Journal of Advances in Medical and Biomedical Research | ISSN:2676-6264

Proposed Method for Predicting COVID-19 Severity in Chronic Kidney Disease Patients Based on Ant Colony Algorithm and CHAID

Firouzeh Moeinzadeh[1], Mohammad Sattari[2*]

1. *Isfahan Kidney Diseases Research Center, Isfahan University of Medical Sciences, Isfahan, Iran*
2. *Health Information Technology Research Center, Isfahan University of Medical Sciences, Isfahan, Iran*

Article Info

10.30699/jambs.30.143.507

Received: 2021/06/06;
Accepted: 2022/07/06;
Published Online: 10 Oct 2022;

Use your device to scan and read the article online

Corresponding Information:
Mohammad Sattari,
Health Information Technology Research Center, Isfahan University of Medical Sciences, Isfahan, Iran
E-Mail: msattarimng.mui@gmail.com

ABSTRACT

Background & Objective: The COVID-19 pandemic is a phenomenon that has infected and killed many people worldwide. Underlying diseases such as diabetes mellitus, heart failure, and chronic kidney disease (CKD) can affect the severity of COVID-19 and aggravate patients' condition. This study aimed to predict the severity of the COVID-19 disease in CKD patients by combining feature selection and classification methods.

Materials & Methods: This study was conducted between March 2021 and September 2021 in Isfahan University of Medical Sciences. The data set includes 83 traits of 72 kidney transplant patients, 231 kidney failure patients, and 105 dialysis patients. The data set has 77 input attributes, including age, sex, diabetes mellitus, hypertension, ischemic heart disease, chronic lung disease, and kidney transplant

In the proposed method, the combination of ant colony algorithm and the CHAID method has been used.

Results: The combination of the ant colony algorithm and CHAID method leads to better performance than CHAID alone. A total of 22 rules were extracted, of which 6 rules with a confidence of more than 60% were introduced as selected rules. The most reliable rule states that if a person has CKD stage 5, is not undergoing dialysis (5ND), and is short of breath, in 81% of cases the type of COVID-19 disease will be severe.

Conclusion: In this study the severity of COVID-19 disease in kidney patients was measured using variables including age, diabetes mellitus, blood pressure, CKD stage, etc. The results showed that high levels of kidney disease can lead to severe COVID-19.

Keywords: Data mining, Prediction, Kidney disease, COVID-19

Introduction

The COVID-19 pandemic is a phenomenon that has plagued and killed many people worldwide (1). The transmission rate of the disease is so high that it has disrupted the normal lives of people. The disease occurs with various intensities in different people. Several studies have shown that underlying diseases such as diabetes, heart failure, ischemic stroke, and kidney failure can adversely affect COVID-19 severity (2-3). Studies have demonstrated that CKD is associated with severe COVID-19 (4). Various studies have examined the relationship between these two diseases.

Cheng examined the relationship between kidney disease and hospital deaths in COVID-19 patients. In a cohort study, he evaluated 701 kidney patients with COVID-19. The results showed that 16% of deaths occurred in the hospital (5). In a case report, Fu reported a patient with end-stage kidney disease (ESRD) who had been admitted to a hospital with COVID-19 infection. The study case was a 75-year-old man with ESRD who had contracted COVID-19. The results showed that patients with chronic kidney disease (CKD) were more likely to get more severe COVID-19 infections due to their weakened immune systems (6). Flythe examined the hospital mortality of dialysis patients with pneumonia due to COVID-19. The results showed that there is a shorter time between the onset of symptoms and ICU admission in dialysis patients. The mortality rate is very high in these patients, and they should take more care to avoid COVID-19 infection (7). Marco demonstrated that simultaneous underlying kidney disease and diabetes mellitus in patients infected with COVID-19 weaken the immune system and may pose higher risk of death (8).

Due to the widespread prevalence of COVID-19, assessing its severity in CKD patients is very important, and it may be useful in decreasing their mortality rate. This study aimed to predict COVID-19 infection severity in CKD patients by combining feature selection and classification methods.

Materials and Methods

This study was conducted between March 2021 and September 2021 in Isfahan University of medical sciences.

Part I - Data Set

This article is part of a research project supported by the Isfahan University of Medical Sciences (ethics code IR.MUI.NUREMA.REC.1400.112).

The data set includes information about patients hospitalized in Isfahan who are at risk for COVID-19 infection. This set includes 72 kidney transplant patients, 231 CKD patients, and 105 dialysis patients. Out of the 72 kidney transplant patients, 16 have severe, 45 have moderate, and 10 have mild COVID-19 disease. Out of the 231 patients with CKD, 105 have severe, 104 have moderate, and 15 have mild COVID-19 disease. Out of the 105 dialysis patients, 32 have severe, 55 have moderate, and 15 have mild COVID-19 disease. The data set used in this study has 77 input attributes. These attributes include age, sex, diabetes mellitus, hypertension, ischemic heart disease, chronic lung disease, kidney transplant, insulin, hydroxychloroquine, angiotensin converting enzyme inhibitor / angiotensin receptor blocker (ACEI/ARB) use, antihyperlipidemic agent (statins) use, smoking, symptoms including cough, shortness of breath, productive cough, myalgia, headache, nausea, diarrhea, abdominal pain, loss of appetite, gastrointestinal bleeding, sore throat, insomnia, fever, multiple organ failure, dialysis during hospitalization, super-infection, hypotension, intubation, interval from onset of symptoms to recovery or death, hospital stay duration, duration of stay in ICU, and laboratory data on admission and discharge or death including hemoglobin, white blood cells (WBC), platelet, sodium, potassium, calcium, lactate hydrogenase (LDH), C-reactive protein (CRP), serum creatinine and maximum creatinine, estimated glomerular filtration rate (eGFR), and creatinine phosphokinase (CPK). The treatment regimens include antivirals (ritonavir, favapirovir, oseltamivir, ribavirin, remdesivir), antibiotics (ceftriaxone, azithromycin, meropnem, linezolide), corticosteroids (dexamethasone, hydrocortisone, methylprednisolone, prednisolone), interferon, hemoperfusion, CKD stages, COVID-19 severity, lung involvement in HRCT, disease outcome, and real time polymerase chain reaction (RT-PCR). This number of attributes is very high for data mining, so the number of attributes had to be reduced.

Part II – Preprocessing

A. Attribute reduction

After expert assessment, the 77 input traits were reduced to 35 traits: age, sex, diabetes mellitus, hypertension, ischemic heart disease, chronic lung disease, kidney transplant, insulin, hydroxychloroquine or ACEI/ARB use, antihyperlipidemic agents (statins), smoking, symptoms including cough, shortness of breath, myalgia, headache, nausea, diarrhea, abdominal pain, loss of appetite, fever, intubation, interval from onset of symptoms to recovery or death, hospital stay duration, duration of stay in the ICU, laboratory data on admission and at discharge or death including, WBC, lymphocyte count, eGFR, antivirals, antibiotics, corticosteroids, CKD stages, COVID-19 severity, lung involvement in HRCT, and outcome were selected, and 42 traits were omitted.

B. Handling Missing Values

Fortunately, the number of missing values in the selected attributes was low in the dataset. Lung involvement in HRCT is the only attribute with missing values. Average values were used to replace the unknown values.

Part III - Modeling

We used RapidMiner software Version 9.1 for statistical analysis.

A. Feature selection method

Due to the multiplicity of attributes, it was necessary to use attribute selection methods to reduce the number of attributes. The proposed method was the ant colony method (9). In this method, different combinations of traits are considered. There is a total of 2 to the power of 35 states. Each of the compounds is considered a node in the graph. Each node represents an ant with a pheromone content. The ants then begin to move and choose paths that are more efficient and less distant. Then the traits that receive less pheromone will be eliminated from the set of traits. By applying the ant colony algorithm, the number of selected traits was reduced from 35 to 18 traits, including age, diabetes mellitus, hypertension, insulin, ACEI/ ARB use, statin use, cough, shortness of breath, myalgia, nausea, fever, intubation, interval from onset of symptoms to recovery or death, hospital stay duration, duration of stay in ICU, antiviral, antibiotic, and corticosteroid use, CKD stage, and severity of COVID-19.

B - CHAID method

CHAID is a tree-based decision-making approach. In this method, data space is divided into different subcategories in different performances. This method starts from the root node and creates two or more nodes in each division. This algorithm uses chi-square statistical test and, based on this test, decides which subset should be used to identify child nodes (10). The algorithm values the values of Statistics are homogeneous or similar depending on the target variable and retain all heterogeneous values. Then the best attribute is selected to form the first branch in the decision tree in a way that each node is made up of a group of homogeneous values of the selected target variable. This reversal process continues until the tree is fully formed. The flowchart of the proposed method is presented in Figure 1.

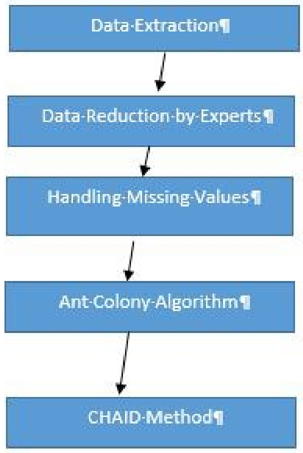

Figure 1. **Flowchart of the proposed method**

Results

Evaluation Measures

This study used test data sets to evaluate the methods. First, the complexity matrix, which includes various criteria, such as true positive (TP), true negative (TN), false positive, and false negative (FN), is calculated. The target is a three-objective class of mild, moderate, and severe. Thus, a 3×3 complexity matrix will be created. The rows correspond to each of the predicted values and the columns correspond to the actual values.

One of these criteria is accuracy (11); the closer the accuracy is to 1, the better the result. This criterion is calculated using the following formula:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Other criteria are sensitivity, specificity (12), and confidence (13); the closer they are to 1, the better the performance. The confidence value indicates how reliable this rule is. The higher the value, the more likely the head items will occur in a group (13). Sensitivity can determine the true positive rate, and specificity can determine the false positive rate.

Eighteen attributes were extracted. People were divided into five groups based on their age. The number of people aged between 56 and 70 was 126 and the number of those aged between 71 and 98 was 187. Two hundred and eleven people had diabetes mellitus, 95 people used insulin, and 89 people had CKD stage 4.

Table 1. **Attributes, values, and count of kidney disease dataset**

Attribute	Value (count)
Age	10–25 (8), 26–40 (23), 41–55 (58), 56–70 (126), 71–98 (187)
Myalgia	Yes (247), No (157)
Diabetes Mellitus	Yes (211), No (194)
Hypertension	Yes (216), No (187)
Insulin	Used (95), Not used (310)
ACEI/ARB	Used (125), Not used (280)
Statin	Yes (127), No (276)
Cough	Yes (197), No (205)
Shortness of breath	Yes (229), No (174)
Intubation	Used (80), Not used (324)
Nausea	Yes (76), No (328)
Fever	Yes (136), No (269)
Antiviral	Yes (46), No (359)
Antibiotics	Yes (313), No (92)
Corticosteroids	Yes (275), No (130)
Interval from onset of symptoms to recovery or death	One day to one week (80) One week two weeks (141) two weeks or more (182)
CKD stage	2 (11), 3a (43), 3b (69), 4 (89), 5 (4), 5nd (47)
Duration of stay in ICU	0 days (243), one day to one week (89), more than one week (64)

To investigate the effectiveness of the ant colony algorithm in reducing the number of traits from 35 to 18, the CHAID method was used once on the 18 traits and once on the 35 traits. Therefore, the proposed method is the combination of ant colony theory and CHAID. This combined method was compared to the CHAID method. The accuracy is shown in Table 2. The results of Table 2 show that the use of the ACO method improved the accuracy of the CHAID method by up to 17%.

Table 2. **The accuracy of CHAID and CHAID+ACO.**

CHAID	CHAID+ACO
70.37%	87.65%

The sensitivity and specificity of the methods are listed in Table 3. The CHAID+ACO method detects 86% of severe cases of COVID-19 correctly. It was able to identify 92% of the moderate cases of COVID-19.

Table 3. **The sensitivity and specificity of CHAID and CHAID+ACO.**

	CHAID		CHAID+ACO	
Class	Sensitivity	Specificity	Sensitivity	Specificity
Mild	65.00%	82.00%	80.00%	97.00%
Moderate	77.46%	70.46%	92.16%	87.76%
Severe	69.33%	78.24%	86.93%	94.00%

A total of 22 rules were extracted, of which 6 rules with a confidence higher than 60% were introduced as the selected rules, which are listed in Table 4. These rules have been approved by kidney disease experts.

Table 4. **Extracted rules about severity of COVID-19 in kidney patients.**

Rule	Condition	Result	Confidence
Rule 1	If the length of hospital stay in the ICU is more than 7 days, age is between 71 and 98, and the person has nausea	Severe COVID-19	85%
Rule 2	If CKD stage is 3b and age is between 41 and 55	Severe COVID-19	66%
Rule 3	If CKD stage is 5ND and the person has shortness of breath	Severe COVID-19	81%
Rule 4	If CKD stage is 2	Moderate COVID-19	63%
Rule 5	If CKD stage is 3b and the person has diabetes mellitus	Moderate COVID-19	66%
Rule 6	If the interval from the onset of symptoms to recovery or death is two weeks and the person has shortness of breath	Mild COVID-19	68%

Discussion

The recent transition from traditional medicine to evidence-based medicine has highlighted the need for data mining (14). One of the issues with COVID-19 is the effect of underlying diseases on its severity. Research has shown that underlying diseases such as kidney disease or diabetes mellitus can exacerbate COVID-19 or even lead to death (7). Therefore, this group of patients have a higher priority over other patients in receiving vaccines (15). In the rules extracted in this study, the effect of kidney disease level on the severity of COVID-19 is evident. If a person has 5ND kidney disease level and also has shortness of breath, in 81% of cases, COVID-19 will be severe. Also, if the CKD stage is 3b and the age is between 41 and 55 years old, the type of COVID-19 disease will be severe in 66% of cases. Ronco stated that kidney involvement in COVID-19 is high and will lead to hospital admission in 40% of cases (16). Also, if the CKD stage is 2, COVID-19 severity will be moderate in 63% of cases. In fact, the lowest level of kidney disease corresponds to moderate COVID-19 severity. This highlights the importance of monitoring CKD patients and their health during the COVID-19 pandemic. Thus, these patients may deal with more serious COVID-19 infections even if their CKD disease level is not very high. Kapanova stated that chronic kidney disease has negative impact on COVID-19 (17). Note that in another rule, even non-diabetes mellitus patients with CKD stage 3b experience moderate COVID-19 severity in 66% of cases.

According to the study findings, the use of ant colony improves the performance of the CHAID method. Therefore, it can be said that basing prediction on some traits does not improve the performance of the method and may even reduce the accuracy of the method. In fact, the ant colony theory method can prevent the destructive effects of confounding variables and improve the performance of the proposed method.

Conclusion

In this study, the COVID-19 severity in CKD patients was measured using various variables including age, diabetes mellitus, hypertension, CKD stage, etc. The results showed that ACO improved the performance of the CHAID method in predicting disease severity in CKD patients with COVID-19. Thus, this combination can be considered in future machine learning studies. Moreover, kidney disease can lead to more severe COVID-19 infections.

Measures should be taken to monitor the health of these patients. Moreover, the importance of kidney disease in prioritizing these patients in hospitals across the country should be considered.

Acknowledgments

None.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

1. Singhal T. A review of coronavirus disease-2019 (COVID-19). Indian J Pediat 2020; 87(4):281-6. [DOI:10.1007/s12098-020-03263-6] [PMID] [PMCID]
2. Merkler AE, Parikh NS, Mir S, et al. Risk of ischemic stroke in patients with coronavirus disease 2019 (COVID-19) vs patients with influenza. JAMA neurology. 2020;77(11):1366-72. [DOI:10.1001/jamaneurol.2020.2730] [PMID] [PMCID]
3. Bohn MK, Hall A, Sepiashvili L, Jung B, Steele S, Adeli K. Pathophysiology of COVID-19: Mechanisms underlying disease severity and progression. Physiology. 2020;35(5):288-301. [DOI:10.1152/physiol.00019.2020] [PMID] [PMCID]
4. Henry BM, Lippi G. Chronic kidney disease is associated with severe coronavirus disease 2019 (COVID-19) infection. International urology and nephrology. 2020;52(6):1193-4. [PMCID] [DOI:10.1007/s11255-020-02451-9] [PMID]
5. Cheng Y, Luo R, Wang K. Kidney disease is associated with in-hospital death of patients with COVID-19. Kidney Int. 2020;97(5):829-38. [DOI:10.1016/j.kint.2020.03.005] [PMID] [PMCID]
6. Fu D, Yang B, Xu J, Mao Z, Zhou C, Xue C. COVID-19 infection in a patient with end-stage kidney disease. Nephron. 2020;144(5):245-7. [DOI:10.1159/000507261] [PMID] [PMCID]
7. Flythe JE, Assimon MM, Tugman MJ, et al. Characteristics and outcomes of individuals with pre-existing kidney disease and COVID-19 admitted to intensive care units in the United States. Am J Kidney Diseases. 2021;77(2):190-203. [DOI:10.1053/j.ajkd.2020.09.003] [PMID] [PMCID]
8. D'Marco L, Puchades MJ, Romero-Parra M, Gorriz JL. Diabetic kidney disease and COVID-19: the crash of two pandemics. Front Med. 2020;7:199. [DOI:10.3389/fmed.2020.00199] [PMID] [PMCID]
9. Dorigo M, Birattari M, Stutzle T. Ant colony optimization. IEEE Computational Intelligence Magazine. 2006; 1(4):28-39. [DOI:10.1109/MCI.2006.329691]
10. Hill DA, Delaney LM, Roncal S. A chi-square automatic interaction detection (CHAID) analysis of factors determining trauma outcomes. J Trauma Acute Care Surg. 1997;42(1):62-6. [DOI:10.1097/00005373-199701000-00011] [PMID]
11. Swets JA. Measuring the accuracy of diagnostic systems. Science. 1988;240(4857):1285-93. [DOI:10.1126/science.3287615] [PMID]
12. Parikh R, Mathai A, Parikh S, Sekhar GC, Thomas R. Understanding and using sensitivity, specificity and predictive values. Indian J Ophthalmol. 2008;56(1):45.[PMCID][PMID] [DOI:10.4103/0301-4738.37595]
13. Astudillo C, Bardeen M, Cerpa N. data mining in electronic commerce-support vs. confidence. J Theor Appl Electron Commer Res. 2014;9(1):1-7. [DOI:10.4067/S0718-18762014000100001]

14. Moeinzadeh F, Rouhani M H, Mortazavi M, Sattari M. Prediction of chronic kidney disease in Isfahan with extracting association rules using data mining techniques. Tehran Univ Med J. 2021;79(6):459-67
15. Magicova M, Zahradka I, Fialova M, et al. Determinants of immune response to anti-SARS-CoV-2 mRNA vaccines in kidney transplant recipients: A Prospective Cohort Study. Transplantation. 2022;106(4):842-52. [PMCID] [DOI:10.1097/TP.0000000000004044] [PMID]
16. Ronco C, Reis T, Husain-Syed F. Management of acute kidney injury in patients with COVID-19. Lancet Respir Med. 2020;8(7):738-42. [DOI:10.1016/S2213-2600(20)30229-0]
17. Kapanova GK, Kaskabayeva AS, Alibekova RI, Botabayeva AS, Muzdubayev DK. Clinical case: symptomatics and efficiency of treatment of coronaviral infection on the background of chronic kidney disease. Sci & Health Care. 2021; 23(1):69-73.