KEMAS 18 (3) (2023) 334-340

Jurnal Kesehatan Masyarakat

http://journal.unnes.ac.id/nju/index.php/kemas

# Decision Tree Prediction Model in Patient Mortality Rate based on Risk Factors

Sri Handayani[1]✉, Ririn Nurmandhani[1], Edi Jaya Kusuma[2], Sadono Wiwoho[3]
[1]Public Health Study Program, Faculty of Health, Universitas Dian Nuswantoro
[2]Medical Record and Health Information Study Program, Faculty of Health, Universitas Dian Nuswantoro
[3]Province Health Office, Central Java

**Article Info**

*Article History:*
Submitted May 2022
Accepted September 2022
Published January 2023

*Keywords:*
Covid-19; comorbid; mortality; decision trees; data mining

**DOI**
https://doi.org/10.15294/kemas.v18i3.36710

**Abstract**

The Coronavirus disease (Covid-19) has become a global problem since WHO declared a pandemic in 2020. The number of deaths due to Covid-19 has increased significantly in many countries. This study aimed to implement decision tree modeling to represent the relationship between risk factors and the mortality rate of Covid-19 patients. This study analyzed secondary data of 83,024 Covid patients from January 2020 to June 2021. Data processing used data mining with the decision tree classification method. The results showed that comorbidity is the leading risk factor for death which is then influenced by age. The higher the age group with comorbidities, the higher the risk of death. Suggested that health services can utilize the results of this study to prevent the severity of Covid-19 infection. Such as the development of comorbid awareness programs and community-based education on managing patients with comorbidities.

## Introduction

Coronavirus diseases (Covid-19) is an infectious disease caused by the SARS-CoV-2 virus (World Health Origanisation, 2022). Until February 25, 2022, the total number of cases globally reached 428,511,601 people infected with the death rate due to Covid-19 reaching 5,911,081 (World Health Origanisation, 2022). In Indonesia alone, deaths due to Covid-19 reached 147,342 people (World Health Origanisation, 2022). The high mortality rate in Indonesia is caused by the delay in responding to the health crisis (Olivia et al., 2020)

The number of Covid-19 in Indonesia has significantly increased; cases initially only concentrated on the island of Java quickly spread to other large islands with a high number of migrant populations, such as East Kalimantan, Papua, and South Sumatra (Olivia et al., 2020). Java island became Indonesia's highest area of Covid-19 distribution (Ilmi & Praptana, 2022). The quick increase in the cases significantly impacts hospital management, such as the lack of protective equipment and health support tools (Merlin & Vanchapo, 2021).

Various studies on risk factors have been carried out. People with hypertension and diabetes mellitus, male gender, and active smokers were more likely to be infected with Covid-19 (Gallo Marin et al., 2021; Susilo et al., 2020). In addition, it is also a risk factor for death due to Covid-19 infection (Nanda Nur Illah, 2021; Satria et al., 2020). A study found that 22.8% of hospitalized Covid-19 patients died. Most of the comorbid were found (Maryati et al., 2022). The most common comorbid in Covid-19 deaths is Diabetes Mellitus. Diabetes Mellitus increased by 2.78 times higher in Covid-19 patients, resulting in death (Corona et al., 2021; Govender et al., 2021; Nandy et al., 2020; Varikasuvu et al., 2021). It is caused by several factors that increase the risk of Covid-19 infection among diabetic patients (Ejaz et al., 2020).

In addition to comorbidities, age and gender had a significant relationship with

✉ Correspondence Address:
Public Health Study Program, Faculty of Health, Universitas Dian Nuswantoro, Indonesia
Email : sri.handayani@dsn.dinus.ac.id

pISSN 1858-1196
eISSN 2355-3596

death from Covid-19. Where men aged 45-60 years are susceptible to Covid-19 and have the potential to increase mortality (Nanda Nur Illah, 2021). A model showed that men are more likely to be infected with Covid-19 with severe symptoms and require hospital treatment than women (Bienvenu et al., 2020; Clark et al., 2020; Mohamed et al., 2021).

Various studies have found an influence between comorbidities, gender, and age on the incidence of death in Covid-19 patients (Bertsimas et al., 2020; Dessie & Zewotir, 2021; Pung et al., 2021). However, no research has found a relationship between these factors and the mortality rate of Covid-19 patients. Therefore, this study aimed to implement decision tree modeling to represent the relationship between these factors and the mortality rate of Covid-19 patients.

## Method

This study used secondary data on Covid-19 patients obtained from the Semarang City Health Service (DKK) from January 2020 to June 2021. The total data obtained were 83,437 and the cleaned data processed was 83,024. The data consisted of gender, age, comorbidities, symptoms, date of being tested positive for Covid-19, date of being declared cured/died, isolation status (self-isolation or hospital treatment), and final status (cured/death). Data processing uses data mining with the decision tree classification method. Decision tree classification methods are commonly used in data mining to form a classification system that uses multiple variables or develops predictive algorithms for the dependent variable (Song & Lu, 2015). The data processing process in this study shows in Fig.1

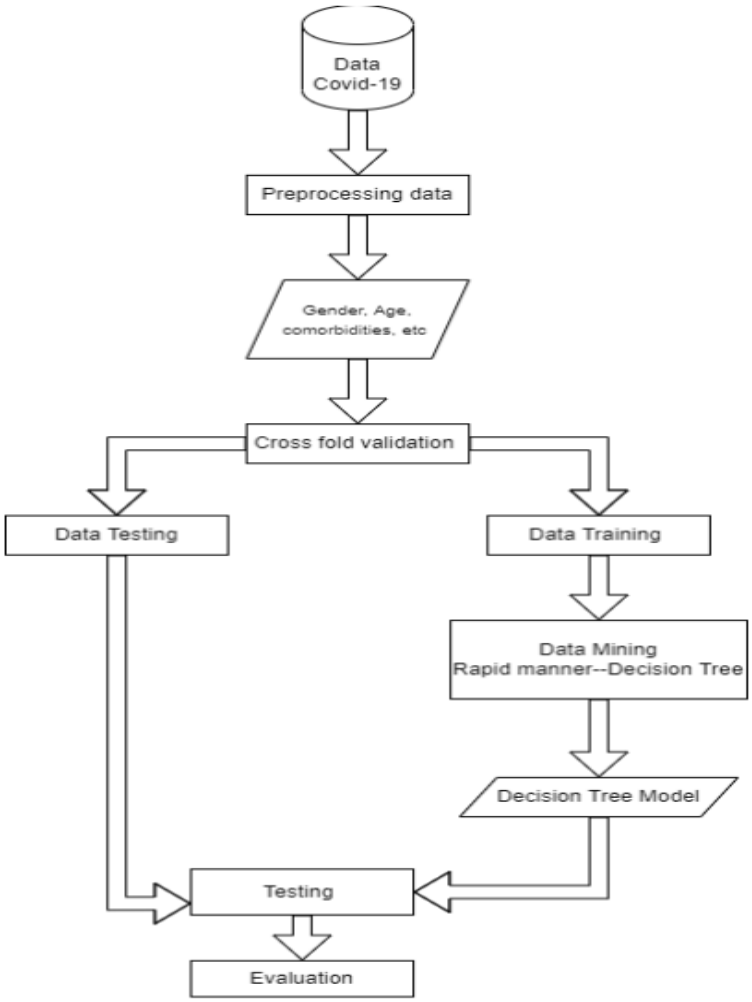


Figure 1. Data Processing

The secondary data for Covid-19 collected is carried out in the pre-processing process. In this stage, complete and incomplete data is sorted, checking data consistency. Meanwhile, in the pre-processing process, the data discretization process is carried out on the age variable with age grouping based on the risk of Covid-19 infection (World Health Organisation, 2022). The data results ready to be processed are then divided into training and test data using the cross-validation technique. This technique helps form a model that is resistant to unknown data. In this stage, the output obtained is a decision tree model trained using training data. Furthermore, the model obtained with the training data is compared with the test data, which produces test parameters in the form of the level of accuracy, precision, and recall of the model that has been generated.

The implementation process utilizes the Rapid Miner application, with the main class target being the patient status variable. The performance evaluation process is carried out using test parameters in the form of accuracy, precision, and recall. Each evaluation parameter shows how the decision tree model identifies patient status by looking at the variables of gender, age range, comorbidities, symptoms, and treatment status.

## Results and Discussion

The results of the data mining experiment process obtained a decision tree model. From the tree rules obtained, the pattern of the interrelated variables in determining patient status can be seen. Overall, the decision tree model will evaluate the comorbid variables as the first node, wherein that variable has two branches, namely "yes" and "no." In the next stage, the variable of care status is observed, wherein this variable has two branches, namely "treated in health facilities" and "self-isolation." The following observed variable is the age range which has eight branch distributions. The age range varies with the criteria "85+" and has additional branches in the form of symptom variables which are distributed into "yes" and "no." Besides that, the symptom variable with a value of "yes" is derived again through the gender variable branch (See. Fig.2).

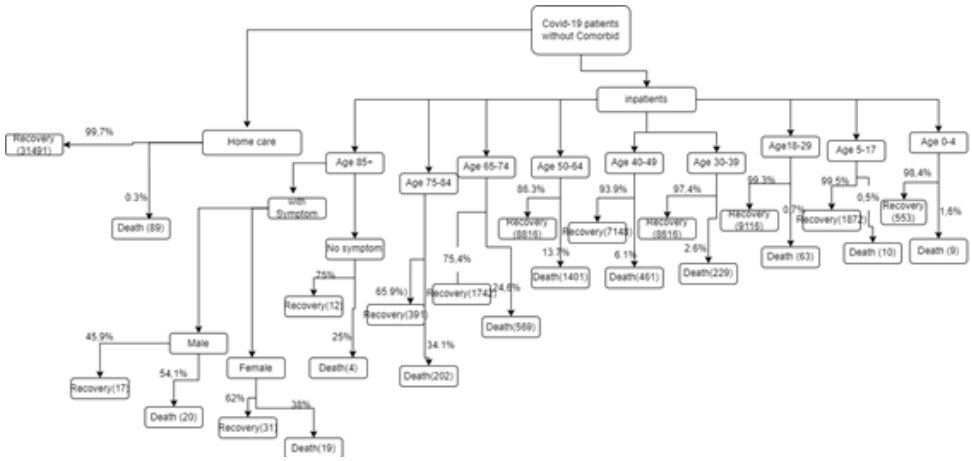


Figure 2. Patient Decision Tree without Comorbid

The decision tree model in patients without comorbidities who received treatment at the health facilities showed an increase in deaths due to Covid-19 in the elderly group. In the age group of 85 and over, the decision tree model showed the role of symptoms and gender that affect the output of patient care. Patients aged 85 and over who are symptomatic and male have a higher mortality percentage (54.1%) than females (38%) in the same age group.

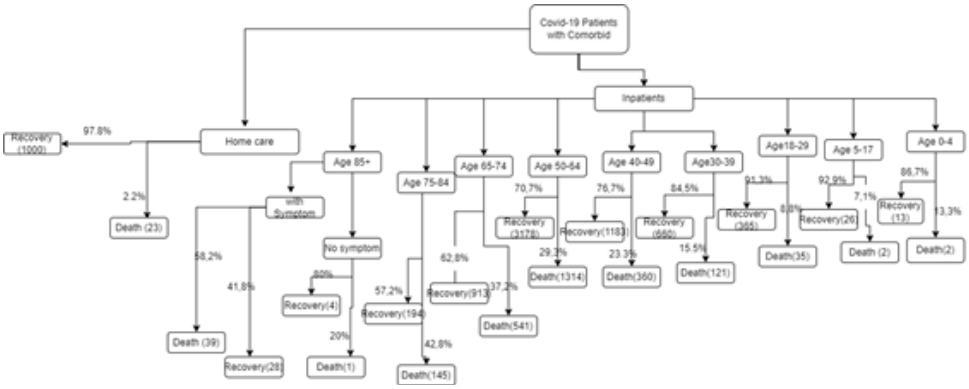


Figure 3. Decision Tree of Patients with Comorbid

Figure 3 shows that most patients with comorbidities require treatment at health facilities. The percentage of deaths increases with the increasing age group. In patients aged 85 and over, symptom factors affect treatment output. Most (58.2%) died in patients who showed early symptoms in this age group.

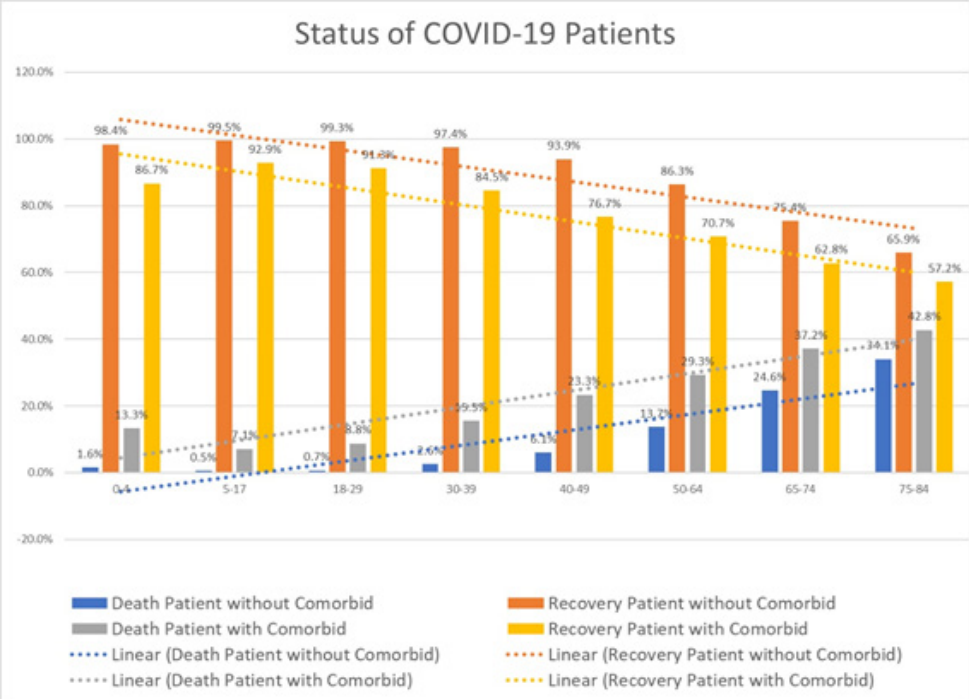


Figure 4. Status of COVID-19 Patients With and Without Comorbid

Based on comorbid variables and the distribution of age range variables, it can be seen that the older, the greater the death rate due to Covid-19. Figure 4 shows that the recovery status of patients without comorbidities has decreased with each increase in the age range, and the death rate has increased where the most significant spike is between 50 to 84 years. Meanwhile, compared with the status of patients with comorbidities shown in Figure 5, there was a significant decrease in the number of patients recovering. The opposite occurred in the number of deaths of patients with comorbidities, where mortality increased significantly, especially starting in the age range of 30 to 84.

From the evaluation results obtained, the value of each parameter (see Table 1). The accuracy value indicates the model's ability to predict the patient's status, either "dead" or "recovered." The decision tree model can produce an accuracy value of 93.18%,

which means that 93.18% of the test data generated from the cross-validation process can be predicted correctly. Then the precision parameter in the decision tree model that is formed can produce a value of 93.22%, which means the model can accurately predict 93.22% of data that has a "cured" class from all test data. Then the recall can reach 99.95%, which means the system can separate the "healed" class data with nearl

**Table 1** Model Evaluation Results

| Evaluation | Score (%) |
|---|---|
| Accuracy | 93.18 |
| Precision | 93.22 |
| Recall | 99.95 |

y 100% accuracy.

The decision tree model showed that the risk grouping of patients in the first stage is the presence or absence of comorbidities. This case described that comorbidity gives different results for Covid-19 patients. A review study showed that chronic non-communicable diseases significantly affect the prognosis of Covid-19 (Anjorin et al., 2021). Patients with comorbidities are more susceptible to infection and have more severe symptoms (Huang et al., 2020; Madjid et al., 2020). The results also found that most patients with comorbidities require treatment in health services.

In this case, comorbidities are vital in preventing deaths from Covid-19. In Indonesia, comorbidities are one of the essential focuses in controlling Covid-19 as stated in the Decree of the Minister of Health of the Republic of Indonesia Number HK.01.07/MENKES/413/2020. In implementing education and community empowerment, it is essential to control comorbidities. In addition, health services are required to report the comorbid status of Covid-19 patients. (Keputusan Menteri Kesehatan Republik Indonesia Nomor Hk.01.07/Menkes/413/2020 Tentang Pedoman Pencegahan Dan Pengendalian, 2020). A study showed that Covid-19 patients with comorbid heart and kidney failure have the potential to have severe symptoms. Meanwhile, patients with comorbid Diabetes Mellitus have a 2.78 times risk of dying from Covid-19 (Nandy et al., 2020).

The second branch of the decision tree model explained that both comorbid and non-comorbid patients have different possibilities for recovery based on age groups. The older the age group, the higher the percentage of patients who died, both comorbid and non-comorbid. A study found that being over 60 years old and having comorbidities was a risk factor for death in Covid-19 patients. Although younger patients with hypertension and diabetes had a risk of having a worse prognosis than the older age group (Bae et al., 2020).

Another study found that those less than 50 years of age had low awareness of hypertension and tended to ignore it compared to older people (Sehestedt et al., 2007). It supports the decision tree model in the comorbid patient group, where mortality increases at 30 years. An interesting thing was found in the group after 85+ without comorbidities, where the decision tree model showed a worse prognosis in the symptomatic group and male gender. A study found that biological and sociocultural factors caused vulnerability in men; many cases found men showed more severe symptoms of Covid-19 infection. (Jin et al., 2020; Papadopoulos et al., 2021; White, 2020)

## Conclusions

The decision tree model describes patients with Covid-19 infection up to 2021 before high vaccine coverage and Omicron-type. The resulting death is likely to be in patients who have not received the vaccine. However, with a high level of accuracy, the resulting decision tree model can be used as a risk assessment step in Covid-19 patients. Especially in the old age group and have comorbidities. The results of this study are evidence that in handling Covid-19, there is a need for a comorbid risk assessment. It is hoped that health services can utilize the results of this study to prevent the severity of Covid-19 infection—the development of comorbid awareness programs and community-based education on the management of patients with comorbidities.

## Acknowledgment

We sincerely thank the Center of

Excellence (CoE) Udinus and Universitas Dian Nuswantoro Research and Community Services Institute (LPPM), which has funded the research through the 2021 PDPT scheme.

## References

Anjorin, A.A., Abioye, A.I., Asowata, O.E., Soipe, A., Kazeem, M.I., Adesanya, I.O., & Raji, M.A., 2021. Comorbidities and the COVID-19 Pandemic Dynamics in Africa. *Tropical Medicine & International Health: TM & IH*, 26(1), pp.2–12.

Bae, S., Kim, S.R., Kim, N., Shim, W.J., & Park, M., 2020. Impact of Cardiovascular Disease and Risk Factors on Fatal Outcomes in Patients with COVID-19 According to Age: a Systematic Review and Meta-¬ analysis. *Heart (British Cardiac Society)*, 107(5), pp.373–380.

Bertsimas, D., Lukin, G., Mingardi, L., Nohadani, O., Orfanoudaki, A., Stellato, B., Wiberg, H., Gonzalez-Garcia, S., Parra-Calderón, C.L., Robinson, K., Schneider, M., Stein, B., Estirado, A., Beccara, L.A., Canino, R., Bello, M.D., Pezzetti, F., & Pan, A., 2020. COVID-19 Mortality Risk Assessment: An International Multi-center Study. *PloS One*, 15(12).

Bienvenu, L.A., Noonan, J., Wang, X., & Peter, K., 2020. Higher Mortality of COVID-19 in Males: Sex Differences in Immune Response and Cardiovascular Comorbidities. *Cardiovascular Research*, 116(14), pp.2197–2206.

Clark, A., Jit, M., Warren-Gash, C., Guthrie, B., Wang, H. H. X., Mercer, S. W., Sanderson, C., McKee, M., Troeger, C., Ong, K. L., Checchi, F., Perel, P., Joseph, S., Gibbs, H. P., Banerjee, A., & Eggo, R. M., 2020. Global, Regional, and National Estimates of the Population at Increased Risk of Severe COVID-19 due to Underlying Health Conditions in 2020: a Modelling Study. *The Lancet Global Health,* 8(8), pp.e1003–e1017.

Corona, G., Pizzocaro, A., Vena, W., Rastrelli, G., Semeraro, F., Isidori, A.M., Pivonello, R., Salonia, A., Sforza, A., & Maggi, M., 2021. Diabetes is Most Important Cause for Mortality in COVID-19 Hospitalized Patients: Systematic Review and Meta-Analysis. *Reviews in Endocrine & Metabolic Disorders,* 22(2), pp.275–296.

Dessie, Z.G., & Zewotir, T., 2021. Mortality-related Risk Factors of COVID-19: a Systematic Review and Meta-analysis of 42 Studies and 423,117 Patients. *BMC Infectious Diseases*, 21(1).

Ejaz, H., Alsrhani, A., Zafar, A., Javed, H., Junaid, K., Abdalla, A.E., Abosalif, K.O. \A., Ahmed, Z., & Younas, S., 2020. COVID-19 and Comorbidities: Deleterious Impact on Infected Patients. *Journal of Infection and Public Health*, 13(12), pp.1833–1839.

Gallo-Marin, B., Aghagoli, G., Lavine, K., Yang, L., Siff, E.J., Chiang, S.S., Salazar-Mather, T.P., Dumenco, L., Savaria, M.C., Aung, S.N., Flanigan, T., & Michelow, I.C., 2021. Predictors of COVID-19 Severity: A Literature Review. *Reviews in Medical Virology,* 31(1), pp.1–10.

Govender, N., Khaliq, O.P., Moodley, J., & Naicker, T., 2021. Insulin Resistance in COVID-19 and Diabetes. *Primary Care Diabetes*, 15(4), pp.629–634.

Huang, I., Lim, M.A., & Pranata, R., 2020. Diabetes Mellitus is Associated with Increased Mortality and Severity of Disease in COVID-19 Pneumonia - A Systematic Review, Meta-analysis, and Meta-regression. *Diabetes & Metabolic Syndrome,* 14(4), pp.395–403.

Ilmi, L.R., & Praptana, P., 2022. Trend for Risk Covid-19: A Case Study in Indonesia. *Jurnal Kesehatan Masyarakat,* 17(3), pp.398–404.

Jin, J.M., Bai, P., He, W., Wu, F., Liu, X.F., Han, D.M., Liu, S., & Yang, J.K., 2020. Gender Differences in Patients With COVID-19: Focus on Severity and Mortality. *Frontiers in Public Health*, 8.

Kementrian Kesehatan Republik Indonesia., 2020. *Keputusan Menteri Kesehatan Republik Indonesia Nomor HK.01.07/menkes/413/2020 tentang Pedoman Pencegahan dan Pengendalian*.

Madjid, M., Safavi-Naeini, P., Solomon, S.D., & Vardeny, O., 2020. Potential Effects of Coronaviruses on the Cardiovascular System: A Review. *JAMA Cardiology,* 5(7), pp.831–840.

Maryati, W., Widyastuti, A., Rizky, A.P.K., Listyorini, P.I., & Aryanti, F.D., 2022. Analisis Karakteristik Pada Pasien Rawat Inap Kasus Covid-19. *Infokes: Jurnal Ilmiah Rekam Medis Dan Informatika Kesehatan*, 12(1), 20–25.

Merlin, N.M., & Vanchapo, A.R., 2021. Readiness Management in Handling COVID-19 Pandemic and Early Detection in The Referral Hospital in East Nusa Tenggara Province. *Jurnal Kesehatan Masyarakat,* 17(2), pp.279–286.

Mohamed, M.S., Moulin, T.C., & Schiöth, H.B., 2021. Sex Differences in COVID-19: the

Role of Androgens in Disease Severity and Progression. *Endocrine,* 71(1), pp.3–8.

Nanda-Nur-Illah, M., 2021. Analisis Pengaruh Komorbid, Usia, dan Jenis Kelamin Terhadap Meningkatnya Angka Kematian pada Masa Pandemi Covid-19. *Jurnal Sosial Sains,* 1(10), pp.1228–1233.

Nandy, K., Salunke, A., Kumar, S., & Pandey, A., 2020. Coronavirus Disease (COVID-19): A Systematic Review and Meta-analysis to Evaluate the Impact of Various Comorbidities on Serious Events. *Diabetes & Metabolic Syndrome,* 14(5), pp.1017–1025.

Olivia, S., Gibson, J., & Nasrudin, R., 2020. Indonesia in the Time of Covid-19. *Bulletin of Indonesian Economic Studies*, 56(2), pp.143–174.

Papadopoulos, V., Li, L., & Samplaski, M., 2021. Why does COVID-19 Kill More Elderly Men than Women? Is There a Role for Testosterone?. *Andrology,* 9(1), pp.65–72.

Pung, R., Park, M., Cook, A.R., & Lee, V.J., 2021. Age-related Risk of Household Transmission of COVID-19 in Singapore. *Influenza and Other Respiratory Viruses*, 15(2), pp.206–208.

Satria, R.M.A., Tutupoho, R.V., & Chalidyanto, D., 2020. Analisis Faktor Risiko Kematian dengan Penyakit Komorbid Covid-19. *Jurnal Keperawatan Silampari,* 4(1), pp.48–55.

Sehestedt, T., Ibsen, H., & Jørgensen, T., 2007. Awareness, Treatment and Control of Hypertension in Denmark. The Inter99 study. *Blood Pressure*, 16(5), pp.312–319.

Song, Y.Y., & Lu, Y., 2015. Decision Tree Methods: Applications for Classification and Prediction. *Shanghai Archives of Psychiatry*, 27(2), pp.130–135.

Susilo, A., Rumende, C.M., Pitoyo, C.W., Santoso, W.D., Yulianti, M., Herikurniawan, H., Sinto, R., Singh, G., Nainggolan, L., Nelwan, E.J., Chen, L.K., Widhani, A., Wijaya, E., Wicaksana, B., Maksum, M., Annisa, F., Jasirwan, C.O.M., & Yunihastuti, E., 2020. Coronavirus Disease 2019: Tinjauan Literatur Terkini. *Jurnal Penyakit Dalam Indonesia,* 7(1), pp.45.

Varikasuvu, S.R., Dutt, N., Thangappazham, B., & Varshney, S., 2021. Diabetes and COVID-19: A Pooled Analysis Related to Disease Severity and Mortality. *Primary Care Diabetes*, 15(1), pp.24–27.

White, A., 2020. Men and COVID-19: The Aftermath Men and COVID-19: The Aftermath. *Postgraduate Medicine*, 132(4), pp.18–27.

World Health Organisation., 2022. WHO Coronavirus Disease (COVID-19) Dashboard With Vaccination Data | WHO Coronavirus (COVID-19) Dashboard With Vaccination Data. *World Health Organization.*