatest risk of developing severe illness Covid-19. Therefore, this research's findings may help people that live in this endemic to better understand which factors contribute to the worsening of the illness. Method used is binary logistic regression. The data used is the 2022 Covid-19 patient data from respondents in the community. Possible factors affecting the stages of their illness have been grouped into demographic factors which are gender, race, and age, occupation and medical background which are smoking status, Covid-19 status, stages of Covid-19, type of vaccine, hospital admission, quarantine center admission, ICU admission, type of symptoms and type of comorbidities to Covid-19 patient. Result shows that factors such as Malay ethnicity, smokers, admitted patients and patients with fever symptoms are more likely to contribute to the variation in Covid-19 stages.

Keywords: Symptoms; Stages of Covid-19; Logistics Regression

1. Introduction

Since the first known case of coronavirus disease 2019 (COVID-19) was announced in China in December 2019, the world has seen the largest single impact to social and global wellbeing [1]. Research has indicated that since December 2019, it has spread throughout the world. COVID-19 appears to have a broad clinical spectrum, ranging from asymptomatic infection to moderately, seriously, and critically affected people [2].

Moreover, corona virus infectious disease (COVID-19) pandemic has resulted in high death, severe acute respiratory failures, and, as a result, overcrowding in emergency departments (EDs) due to insufficient laboratory testing capacity [3]. The creation of decision support systems for real-time clinical identification of COVID-19 is critical in assisting patients with categorization and allocating resources to those who are at risk. In the context of the COVID-19 response, health workers may be exposed to workplace diseases that put them at risk of sickness, injury, and even death. Besides, in the article of World Health Organization, they examine the chance of coming into direct, indirect, or close contact with a virus-infected person can be predicted [4]. Meanwhile, this is not the first-time massive death occurs in worldwide since in China each year have reported heart disorders which account for more than 1,500,000 deaths, making it the country's biggest cause of mortality [5]. As the results, prevalence of disease can cause total loss of financial, well-being and family members.

Furthermore, in the study of a multi-center retrospective, it encompassed COVID-19 patients classified as being moderately, severely, and critically ill. The National Health Commission (V.5) defined the illness severity of COVID-19 based on the Guidelines on the Diagnosis and Treatment of COVID-19. It shows, the presence of at least one comorbidity in 51.54 percent of the patients. Hypertension was the most common comorbidity (33.90 percent), followed by diabetes (13.18 percent) and cardiovascular disease (6.51 percent). There were only a few cases with a current (7.53 percent) or previous (2.57 percent) smoking habit. Fever (81.22 percent) and dry cough (33.63 percent) were the most prevalent symptoms on admission, followed by sputum production (30.77 percent) and shortness of breath (26.65 percent). When comparing the critical (92.45 percent) and severe (95.24 percent) groups, chest CT scans revealed a significantly higher percentage of bilateral lungs involvement in the critical (92.45 percent) and severe (95.24 percent) groups (80.07 percent) [2].

However, “An Introduction to Logistic Regression Analysis and Reporting”, studied about the execution of the investigation. They wrote that a mathematical approach that will be used to calculate and analyze the data. That mathematical approach is logistic regression. Generally, logistic regression is well suited for describing and testing hypothesis about relationships between a categorical outcome variable and one or more categorical or continuous predictor variables. Logistic regression models have a high ability in the diagnosis of COVID-19 infection because it use for binary dependent variables such as they in mild stages or severe stages [6].

2. Literature Review

2.1. Aspect that Influences the Severity of Covid-19

Many case studies and articles about Covid-19 have been written, with topics such as the impact of Covid-19 on the globe and simple statistics or data mining about current events [8]. Women and men are affected differently by pandemics and outbreaks. Individuals’ experiences are expected to vary depending on their biological and gender features, as well as their interactions with other social determinants, from the rising of exposure and biological susceptibility to infection to the social and economic implications, [8]. The study of Yupari et al., [9] shows the overall community is susceptible to this virus; nevertheless, males and those with comorbidities have been more apparent among those affected.

Besides, because government agencies normally reveal average Covid-19 data as summary statistics of patient demographics, this has been difficult to see the accurate results of coronavirus illness. These data may reveal differences in Covid-19 results across broad population groupings. but they do not allow for comparisons between more detailed population subgroups defined by several demographics.

Although elderly people infected with COVID-19 have a higher risk of death, infection rates measured by tests do not rise in synch with age. Children, teenagers, and other senior adults in their 60s and 80s had a greater test-based infection rate than the work force population. Latinos and African Americans have greater infection rates based on tests than other races and ethnicities. All-male groups with race as Asian, Latin X, African American, White and Multi-race have the highest 5-test-based case fatality rates, followed by African American females, demonstrating that African Americans are an especially vulnerable California subpopulation. Males and older people have much higher case fatality rates, according to gender and age-disaggregated national case data from a wide range of nations around the world [10].

However, there is no data on how the factors linked to Covid-19 mortality compare to those linked to death from other causes, and thus on the amount to which a person’s risk of dying from Covid-19 is likely to be driven by their overall mortality risk. We know that getting older is a huge risk factor for death from any cause. It’s possible that Covid-19 merely increases everyone’s chance of death by a constant factor, or that some things have a different impact on Covid-19 deaths. A greater understanding of this would aid in developing methods to identify and protect those who are most vulnerable to negative outcomes during the epidemic approach recently [11].

The symptoms of Covid19 infection are more like those of MERSCoV infection, as most confirmed MERSCoV cases have included fever, cough, shortness of breath, and nausea, vomiting, and diarrhea [12]. Diabetes, cancer, chronic lung disease, chronic heart disease, and chronic kidney disease are the most common underlying disease among MERSCov patients [13], while hypertension, diabetic, chronic obstructive pulmonary disease, coronary heart disease, and chronic renal disease were the most common underlying diseases among Covid-19 patients [12].

Besides, effective diagnosis using several symptoms or attributes of suspected patients is critical to helping health care systems. Until recently, a variety of models have been developed and published, ranging from rule-based scoring to complex machine learning models, [12]. The first coronavirus case was reported in Iran in February 2020. According to the World Health Organization's (WHO) most recent data, the global number of instances of coronavirus or Covid-19 infection has surpassed 63,000,000 persons and resulted in the deaths of over 1,466,000 people. Iran is responsible for more than 948,749 confirmed sick patients and 47,874 deaths [12].

In addition, people with underlying heart disease may be at a higher risk for serious infections and mortality. Tehran and Isfahan, both industrial and populated regions, have higher death rates than the others. They are frequently involved in environment issues such as air pollution, and they prevalence of pulmonary and heart ailment is higher in these regions, [12].

2.2. Mathematical Modelling of Covid-19 cases and severity

A mathematical model that describes the relationship between one or more independent variables and a qualitatively dependent variable is known as logistic regression. There are two or more categories in this dependent variable. A binary logistic regression model is one in which the dependent variable has two categories. The model is called a multinomial or ordinal logistic regression model if the dependent variable contains more than two categories. Other modelling approaches are also feasible, but the logistic model, which is estimated using a maximum like hood, is the most used. The logistic function, which specifies the mathematical from on which an exceedingly flexible and easily useful function is based, is what makes it so popular [14]

Recently, a new study intended to use logistic regression to uncover the factors of Covid-19 sickness severity based on ordinal replies. All the patients were split into three groups based on their severity of illness: moderate, severe, and critical. The following case characteristics on admission were potentially predictive variables demographic and epidemiological aspects, comorbidities, clinical signs and symptoms, laboratory findings, and chest imaging data [2]. Logistic regression is a machine learning classification technique that is widely utilized in clinical analysis [15]. In general, logistic regression is best suited to expressing and testing hypothesis concerning correlations between one or more categorical or continuous predictor variables [6]. In an epidemiological study with a moderate sample size typical of many studies with a limited number of incident occurrences and a limited set of simple clinical indicators, we showed that logistic regression performs as well as machine learning models in predicting the probability of major chronic diseases [16].

No longer after the research, other researchers did research on women, younger persons, and people with chronic diseases, according to their findings, who are at a higher risk of reporting mental health difficulties during COVID-19. One-quarter of the participants (25.1 percent) experienced severe depression symptoms, according to the findings. One-third (34.1 percent) experienced mild to moderate anxiety symptoms, while almost one-sixth (18.7 percent) had mild depressive symptoms. Age, gender, and virus- infected friends were the three most critical predictors of depression and anxiety [11].

Some variables were found to be associated with illnesss severity, including age, gender, hypertension, diabetes, interval between illness onset and diagnosis, interval between illness onset and admission, pharyngodynia, shortness of breath, and early antiviral administration. When patients aged 40–69 and 70+ years were compared to patients aged less than 40 years, the risks of having a more severe disease were 1.586 (95 percent CI: 0.824–3.053) and 3.419 (95 percent CI: 1.596–7.323) times greater in the age groups 40–69 and 70+ years, respectively. When compared to individuals without hypertension, patients with hypertension had a 3.372 (95 percent CI: 2.185–5.202) times higher chance of worsening illness severity [2].

2.3. Software and Type of Analysis

This analysis will use IBM SPSS Statistic to analyze the data. One technique to show a relationship between two categorical variables is to use a Chi-square statistic. Numerical (countable) variables and non-numerical (categorical) variables are the two types of variables in statistics, [10]. ANOVA is a statistical tool for determining whether there are differences between the two population means. To evaluate variation with the purpose of detecting possible variations among group means, the overall variation is divided into variation due to differences between groups and variations due to differences within groups using one-way ANOVA, [10]. In this study, logistic regression will be used to analyze symptoms severity and the effect of Covid-19. Software that will be used is IBM SPSS Statistic 27 and this thesis used binary logistic regression.

3. Methodology

3.1. Data Acquisition

A questionnaire was designed using Google Forms, and the link was sent through a social media group. Questions detailed in inquiring about their demographic, background and medical background and the identity of these survey participants will be kept confidential. The questionnaire has thirteen questions. In the Google Form, it explains the purpose of the survey and question. Next are questions about gender, age, race, occupation, smoking status, Covid-19 status, stages of Covid-19, type of vaccine, hospital admission status, quarantine center admission status, ICU admission status, type of symptoms, and type of comorbidities. This was done to gather sufficient and relevant data to address the study's research objectives. The data is collected by using Google Form as the platform and the link is distributed to the community platform. The survey was conducted using two languages, which are Malay and English. There are 77 respondents who have answered the Google form. This chapter will provide the overview of the data as well as the methods that will be used to achieve the purposes and goals of this study. Some of the approaches that will be discussed are general linear models, probability distributions, logistic models, sample data testing, and logistic regression output results.

3.2. Data Analysis

3.2.1. A Chi-Square Test for Independence

A Chi-Square test for independence is a test that is used to determine whether symptoms and stages of Covid-19 in a patient have an independent relationship. We made the following assumptions for the hypothesis:

H_0: Row and column variables are independent
H_1: Row and column variables are not independent

3.2.2. Analysis of Variance (ANOVA)

ANOVA is a statistical tool for determining whether there are differences between different population means. To evaluate variation with the purpose of detecting possible variations among group means, the overall variation will be divided into variation related to differences between groups and variation due to differences within groups using one-way ANOVA. Inside-group variation (SSW) is a measure of random variation within a group. The among-group variation (SSB) is a measurement of differences between groups. The number of values in all groups is represented by the symbol n, while the number of groups is represented by the symbol *c*.

4. Results and discussion

4.1. Overall Analysis

Most respondents were female Chinese, representing 48.0 percent of the total participants. This was followed by male Chinese, 24.0 percent; female Malay 22.67 percent, and male Malay and female Indian 2.67 percent each. Lastly male Indian with 0 percent. By race, Chinese represent 72.0 percent; meanwhile, Malays are 25.3 percent, followed by the Indians, which are 2.67 percent. As for gender-wise, most participants were female, consisting of 73.3 percent of the total, and males only represented 26.7 percent.

The highest data for respondents of age was from respondents of 13-39 years. The lowest results are respondents aged 60 and above. Respondents aged 40-59 years results with 31 respondents. Besides, the highest results of occupation are students with 27 out of 77 percent. Followed by non-government workers with 16 respondents; government workers with 13 respondents; respondents who did not work with 10 respondents. Furthermore, retired, and other occupations showed the same results with 5 respondents and factory workers with 1 respondent.

Smoking status of patients shows that 70 respondents out of 77 did not smoke, while 7 respondents in this survey did smoke. The results of stages of Covid-19 patient analysis which was the highest results go to respondents who have stage 1 and stage 2 of Covid-19, while others were respondents with stage 3 and stage 4.

The type of vaccine analysis of the 77 respondents. Pfizer vaccine shows the one that respondent took with 31 respondents, while Sinovac had 28 respondents that took the vaccine. Respondents that did not take the vaccine were lower than respondents that took the vaccine; this could be because they were not encouraged to take the vaccine because of their health condition, or they did not take the vaccine because of self-preference.

The result analysis of hospital, quarantine center, and ICU admission. It shows that quarantine center admission had the highest result for respondents that were admitted than in the hospital, while ICU results show that respondent that did not admit in ICU was the highest than in respondent not admitted to the hospital. Other than that, the symptoms are analyzed. There are 12 symptoms that have been analyze. The highest symptoms occurring by the respondent were fever and the lowest symptoms occurring by the respondents was abdominal pain.

The comorbidities result showed respondents with diabetes were 9 out of 77 while high blood pressure with 8 respondents. The results of low blood pressure and obesity were the same with 2 respondents. These results might be different if more respondents could be participating.

4.2. Test of Statistic

4.2.1. Chi-Square Test

Hospital admission (*p*-value = < 0.001), quarantine center admission (*p*-value = 0.049), fever (*p*-value = 0.004), and shortness of breath (*p*-value = 0.009) symptoms variables tested have the *p*-value less than 0.05. So, H_0 has not failed to be rejected. Hence, we can conclude there is a correlation between these variables and the stages of Covid-19 patients.

4.2.2. Analysis of Variance (ANOVA)

hospital admission (*p*-value = < 0.001), quarantine center admission (*p*-value = 0.050), fever (*p*-value = 0.004), shortness of breath (*p*-value = 0.009), variables tested have *p*-value less than 0.05. Hence, we can conclude there is a statistically significant difference between these variables to the stages of Covid-19 patients.

4.3. Stages Prediction Analysis

The univariate analysis shows that for confidence interval a = 0.05, the predictor hospital admission, x_{11}, fever, x_{14} and shortness of breath symptoms, x_{17}, has a significant p-value less than 0.05.

In hospital admission, x_{11}, variable, patient choose Yes is coded with '1' meanwhile patient choose No is coded with '0'. The exp(β) value shows that patients that admit in hospital are 8.000 at odds with patients that do not admit in hospital to have higher severity of Covid-19 stages. Taking the log of odds:

$$\ln (8.000) = 2.0794$$

Therefore, patients who are admitted to hospital are 2.0794 times more likely to have high severity stages of Covid-19 than those who do not. Fever, x_{14} variable shows the exp(β) value for patients who have fever symptoms are 14.000, at odds to patients who have not had fever symptoms. Taking the log odds:

$$\ln (14.000) = 2.6391$$

From the log of odds values, we can conclude that patients with fever symptoms have 2.6391 more tendencies to have high severity in stages of Covid-19 than those who do not have fever symptoms.

Lastly, shortness of breath symptom, x_{17} variable shows the exp(β) value for patient who has shortness of breath symptom is 4.444, at odds to patient who has not has shortness of breath symptom. Taking the log odds:

$$\ln (4.444) = 1.4916$$

From the log of odds values, we can conclude that patients with shortness of breath symptoms have 1.4916 more tendencies to have high severity in stages of Covid-19 than those who do not have shortness of breath symptoms.

The multivariate analysis in stages of Covid-19, the -2 Log Likelihood (-2LL) for all three methods have significant value, having less than 100 percent. Therefore, all three methods, Enter Method, Forward and Backward, show good fitting fitness. Nagelkerke's R-Square represents the proportional reduction in the proportional decrease in the absolute value of the Log-Likelihood. It adjusted the scale to the range from 0 to 1. For each method, based on Nagelkerke's R-Square value, the Enter Method is 84.30 percent; Forward 44.10 percent, and Backward 73.70 percent of the variation in the outcome variable for the probability to have severity in stages of Covid-19. Since Forward has below 50 percent Nagelkerke's R-Square value, Forward does not fit this dataset very well. Meanwhile, Enter Method and Backward methods more than 50 percent Nagelkerke's R-Square value, the two methods fit the data set very well. All methods are shown to have a p-value > 0.05 from the Hosmer and Lemeshow Test. The greater the p-value, the better the method matches to fit the model. Since the Backward has a greater p–value = 0.998 than the Enter Method p-value = 0.696, the Backward will be used to model the data.

Table 4.1 Backward method results

Variables in the equation								
	β	S.E	Wald	df	Sig.	exp(β)	Lower	Upper
Malay	3.587	1.223	8.608	1	0.003*	36.133	3.290	396.857
Smoking status	4.078	1.974	4.270	1	0.039	59.034	1.234	2624.804
Hospital admission	5.614	1.617	12.058	1	<0.001	274.316	11.535	6523.676
ICU admission	-23.804	40192.970	0.000	1	1.000	0.000	0.000	-
Fever	3.741	1.368	7.474	1	0.006	42.143	2.884	615.891
Hypertension	-26.145	21554.708	0.000	1	0.999	0.000	0.000	.
Diabetes	-24.520	9527.121	0.000	1	0.998	0.000	0.000	.
SLE	24.216	40192.969	0.000	1	1.000	3287	0.000	.
Constant	-6.754	1.690	15.967	1	<0.001	0.001	-	-

Based on Table 4.1, some results, just like in the univariate analysis, Malay patients, x_5 are more likely to be in high severity stages of Covid-19 than the other races. The Malay patients are 0.003 at odds with another race. Meanwhile, for patients who are smokers, x_9 are at 0.039 odds more than those who are not smokers. Other than that, patients who have admitted to the hospital, x_{11} *are more* likely to have high severity in stages of Covid-19 with <0.001 odds than those who have admitted. Lastly, patients with fever symptoms, x_{14} are 0.006 times better than patients who do not have fever symptoms.

From the values in Table 4.1, we can model:

$$\log \frac{\pi}{1-\pi} = \sum_1^n \quad b_i x_i$$

$$= \log \frac{\pi}{1-\pi} \qquad (1)$$

$$= 6.754 + 3.587x_5 + 4.078x_9 + 5.614x_{11} + 3.741x_{14}$$

Where π is the stages of Covid-19 patients, x_5 = race (Malay), x_9= smoker status (smoker), x_{11}= hospital admission (admit hospital), x_{14}= symptom (fever).

Therefore, the stages of Covid-19 patient are

$$\log \frac{\pi}{1-\pi} = 6.754 + 3.587x_5 + 4.078x_9 + 5.614x_{11} + 3.741x_{14} \qquad (2)$$

where,

$$\pi = \frac{\boldsymbol{exp}(\,6.754 + 3.587x_5 + 4.078x_9 + 5.614x_{11} + 3.741x_{14})}{\mathbf{1} + \boldsymbol{exp}(\,6.754 + 3.587x_5 + 4.078x_9 + 5.614x_{11} + 3.741x_{14})} \qquad (3)$$

Conclusion

Factors that affect the stages of Covid-19 patients are better visualized and understood. From the analysis, it reveals that Malay ethnicity, smoking patients, hospital admission, and symptoms of fever have high probability to affect the stages of Covid-19. From the results, hospital admission does not determine one's stages of Covid-19. It also applies to another categorical factor. Logistic regression shows that patients that have symptoms of fever are more likely to affect the severity of stages of Covid-19. In other words, patients with Covid-19 symptoms will determine the severity of the stages.

Given the ongoing global pandemic of COVID-19, this study will contribute to early identification of patients with high risk of developing critical illness and optimize the arrangement of health resources. This will help medical teams to identify the most severe symptoms to help patients treat their illness before it gets worse in the future. To improve this research, further studies should be taken with more patients of Covid-19 participation in the survey process. On that note, respondents' information should always be kept updated annually of their whereabouts. The information gathered would give us better insights into the current effect on stages of Covid-19. This research only shows the preliminary research of logistic regression of Covid-19 patients with their symptoms in Malaysia. This would assist the health authorities in managing and better facilitating patients of Covid-19.

Acknowledgement

The researcher would like to thank all the people who have supported the research. I wish to express my sincere appreciation to my supervisor, Dr Haliza binti Abd Rahman, for encouragement and guidance. My sincere appreciation also extends to all my colleagues and others who have helped at various occasions. Their views and tips are useful indeed. Lastly, I am grateful to all my family members for their support.

References

[1] Marzo, R. R., Vinay, V., Bahari, R., Chauhan, S., Ming, D. A. F., Fernandez, S., Fernandez, P. N., Johnson, C. C. P., Rahman, M. M., Goel, S. et al. Depression and anxiety in Malaysian population during third wave of the COVID-19 pandemic. Clinical epidemiology and global health, 2021. 12:100868.

[2] Xu, K., Zhou, M., Yang, D., Ling, Y., Liu, K., Bai, T., Cheng, Z. and Li, J. Application of ordinal logistic regression analysis to identify the determinants of illness severity of COVID-19 in China. Epidemiology & Infection, 2020.148.

[3] Saegerman, C., Gilbert, A., Donneau, A.-F., Gangolf, M., Diep, A. N., Meex, C., Bontems, S., Hayette, M.-P., D'Orio, V. and Ghuysen, A. Clinical decision support tool for diagnosis of COVID-19 in hospitals. PloS one, 2021. 16(3): e0247773.

[4] Organization, W. H. COVID-19: Occupational health and safety for health workers. Technical report. World Health Organization. 2021. URL http://www.jstor.org/stable/resrep33225. [5] Bartik, A. W., Bertrand, M., Cullen, Z., Glaeser, E. L., Luca, M. and Stanton, C. The impact of COVID-19 on small business outcomes and expectations. Proceedings of the national academy of sciences, 2020. 117(30): 17656–17666.

[6] Peng, C.-Y. J., Lee, K. L. and Ingersoll, G. M. An introduction to logistic regression analysis and reporting. The journal of educational research, 2002.96(1): 3–14.

[7] Organization, W. H. Clinical management of COVID-19: Interim guidance. Technical report. World Health Organization. 2020. URL http://www.jstor.org/stable/resrep27972.9.

[8] Organization, W. H. Gender and COVID-19: Advocacy brief. Technical report. World Health Organization. 2020. URL http://www.jstor.org/stable/resrep28109.

[9] Luz Yupari, I., Bardales Aguirre, L., Rodriguez Azabache, J., Barros Sevillano, J. and Rodríguez Díaz, A. Risk Factors for Mortality from COVID-19 in Hospitalized Patients: A Logistic Regression Model. Revista de la Facultad de Medicina Humana, 2021. 21(1): 4.

[10] Xiong, D., Zhang, L., Watson, G. L., Sundin, P., Bufford, T., Zoller, J. A., Shamshoian, J., Suchard, M. A. and Ramirez, C. M. Pseudo-likelihood based logistic regression for estimating COVID-19 infection and case fatality rates by gender, race, and age in California. Epidemics, 2020. 33: 100418.

[11] Bhaskaran, K., Bacon, S., Evans, S. J., Bates, C. J., Rentsch, C. T., MacKenna,B., Tomlinson, L., Walker, A. J., Schultze, A., Morton, C. E. et al. Factorr associated with deaths due to COVID-19 versus other causes: population-based cohort analysis of UK primary care data and linked national death registrations within the OpenSAFELY platform. The Lancet Regional Health-Europe, 2021.6: 100109.

[12] Mohammadi, F., Pourzamani, H., Karimi, H., Mohammadi, M., Mohammadi, M., Ardalan, N., Khoshravesh, R., Pooresmaeil, H., Shahabi, S., Sabahi, M. et al. Artificial neural network and logistic regression modelling to characterize COVID-19 infected patients in local areas of Iran. Biomedical journal, 2021.44(3): 304–316.

[13] Dong, C., Qiao, Y., Shang, C., Liao, X., Yuan, X., Cheng, Q., Li, Y., Zhang, J., Wang, Y., Chen, Y. et

al. Non-contact screening system based for COVID-19 on XGBoost and logistic regression. Computers in biology and medicine, 2022. 141: 105003. [14] Josephus, B. O., Nawir, A. H., Wijaya, E., Moniaga, J. V. and Ohyver, M. Predict mortality in patients infected with COVID-19 virus based on observed characteristics of the patient using logistic regression. Procedia computer science, 2021. 179: 871–877.

[15] Rajendra, P. and Latifi, S. Prediction of diabetes using logistic regression and ensemble techniques. Computer Methods and Programs in Biomedicine Update, 2021. 1: 100032.

[16] Nusinovici, S., Tham, Y. C., Yan, M. Y. C., Ting, D. S. W., Li, J., Sabanayagam, C., Wong, T. Y. and Cheng, C.-Y. Logistic regression was as good as machine learning for predicting major chronic diseases. Journal of clinical epidemiology, 2020. 122: 56–69.