

**PREDICTION COVID-19 POSSIBILITY WITH COMORBIDITY AND  
RECURSIVE FEATURE ENGINEERING**

**BY**

**MD. SHOJOL ISLAM**

**ID: 161-15-7452**

This Report Presented in Partial Fulfillment of the Requirements for the Degree of  
Bachelor of Science in Computer Science and Engineering

Supervised By

**DR. SHEAK RASHED HAIDER NOORI**

Associate Professor & Associate Head  
Department of CSE  
Daffodil International University



**DAFFODIL INTERNATIONAL UNIVERSITY**

**DHAKA, BANGLADESH**

**JANUARY 2021**

## **APPROVAL**

This Project titled "**Prediction COVID-19 Possibility with Comorbidity and Recursive Feature Engineering**", submitted by **Md. Shojol Islam** to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on **January 27, 2021**

### **BOARD OF EXAMINERS**



---

**Dr. Touhid Bhuiyan**  
**Professor and Head**  
Department of Computer Science and Engineering  
Faculty of Science & Information Technology  
Daffodil International University

**Chairman**



---

**Gazi Zahirul Islam**  
**Assistant Professor**  
Department of Computer Science and Engineering  
Faculty of Science & Information Technology  
Daffodil International University

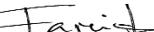
**Internal Examiner**



---

**Raja Tariqul Hasan Tusher**  
**Senior Lecturer**  
Department of Computer Science and Engineering  
Faculty of Science & Information Technology  
Daffodil International University

**Internal Examiner**



---

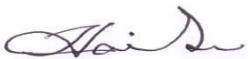
**Dr. Dewan Md. Farid**  
**Associate Professor**  
Department of Computer Science and Engineering  
United International University

**External Examiner**

## **DECLARATION**

We hereby declare that, this project has been done by us under the supervision of **Dr. Sheak Rashed Haider Noori, Associate Professor & Associate Head, Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

**Supervised by:**



---

**Dr. Sheak Rashed Haider Noori**  
Associate Professor & Associate Head  
Department of CSE  
Daffodil International University

**Submitted by:**



---

**Md. Shojol Islam**  
ID: 161-15-7452  
Department of CSE  
Daffodil International University

## **ACKNOWLEDGEMENT**

First we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year project/internship successfully.

We really grateful and wish our profound our indebtedness to **Dr. Sheak Rashed Haider Noori, Associate Professor & Associate Head**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of Machine to carry out this project. His endless patience ,scholarly guidance ,continual encouragement , constant and energetic supervision, constructive criticism , valuable advice ,reading many inferior draft and correcting them at all stage have made it possible to complete this project.

We would like to express our heartiest gratitude to Dr. Syed Akhter Hossain, Professor and Head, Department of CSE, for his kind help to finish our project and also to other faculty member and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

## ABSTRACT

People of the whole world are now just stopped because of the corona virus. People are became helpless regarding COVID-19. We forgot living our everyday life. Most people are not aware that they are suffering from comorbidity which weakens their immunity to fight corona virus. Corona virus is becoming more lethal to those people. So if we want to be safe then we must have to be cautious about comorbidity and start maintaining it. Our organs get weak because of comorbidity and our immunity system gets weaker. If corona virus attacks the people who have comorbidities then there's a possibility of organ failure which increase the risk of death .My research on "**Prediction COVID-19 Possibility with Comorbidity and Recursive Engineering**" will help to understand how we are at risk while having comorbidity and the relation between comorbidity. Also I have applied the recursive feature engineering technique to find the relation between columns of my dataset. Then we will able to find the possibility of getting infected with other than corona virus. So in this research it briefly discussed that how a patient can easily get infected with corona virus with comorbidity syndrome. Also we can analyze the chances of getting infected with one comorbidity syndrome to another syndrome.

## **TABLE OF CONTENTS**

<b>CONTENTS</b>	<b>PAGE</b>
Board of examiners	i
Declaration	ii
Acknowledgements	iii
Abstract	iv
<b>CHAPTER</b>	
<b>CHAPTER 1: INTRODUCTION</b>	<b>4-5</b>
1.1 Introduction	4
1.2 Motivation	4
1.3 Rational of the Study	4
1.4 Research Questions	5
1.5 Expected Outcome	5
<b>CHAPTER 2: BACKGROUND</b>	<b>6-8</b>
2.1 Introduction	6
2.2 Related Works	6
2.3 Research Summary	7
2.4 Scope of the Problem	7
2.5 Challenges	8

<b>CHAPTER 3: RESARCH METHODOLOGY</b>	9-13
3.1 Research Subject and Instrumentation	9
3.2 Data Collection Procedure/ Data Utilization	11
3.3 Statistical Analysis	12
3.4 Recursive Feature Engineering	12
3.5 Implementation Requirement	13
<b>CHAPTER 4: EXPERIMENTAL RESULTS AND DISCUSSION</b>	14-15
4.1 Introduction	14
4.2 Experimental Results and Analysis	14
4.3 Discussion	15
<b>CHAPTER 5: SUMMARY, CONCLUSION RECOMMENDATION AND IMPLICATION FOR FUTURE RESEARCH</b>	16-17
5.1 Summary of The Study	16
5.2 Limitations	16
5.3 Conclusion	16
5.4 Recommendations	16
5.5 Implication for Future Research	17
<b>APPENDIX</b>	18
<b>REFERENCES</b>	19
<b>PLAGIARISM REPORT</b>	20

## **List of Figures**

<b>FIGURES</b>	<b>PAGE</b>
Figure 1: Using Recursive Feature Engineering	12
Figure 2: COVID-19 Possibility with Comorbidity	12

## **List of Table**

### **Tables**

Table 1: With comorbidity Possibility	14
Table 2: Applying Recursive Feature Engineering	15

# **CHAPTER 1**

## **INTRODUCTION**

### **1.1 Introduction**

In this 21<sup>st</sup> century we are facing the massive disaster which is COVID-19. It just stops our daily life in this whole world. Corona virus is the most deadly virus among us. Because it's a new virus the scientist are researching this daily. This is kind of virus which we didn't expect or predicted. This virus is so unpredictable because the symptoms are very normal like fever, dry cough and sickness. The whole world is now just stopped because of this virus. It is inheritance of SARS virus but more dangerous. 80.4 million\* people are infected with this virus and even 1.76 million\* people are already dead. So COVID-19 virus is a great threat to human being.

### **1.2 Motivation**

The COVID-19 virus is spreading rapidly. So I have decided to research to those infected and non-infected people who have comorbidity. Comorbidity makes our immunity system weak and easily harm our body so I've decided to predict COVID-19 possibility to those people who are suffering from comorbidity. Observe shows that comorbidity directly influence the possibility of getting infected by COVID-19 [1].

### **1.3 Rational of the Study**

Computer science really means to solve our real life problems using computer technology. In this research I have used several algorithms to predict a value which can indicates the possibility of getting infected with COVID-19 or corona virus. In this whole world there are lot of people who are suffering from comorbidity. Corona infected people can easily face organ failure problem. So this study will help them to be cautious from corona virus.

## **1.4 Research Questions**

- What is the relation between comorbidity and corona virus?
- What is recursive feature engineering in this research?
- What are the benefits of those methodology?
- What type of algorithm is used to predict the possibilities?
- What type of value is in this research?
- What type of result generate in this research? Which platform is used to proceed this research?
- What are the chances that the dataset is valid?

## **1.5 Expected Outcome**

My main focus is to aware people who are suffering from heart disease, kidney disease, diabetes, blood pressure, liver disease, asthma and who are addicted to smoking or alcohol.

These kind of people have some weak immunity system which can increase the possibility of getting infected by corona virus easily and it can be lethal to them because of those weak immunity system. According to my method my calculations for each people can know if they have to be more cautious. I have got my expected outcome during this research. This research is not only focused on covid or non-covid patients it's also focused on the chances that if you are suffering from one comorbidity disorder then there might be chance that anyone can suffer various comorbidity following that syndrome.s

## **CHAPTER 2**

### **BACKGROUND**

#### **2.1 Introduction**

In addition to corona virus is spreading rapidly it seeks for no ages or limitations. No matter how young or older peoples are. As long as they are in touch with infected people they will get infected too and this might prove lethal for them.

Following this above circumstances I have decided to work on non-infected people who are at great risk that they might infected by corona virus. I have studied several algorithms where I can get a result that shows how much a people is at risk with comorbidity.

Hence, I think my work proposal will be convenient to aware those people who have comorbidity like diabetes, liver disease, kidney disease, asthma, heart disease and habit of alcohol or smoking from getting infected buy lethal corona virus.

#### **2.2 Related Works**

Basically there are lot of COVID-19 related work took place but none of them relates the comorbidity I have chosen. This is original and new concept. However I have created my own dataset which is unique for this research and it creates a new dimension to fight with novel corona virus.

The most unique part of my research is I have chosen a very common comorbidities which indirectly affect people in several ways. I have used several algorithms to get accurate value where anyone can easily analyze how much they are in risk.

## **2.3 Research Summary**

In my study I aimed to find a way to aware the people who are suffering from covid or non-covid and Comorbidity through computer technology. I have studied sufficient behind my work like how comorbidity works and how corona virus weaken our immunity system and if our organs are already been damaged then for the infection of COVID-19 we might face organ failure problem that leads to death.

I have studied several algorithms such as naïve Bayes, Linear and Logistic Regression, Decision Tree, Support Vector Machine (SVM), Random Forest basically algorithms follows the dataset and apply their method to give us a proper prediction with accuracy.

In this paper the following algorithms are applied on 2651 data of random patients who are infected and non-infected from COVID-19 and also their records of comorbidity.

In my efforts, I am going to introduce a new and innovative way to acknowledge the risk of getting infected with corona virus.

## **2.4 Scope of the Problem**

This study concentrates on gaining information about corona virus infection chances and also for recursive feature engineering we can get to know the relation between comorbidity or in simple words how a disease influenced by another disease.

The columns of infected, liver disease, kidney disease, heart disease, asthma, diabetes and smoking or alcohol habits are interrelated between each other that's what is shown in this paper with recursive feature engineering which is the outside work of getting infected or non-infected by corona virus.

In future work people can get an application of this work where they can get a result of their COVID-19 prediction. They will also know what should follow to not being a

Victim of comorbidity. Comorbidity causes increase of covid-19 cases. Basically covid-19 creates a major problem to human organs and also the people who have comorbidity they are already sufferer of weak organs so logically they are more at most danger by getting infected with corona virus and it might be possible that they might lose their lives during that covid-19 infection period.

## **2.5 Challenges**

The main challenge I have faced during this research is collecting dataset. The whole dataset was collected during corona with a large time period. I had to travel to the hospitals and also survey through online procedure. People are not aware of corona virus so much and also they don't know about that they have comorbidity or not. So collecting data and applying it to prediction algorithm was a great challenge for me. The hospitals was so classified to share their patient data with me. So I have collect data personally form every patient. The doctors assisted a lot in this research and they have verified the data also. They enlightened me with the comorbidity and its dangerous parts also they focus on how we can maintain a good condition during covid-19 and comorbidity suffering. The hospitals are more confidential about their data of any patients. Also they are afraid of the corona virus. The most challenging fact of my work is to collect data from various patient. Basically the people who have corona symptoms they don't have to check their diagnostic test. Somehow the doctors who are really helpful during these period they assist me a lot from their diagnostic report so that I can fulfil my research. These are the major challenges to proceed research among the covid and non-covid patients.

## CHAPTER 3

### RESEARCH METHODOLOGY

#### **3.1 Research Subject and instrumentation**

##### 3.1.1 Converting the dataset to categorical

The whole dataset was converted to binary format or 0 = negative and 1 = positive. The dataset was collected as a diagnostic form and then processed as a binary or categorical format.

##### 3.1.2 Split dataset into train, test

The dataset was divided as train and test. I have trained 0.6 or 60% of data.

##### 3.1.3 Algorithms

The algorithms worked differently on my dataset. I have tried to apply as many algorithms as possible to get a good accuracy from my dataset.

This is the equation of linear regression model. This equation basically reduce the distance between the data and it is continuous

$$y = m_1x_1 + m_2x_2 + \dots + b$$

This is linear regression equation. It basically reduce the error between data points.

$$y = \frac{1}{1+e^{-(m_1x_1+m_2x_2+\dots+b)}} \quad [ e \approx 2.71828 ]$$

This is the logistic regression. It is the technique used for classification and the predicted value of it is categorical.

SVM algorithm is used to margin the distance between data points. Decision tree algorithm basically splits the dataset repeatedly to predict the value.

$$P\left(\frac{A}{B}\right) = \frac{P\left(\frac{B}{A}\right) * P(A)}{P(B)}$$

This is the naïve bayes theorem. This basically runs a conditional probability to predict a result.

Random forest basically divides the dataset into random number and then build several decision trees. It basically took the majority from those decision trees and give the probable output. This algorithm proves more efficiency 77% for my dataset.

### 3.1.4 Cross Validation

After applying cross validation for the mean of 10 values –

Logistic Regression – 76%

K Neighbors – 69%

SVM – 76%

Decision Tree – 76%

Naïve Bayes – 66%

Random Forest – 77%

I got these accuracy after applying cross validation.

### 3.1.5 Stratified K Fold

After folding dataset into 10 splits I got –

**Decision Tree** – 82.63% (maximum), 76.63 (average)

**Naïve Bayes** – 75.6% (maximum), 71.9% (average)

**Random Forest** – 82.64% (maximum), 76.64 (average)

From all the algorithms above i got the most accuracy from Random Forest algorithm.

### 3.1.6 Recursive Feature Engineering

My first motive is to predict the COVID-19 possibility with comorbidity but from analyzing data with recursive feature engineering I can also predict the relation between the columns. Suppose if a person is addicted to smoking or alcohol then the probability of getting infected with heart disease, asthma, liver disease

## 3.2 Data Collection Procedure/Data Utilize

For the purpose of my research I did field work for data collection and also did a survey.

My dataset contains information about Infected, diabetes, liver disease, kidney disease, heart disease, asthma and smoking or alcohol habit these are the basic information of comorbidity. I have collected data from several hospitals inside and outside of Dhaka with the massive help of the doctors.

### 3.3 statistical analysis

The statistical view of my dataset is-

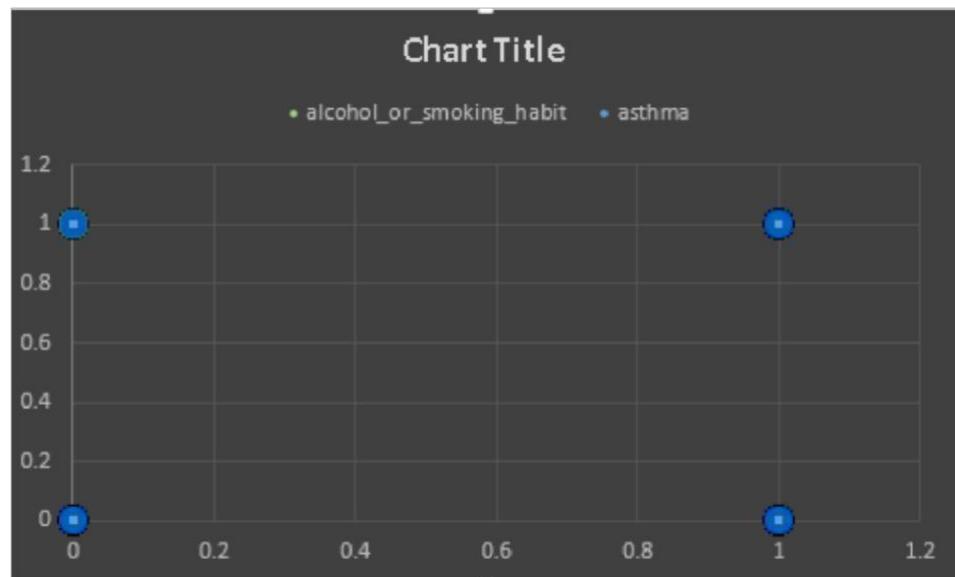


Figure 1: Recursive Feature Engineering

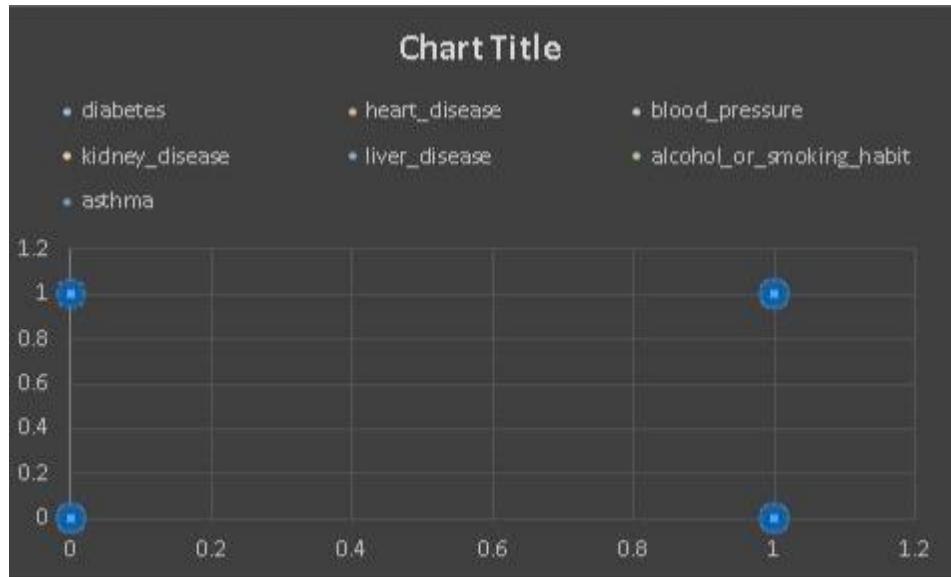


Figure 2: COVID-19 Possibility with comorbidity

My dataset is completely contains categorical data so this is the view of statistical analysis of my data. But the data shows all the information about core comorbidity symptoms. I have collected data from several cities and several peoples. So that I can found out the real possibility across the country. But many other statistical views shows that the people who have comorbidity they are more risk than the people who have good immune system.

### **3.4 Implementation Requirement**

I have implemented my research methodology using python language with the platform Jupyter notebook and anaconda. I have chosen these software's for developing my prediction model. These application are used for applying several algorithms for predictions and possibility. In future there will an application where I any people can find their infection chances. Other than these it required close observation to fulfill the process of my research. The validity of this research is authentic

# CHAPTER 4

## EXPERIMENTAL RESULTS AND DISCUSSION

### 4.1 Introduction

I have basically experimented several users to ensure my data integrity. I experimented to all those people who have comorbidity.

### 4.2 Experimental Results

COVID-19 with comorbidity using Random Forest Algorithm-

0 = No

1 = Yes

Table 1: Covid-19 Possibility experimented sample

Diabetes	Heart Disease	Blood Pressure	Kidney Disease	Liver Disease	Alcohol or smoking habit	Asthma	Possibility(after applying algorithms)
0	0	1	0	1	0	1	1
1	1	1	0	0	0	0	1
1	0	0	0	0	1	1	1
0	0	0	1	1	0	0	0
1	0	1	0	0	0	0	1
0	1	0	0	0	1	0	0
0	0	0	0	0	0	1	1
0	1	0	0	0	1	0	0
0	0	0	0	0	0	0	0

Table 2: Using recursive feature engineering

Heart Disease	Kidney Disease	Liver Disease	Asthma	Possibility(Smoking or Alcohol)
0	0	0	1	1
1	0	0	1	1
1	1	0	0	1
1	1	1	0	1

From the above scenario we can analyze that the people with asthma is at a great risk of COVID-19. The experiment also shows the possibility of getting infected by COVID-19 with comorbidity. But those who already have asthma are in great danger because corona virus brings down their oxygen level on blood. Patients with moderate to intense asthma are at a disadvantage because this virus affects their respiratory tracts, leading to increased asthmatic attacks [2].

Also from recursive featuring engineering we have got to know that it increase the possibility of getting the disorder asthma, kidney or liver disease because of smoking or alcohol habit. People with diabetes are inclined to get infections due to impaired phagocytic cell capabilities. Further, several other factors increase the risk of COVID-19 in diabetic patients. An elevated level of ACE-2 receptors found to be causally related to diabetes by Mendelian randomization analysis; this might prejudice people with diabetes to SARS-CoV-2 infection [4].

### 4.3 Discussion

At last I can recommend this procedure to find out the possibility of getting infected by corona virus with comorbidity. Form my analysis the people who have asthma are at risk dangerously. Also after recursive feature engineering we can analysis the relation between comorbidity.

# **Chapter 5**

## **Summary of the Study, Limitations, Conclusion, Implication for Future Research**

### **5.1 Summary of the Study**

The world is facing a crisis that no one ever seen. It tells us how much weaker we are at our medical sectors. We have to be more concerned about medical technology. People are not careful about their health. They even do not know that they are victims of comorbidity. So that's why I decided to research COVID-19 with comorbidity. If a disease comes to the mother earth then it doesn't vanish that easily.

So we have to raise more consciousness among all of the world. The only thing that can fight any disease is our inner immunity system. But due to comorbidity the immunity system gets hampered heavily. So we have to raise our immunity system to fight any kind of viruses which may come in future or covid-19.

### **5.2 Limitations**

My methodology has some drawbacks like –

- We need more data to predict accurate value
- For now there is no framework where people can use this methodology to predict
- We can't get the death prediction from my dataset because of hospital privacy

### **5.3 Conclusion**

I analyze this methodology to predict possibility of getting infected by COVID-19 with comorbidity. There are many research on COVID19 but mine one's data is completely different than them. I worked with the common comorbidity symptoms. I also collected data from various places. I got an accuracy of 82% which is good for the current number of dataset. For the privacy I couldn't get more than 2651 data. But my research purpose can be fulfilled by them.

## **5.4 Recommendation**

The whole research was based on medical science through computer technology. So our honorable doctors can use this methodology to predict patient easily. They can easily predict the corona virus by splitting comorbidity and non-comorbidity people. So this methodology could be useful for medical science. It is known for almost 18 years that asthmatic people are more prone to develop viral infections. If left uncontrolled, these viral infections can develop severe symptoms. People with asthma have a delayed innate antiviral immune response and impaired secretion of IFN- $\lambda$ , which makes people more susceptible to develop severe complications [3].

## **5.5 Implication for Future Research**

We all want to make our life better in an easiest way. Corona virus is not the end of problems. There might more viruses waiting for us. But we know nothing creates new. Even corona virus have trace back. We all know corona virus is a new form of SARS virus. So I want to research more about corona virus and also I want to predict the deaths cause corona virus and extend this paper then.

I want to research about any virus that can be harm us in future like corona virus and I also want to research SARS inheritance.

## **APPENDIX**

### **Appendix A : Research Reflections**

This part is about to tell the research reflections. It's a massive challenge for me to work on this medical based research. It took a lot of hard work in that research. In my educational life I have learned several thing but this one is completely different than them. I wanted to solve a real life medical problem with computer science and raise awareness among the people who is at risk by getting infected with COVID-19.

A proper planning with hard work and time distribution is key to success of our research. I go through a lot of processing during this research. I gathered a lot of new experience through my research. I met so many people and gathered many information. In future it will really help me to come up something new.

### **Appendix B : Related Issues**

I had to visit physically to collect data. It was quite tough during the whole period of corona virus. I had to confirm more and more to collect a sample data of patients i had to make sure the dataset was true otherwise we will not get the right prediction. So I had to recheck every data.

## **REFERENCES**

- [1] C. O. A. M. R. P. K. Y. P. D. Z. H. I. P. J. M. & M. A. Adekunle Sanyaolu, "Comorbidity and its Impact on Patients with COVID-19," *SN Comprehensive Clinical Medicine*, pp. 1-8, 2020
- [2] "Centers for Disease Control and Prevention.,," 22 december 2020. [Online]. Available:<https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html>. [Accessed 25 december 2020]
- [3] S. M. S. L. M. E. P. W. N. B. M. Contoli, "Asthma and COVID-19," *Journal of Infection and Public Health*, vol. 13, no. 12, 2020.
- [4] A. L. H. S. S. Rao, "Exploring diseases/traits and blood proteins causally related to expression of ACE2, the putative receptor of SARS-CoV-2: a Mendelian randomization analysis highlights tentative relevance of diabetes-related traits," *Diabetes Care*, 2020.

# PLAGIARISM REPORT

