

Epidemiological Analysis of MERS-CoV Using NN and SVM in Respect to Applicability of AI in Multiple Classes

Soyoung Hong, Sungwoo Choi, Donghyun Kim, and Taeseon Yoon

Natural Science: Hankuk Academy of Foreign Studies, Yongin, Republic of Korea

Faculty of Computer: Hankuk Academy of Foreign Studies, Yongin, Republic of Korea

ghdthdud99@gmail.com, swchoi21@gmail.com, k1978022@gmail.com, tsyoon@hafs.hs.kr

Abstract— Middle East Respiratory Syndrome is a respiratory disease spread by MERS coronavirus, which had its first occurrence in Kingdom of Saudi Arabia. In this paper, we analyzed the transmission route of MERS-CoV with a novel molecular approach. We collected the amino acid sequence of MERS-CoV from 15 different regions, which were then analyzed with Neural Network and Support Vector Machine. With the usage of artificial intelligence algorithms, while there was a failure on extracting a rule which can distinguish each virus and figuring out the transmission route, the common features of the sequences were detected. As a result, we could justify the previously studied result on the transmission route. During the procedure, we investigated the basic cause of the failure in consideration of the principle of the algorithms used, and suggested a better path of analysis using artificial intelligence applied to epidemiological studies.

Keywords— Middle East Respiratory Syndrome, MERS, Neural Network, Support Vector Machine, transmission route, epidemiological investigation

I. INTRODUCTION

In 2012, Middle East Respiratory Syndrome (MERS) was first discovered in Saudi Arabia being registered as the first official report of MERS on human. MERS coronavirus, the pathogen of the disease, is known to be an RNA virus in the class of coronaviridae [1], Causing 186 certain cases and 36 deaths from May to July 26, first Korean MERS patient was discovered in Republic of Korea on May 20, 2015 [2][3].

In previous cases of Saudi Arabian MERS-CoV, transmission of the pathogen was relatively well quarantined in contrast with Korean outbreak, where MERS-Cov was spread widely and rapidly by direct transmission among hospitals [2][3]. With the advent of This type of MERS-Cov contagion, keeping track of the genetic mutation of the pathogen through the course of transmission and its transmission route became matters of importance.

In precedential researches, certain samples of MERS-CoV went through RNA extraction and RT-PCR amplification was performed with the extracted RNA to sequence the pieces of them [4]. The data of newly sequenced RNAs and RNAs that are previously sequenced were used as inputs in Molecular Evolutionary Genetics Analysis(MEGA) software, version 5,

to proceed phylogenetic researches and distance calculations [4][5]. Genetic divergence among the RNA samples was also identified by Bayesian inference of MrBayes version 3.2.1. [5].

MERS-Cov samples within different local parts of KSA are analyzed in many studies in order to identify genetic diversity [5]. In the paper of Kim et. al, the genetic mutation among the viruses of KSA and different regions of the world including the pathogen of the outbreak in Republic of Korea was studied to analyze the overall transmission route [6]. But we considered that the rules extracted from amino acid sequences of different MERS-CoV in the study was not accurate enough and incomplete, using only apriori and decision tree algorithm, and apriori algorithm only suggesting the vast majority of Leucine in the sequences.

In this study, the raw data from 15 classes of MERS-CoV was analyzed with Neural Network and Support Vector Machine to produce some useful rules and compared them with previous rules from Apriori and Decision Tree. Then, we tried to draw corresponding rules and apply it to the actual raw data to verify if it has better accuracy and efficiently represent certain part of a sequence. But through the process of algorithmic analysis, it was confirmed that the results from NN and SVM was unsuitable for analyzing MERS-CoV sequences, NN being unable to find distinguishable rules among different viruses and likewise the results of SVM being too close to the normal value, thus lacking classification function.

II. MATERIALS AND METHODS

A. Pathogenesis of MERS-CoV

MERS-CoV, the Middle East Respiratory Syndrome Coronavirus, was first isolated from a patient on September 2012. A pancoronavirus reverse-transcriptase polymerase chain reaction (RT-PCR) assay was used in the isolation process. [7] The infection of the virus can be determined from the simple symptoms of cough, acute respiratory tract infection to multiorgan dysfunction that may lead to death of the patient. [8] In addition, it has been studied that the disease causes renal failure, acute pneumonia and consumptive

coagulopathy, and that most patients with the disease showed respiratory problems [9].

B. Transmission of MERS-CoV

Based on phylogenetic studies, it was originally thought that MERS-CoV is spread from zoonotic events via bats [10]. However, MERS-CoV was later discovered to take the Arabian Camel, *Camelus Dromedarius* as its intermediate host, according to serological studies [11][12]. MERS-CoV enters the host through its spike protein located on the virion surface, which is a type I transmembrane glycoprotein composed of 1353 amino acids [13].

C. Neural Network Algorithm

Artificial NN is a statistical learning algorithm inspired by actual neural network in biology. In general, NN refers to models with problem solving abilities by modulating the intensity of synaptic linkages of artificial neuron (node) network through machine learning [14]. NN could be classed as whether it requires supervised signals (answers) or not. Supervised learning optimize to the problem when provided with supervised signals, and is usually applied to problems that has certain answers. Unsupervised learning is applied to cluster analysis to find representative values in a dataset. NN consists of various perceptron, which produces a binary output from various binary inputs, each with its own weight value according to relative importance. In the process of output calculating from a simple perceptron, each weight and input value are multiplied and summed up, then compared with a certain threshold value to determine the output depending on whether the sum is larger or smaller than the value. By letting the algorithm automatically regulate the weight values, and with the connection of many various perceptron, an artificial NN can react to external stimulations without a programmer's direct arbitration.

D. Support Vector Algorithm

Support Vector Machine(SVM) is a supervised learning model for pattern recognition, data analysis as a field of machine learning, and is usually applied in classification and regression analysis. Given a set of data included in one side of category between the two of them, SVM forms a non-probabilistic binary linear classifier which assigns new data to certain categories based on the previous data set. The classifier is represented by points in space, which clearly divides data from separate categories with a wide gap, and SVM finds the gap that is wide as possible. SVM can be used in linear classification along with non-linear classification. In non-linear classification, the job of mapping the input data into high-dimensional feature spaces is necessary, and Kernel trick is often used to make it efficient.

Generally, SVM is comprised of a hyperplane or a set of hyperplanes that can be utilized in classification or regression analysis [15]. Intuitively, when a hyperplane represents relatively large difference between the closest learned data dot, it can be considered that the classifier error is rather small,

leading to the need to find a hyperplane which represents the farthest position from the closest learned data for a well-classified result.

E. Past Studies

Studies with artificial intelligence were not commonly seen in the field of epidemiology. Then, Kim, et al.(2015) published the study of transmission route of Middle East Respiratory Syndrome-Coronavirus (MERS-Cov) using artificial intelligence algorithm. [6] In the study, researchers have found 15 full DNA sequences collected from different places in the National Center for Biotechnology Information (NCBI). Using Apriori and Decision Tree algorithm, they have analyzed the similarities and differences 15 sequences. From the Decision Tree algorithm, relevance between several sequences have been found. For example, Makkah (Saudi Arabia), Riyadh (South Arabia), Bisha (Saudi Arabia) has been selected by the study as a first source of MERS-Cov.

Our research team has focused on the conclusions made by the Kim, et al.. Those include the similarities of MERS-Cov from Jeddah (Saudi Arabia) and China, those from Hafr-Al-Batin (Saudi Arabia) and Korea, and those from Makkah (Saudi Arabia) and Riyadh (Saudi Arabia). In our study, we have used the Neural Network (NN) Algorithm and the Support Vector Machine (SVM) Algorithm to examine and justify the conclusions of Kim, et al.

III. RESULTS

A. Neural Network

The result analyzed by neural network expresses these kinds of patterns.

MERS coronavirus in South Korea, yielded with 9 windows

class 1 IF : amino1 in {L} ^ amino2 in { } ^ amino3 in {V} ^ amino4 in {W} ^ amino5 in {R} ^ amino6 in {P} ^ amino7 in {H} ^ amino8 in {Q} ^ amino9 in {L} (1)

Designating each line an exemplar, 5072 exemplars were extracted with 9 windows, 4287 with 13 windows, and 3280 with 17 windows. From this data, we intended to extract some rules and apply those to the raw data to check the accuracy. To deal with the vast amount of data with higher brevity and clarity, we calculated the frequency of each amino acid in the 15 classes and took an amino acid with highest frequency in each position.

Table 2 in the Appendix shows the rules of class c(France) we extracted under 9, 13 and 17 windows. The numbers on the tables are the number of the corresponding amino acid in each position. L indicates amino acid leucine, and S indicates serine. In case of position 7 in 17 windows, leucine and serine appeared with the same frequency, 48 times. While rules including C(cysteine), V(valine) were extracted in some classes, a majority of the rules expressed L with S following the lead, not only in class c but also in the other classes. These

rules are just provisional ones, and we have to extract valid rules by analyzing these.

Considering that the rules extracted in preceding researches includes 1 to 3 subrules, we took 2 amino acids from the provisional rules shown above and calculated their frequency. Frequency is yielded by dividing (the number of units satisfying the rule in the class) / (the number of units satisfying the rule in the 15 classes).

TABLE 1. FREQUENCY OF RULES OF CLASS C UNDER 9 WINDOW

Rule	class c	all classes	Frequency
Position 1 = S Position 4 = L	14	121	0.12

While rules are considered valid and meaningful when their frequency exceed 0.7, the rule above shows the frequency of 0.12, which is not high enough for distinguishing the class from the other classes. Not only the rule above, but also most of the rules extracted shows the frequency around 0.1 or 0.2.

This implies that this rule-extraction method is not valid and effective. So we analyzed the reason and the solution considering the qualification for effective rules. Since the frequency value is calculated by comparing the number of corresponding units in the class and in all classes, it is required for the valid rules to include features which are specialized for the class. It means that we have to find particular features of the class being analyzed, distinctive from the other classes. However, the method we used does not reflect this essential qualification for rules. Since MERS-CoV has a number of leucine in its amino acid sequence, most of the rules include L and therefore they cannot work as a criteria of division between the class and the other classes.

In conclusion, it seems that we could get to the conclusion that MERS-CoV sequences have a number of leucine in common, followed by serine. However, the clear differences between the sequences were not found conclusively in this experiment. The reason seems to be that this rule-extraction method is not useful for analyzing amino acid data, especially when dealing with many classes. To solve this inadequacy, a program for calculating conditional probability might be required.

B. Support Vector Machine

After it was found out that it is inappropriate to analyze multi-classes with Neural Network Algorithm, our research team tried to perform the same task using Support Vector Machine Algorithm. In our SVM analysis, we tried to use three types of functions to separate the amino acids into 15 classes as training data and collect some rules, which are normal function, polynomial function, and the Gaussian radial basis function (GRB). However, the computer we used to complete the task was not fast enough to stand operations using polynomial function and GRB function, so there was no choice but to use the normal function.

When we use normal function, the goal is to find a hyperplane that most successfully divides the data sets. In order to achieve this, the algorithm makes each training data into a vector in space. Then it finds some equation of plane

$$\vec{w} \cdot \vec{x} - b = 0$$

that divides the training data successfully, where $\vec{w}$ is the normal vector of the plane and b is some constant. If the data set is successfully divided, the value $\vec{w} \cdot \vec{x} - b$ of each data set should be over 1 or below -1. The two hyperplanes

$$\vec{w} \cdot \vec{x} - b = \pm 1$$

and is called a (hard) margin [14].

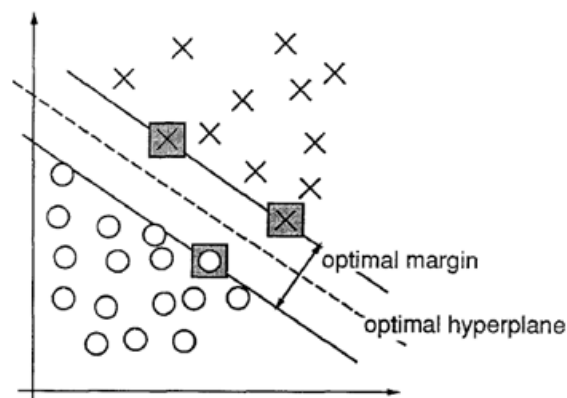


Figure 1. Structure of SVM Algorithm and Margin [14]

However, in our experiment there was no data that matched this criterion, which means that the value $\vec{w} \cdot \vec{x} - b$ of each data was between -1 and 1. From this, we concluded that the analysis was inconclusive, suggesting that our multi-class data sets were too random to get a specific rule to separate them. Also, the algorithm gave us the error percentage of the analysis as 93% so we decided that the data should not be trusted in our project. These difficulties in our analysis means that in using SVM algorithm in epidemiology (at least SVM using normal function), if there are too many classes, it is inappropriate to use multi-class analysis due to the large size of the data and exceptions to consider, which led us to devise a new method using SVM.

C. Results from Double-Class Analysis from SVM

The results from multi-class analysis using SVM was inconclusive because there were too much classes to analyze which meant that there was no appropriate hyperplane that could divide the data into parts. Therefore, our research team selected one of the conclusions from Kim et. al, which is similarity between MERS-CoV from South Korea and Haffr-Al-Batin, Jeddah and China, Makkah and Riyadh [1] and used double-class analysis from SVM to verify the conclusion. In the analysis, we used Gaussian Radial Basis Function (GRB or RBF), and Sigmoid Function (Sig).

1) South Korea and Hafr-Al-Batin

During the experiment, we made data types of <South Korea : +1, Hafr-Al-Batin : -1>. Table 3, Table 4 and Table 5 show the results from SVM with 9 windows, 13 windows, 17 windows respectively. It is easily seen that the accuracy rate of the rules is low, about 50 percent, which entails that the genes of the two viruses are not clearly divided into parts. From the result, we can conclude that the two genes are very alike.

TABLE 2. 9 WINDOW RULE OF SOUTH KOREA AND HAFS-AL-BATIN (ACCURACY ON TEST (%))

	1st	2nd	3rd	4th	5th	Avg
RBF	49.78	53.78	52.89	49.56	46.67	51.51
	50.22	53.11	52.22	54.44	52.44	
Sig	49.59	51.56	54.22	46.22	48.22	50.85
	49.78	53.78	48.67	53.33	53.11	

TABLE 3. 13 WINDOW RULE OF SOUTH KOREA AND HAFS-AL-BATIN (ACCURACY ON TEST (%))

	1st	2nd	3rd	4th	5th	Avg
RBF	52.92	49.58	57.50	57.08	49.58	53.17
	53.75	53.33	53.33	53.33	51.25	
Sig	47.92	47.08	45.83	54.17	47.08	49.54
	50.00	46.25	52.50	49.58	55.00	

TABLE 4. 17 WINDOW RULE OF SOUTH KOREA AND HAFS-AL-BATIN (ACCURACY ON TEST (%))

	1st	2nd	3rd	4th	5th	Avg
RBF	51.33	53.00	53.00	55.00	51.33	51.37
	50.33	54.00	46.33	48.00	51.33	
Sig	51.33	49.00	49.33	50.33	51.67	50.20
	50.33	53.67	46.67	50.67	49.00	

2) Jeddah and China

Corresponding with the experiment in South Korea and Hafar-al-batin, we set the data types using the standard of Jeddah as +1 and China as -1. In table 7, table 8, and table 9, the 9, 13, and 17 window results of SVM algorithm are represented in percentage unit. Among the accuracy rates from 10 experiments, average accuracy rate of RBF was the highest in all windows. The accuracy rates from most of the functions(Normal, Polynomial, Sig) are relatively low closing near 50.00, stating that data of MERS-Covs from Jeddah and China are difficult to clearly divide into certain two sections, which means the two viruses are likely to have similar amino acid sequences, thus genetically related. As an exception, The results from RBF in 17 windows show relatively larger accuracy rates than other figures.

TABLE 5. 9 WINDOW RULE OF CHINA AND JEDDAH (ACCURACY ON TEST (%))

	1st	2nd	3rd	4th	5th	Avg
Normal	54.44	50.00	48.44	45.78	47.78	49.84
	52.44	48.00	50.89	47.56	53.11	
Poly	54.44	50.00	48.44	45.11	47.78	49.82
	53.33	48.67	50.00	47.11	53.33	
RBF	58.00	50.22	50.44	49.78	50.00	51.56
	49.56	51.33	53.56	53.33	49.33	
Sig	54.44	50.00	49.56	48.44	47.78	49.98
	48.22	46.67	51.33	51.33	52.00	

TABLE 6. 13 WINDOW RULE OF CHINA AND JEDDAH (ACCURACY ON TEST (%))

	1st	2nd	3rd	4th	5th	Avg
Normal	50.00	51.00	54.33	54.33	50.00	51.17
	51.00	49.00	49.00	55.33	47.67	
Poly	50.33	51.67	54.67	55.67	51.67	51.77
	49.33	50.33	49.00	57.00	48.00	
RBF	53.33	56.67	53.67	57.67	54.33	53.10
	50.67	54.67	50.33	51.33	48.33	
Sig	50.00	49.67	49.33	49.67	44.00	50.13
	53.67	50.33	50.33	53.33	49.00	

TABLE 7. 17 WINDOW RULE OF CHINA AND JEDDAH (ACCURACY ON TEST (%))

	1st	2nd	3rd	4th	5th	Avg
Normal	57.08	50.42	48.75	51.67	55.83	53.29
	58.33	49.58	58.33	52.50	50.42	
Poly	56.25	50.83	50.83	50.00	56.25	54.21
	60.42	52.50	60.42	53.33	51.25	
RBF	77.08	73.75	72.50	75.42	76.67	76.25
	75.83	84.17	74.17	77.08	75.83	
Sig	52.08	52.92	53.75	50.83	53.33	50.25
	45.00	45.00	50.00	52.50	47.08	

3) Makkah and Riyadh

The last result in Kim et al. that we focused on was relationship between data from Makkah (Saudi Arabia) region and Riyadh (Saudi Arabia). The research argued that the two regions have high co-relationship not only between each other but also with many other regions so there is high chance that the two regions are the center of the epidemic. In our study, we decided to verify the fact that data from the two regions have high similarity between each other.

In the experiment, double-class SVM analysis were used to do the operations. Specifically, the normal function, polynomial function, radial basis function and sigmoid function were used in SVM and we were able to get the accuracy value of each operation. Also, we made data types of <Makkah : +1, Riyadh : -1>. Table 9 represents the 9 window results, Table 10 represents the 13 window results, and Table 11 represents the 17 window results from the analysis.

The accuracy results are about half, like the results of previous experiments, which indicates that data from the two classes are similar to one another. Notable feature from the result was that accuracy yielded using RBF is unusually high.

TABLE 8. 9 WINDOW RULE OF MAKKAH AND RIYADH (ACCURACY ON TEST (%))

	1st	2nd	3rd	4th	5th	Avg
Normal	48.44	49.33	50.44	49.78	48.44	49.84
	51.11	51.78	51.11	49.11	48.89	
Poly	48.67	49.33	50.67	48.67	48.89	50.22
	53.56	51.33	51.33	52.89	46.89	
RBF	75.11	77.56	73.56	75.56	76.44	76.07
	76.44	75.33	76.44	77.11	77.11	
Sig	48.00	47.56	54.00	51.78	48.22	49.69
	53.11	50.44	45.11	52.89	45.78	

TABLE 9. 13 WINDOW RULE OF MAKKAH AND RIYADH (ACCURACY ON TEST (%))

	1st	2nd	3rd	4th	5th	Avg
Normal	47.33	49.67	55.67	47.00	54.67	51.03
	53.00	45.00	53.00	54.67	50.33	
Poly	48.67	50.00	56.33	47.00	55.00	51.50
	54.00	46.33	53.00	53.67	51.00	
RBF	76.33	76.33	81.33	77.00	77.33	77.00
	75.00	77.67	76.33	78.33	74.33	
Sig	46.67	51.00	47.67	47.00	51.67	49.80
	51.67	51.33	48.00	50.00	53.00	

TABLE 10. 17 WINDOW RULE OF MAKKAH AND RIYADH (ACCURACY ON TEST (%))

	1st	2nd	3rd	4th	5th	Avg
Normal	57.08	50.42	48.75	51.67	55.83	53.29
	58.33	49.58	58.33	52.50	50.42	
Poly	56.25	50.83	50.83	50.00	56.25	54.21
	60.42	52.50	60.42	53.33	51.25	
RBF	77.08	73.75	72.50	75.42	76.67	76.25
	75.83	84.17	74.17	77.08	75.83	
Sig	52.08	52.92	53.75	50.83	53.33	50.25
	45.00	45.00	50.00	52.50	47.08	

IV. CONCLUSION

Virus. The bitterest enemy of the mankind. It comes without any sound, killing thousands and thousands. Viruses are dreadful in that we do not know where they are now at this precise moment, and where they will spread in the recent future. And that is why viral epidemiology is of great value.

With epidemiological studies, we can review former infectious diseases about the process of viral mutation and in some cases, figure out and predict the transmission of an ongoing epidemic.

The study of Kim et. al [6] devised a new way of analyzing transmission route of MERS-Cov, focusing on the differences in DNA sequences. In their study, they used Apriori Algorithm and Decision Tree to compare 15 different MERS-Cov from different regions. However, through decades of development, many types of artificial intelligence algorithm were developed, other than the two used, which led us to this study, where we used two other algorithms of artificial intelligence to verify the results from different algorithms and to find out which algorithm is best suitable for the epidemiological analysis.

In our study, we used Neural Network and Support Vector Machine. However, the result was that the two algorithms we used are not appropriate for the epidemiological study. First, for the Neural Network Algorithm, the extracted rules did not meet the qualification required of a rule, that it has to figure out how a certain class is different from the others. The rules appeared similar in all classes. In the Support Vector Machine, using 15 classes as an input, the algorithm could not produce meaningful results.

However, using SVM with double class analysis, we were able to find relevance between certain two sequences, verifying the conclusions made from Kim et. al. The results of all 4 SVM functions suggested that the amino acid sequences of MERS-Cov from South Korea and Hafar-al-Batin, Jeddah and China, Riyadh and Makkah are difficult to distinguish per each pair, thus genetically analogous, which proves the conclusion from the Kim et. al. that these MERS-Covs from each pair are genetically linked.

In this research, we discovered that in using SVM in epidemiology, it is more efficient to apply the algorithm in a double class analysis of 2 viruses than the comparative research of many viruses at the same time. And as a further research, it is thought that using SVM to find out the differences between two certain classes would be worthwhile. Our research team hopes the analysis of applying artificial intelligence and machine learning in epidemiology used in this study to be useful in other epidemiological studies of Big Data field.

REFERENCES

- [1] de Groot RJ, Baker SC, Baric RS, Brown CS, Drosten C, Enjuanes L, Fouchier RA, et al. "Middle East respiratory syndrome coronavirus (MERS-CoV): announcement of the Coronavirus Study Group." *J Virol*. 2013;87:7790–7792.
- [2] Moran Ki, "2015 MERS outbreak in Korea: hospital-to-hospital transmission", 2015; 37: e2015033
- [3] WHO 2015, "Summary of Current Situation, Literature Update and Risk Assessment", WHO/MERS/RA/15.1
- [4] Esam I. Azhar, Ph.D., Sherif A. El-Kafrawy, Ph.D., Suha A. Farraj, M.Sc., Ahmed M. Hassan, M.Sc., Muneera S. Al-Saeed, B.Sc., Anwar M. Hashem, Ph.D., and Tariq A. Madani, M.D. 2014, "Evidence for Camel-to-Human Transmission of MERS Coronavirus." *N Engl J Med* 2014; 370:2499-2505
- [5] Cotten M, Watson SJ, Zumla AI, Makhdoom HQ, Palser AL, Ong SH, Al Rabeeah AA, Alhakeem RF, Assiri A, Al-Tawfiq JA, Albarrak A, Barry M, Shibl A, Alrabiah FA, Hajjar S, Balkhy HH,

- Flemban H, Rambaut A, Kellam P, Memish ZA. 2014, "Spread, circulation, and evolution of the Middle East respiratory syndrome coronavirus." mBio.01062-13.
- [6] Kim, Donghyeon, Soyoung Hong, Sungwoo Choi, and Taeseon Yoon. "Analysis of Transmission Route of MERS Coronavirus Using Decision Tree and Apriori Algorithm." *2016 18th International Conference on Advanced Communication Technology (ICACTION)*(2016): n. pag. Web.
- [7] Zaki AM1, van Boheemen S, Bestebroer TM, Osterhaus AD, Fouchier RA, "Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia." N Engl J Med. 2012; 367(19):1814-20.
- [8] Prasannavenkatesh Durai, Maria Batool, Masaud Shah, and Sangdun Choi, "Middle East respiratory syndrome coronavirus: transmission, virology and therapeutic targeting to aid in outbreak control"
- [9] Who "Mers-Cov Research G State of knowledge and data gaps of Middle East respiratory syndrome coronavirus (MERS-CoV) in humans" PLoS Curr
- [10] Woo PC, Lau SK, Li KS, Tsang AK, Yuen KY. "Genetic relatedness of the novel human group C betacoronavirus to Tylonycteris batcoronavirus HKU4 and Pipistrellus bat coronavirus HKU5." Emerg Microbes Infect. 2012;1:e35.
- [11] Kayali G, Peiris M. "A more detailed picture of the epidemiology of Middle East respiratory syndrome coronavirus." Lancet Infect Dis. 2015;15:495-497.
- [12] Milne-Price S, Miazgowicz KL, Munster VJ. "The emergence of the Middle East respiratory syndrome coronavirus." Pathog Dis. 2014;71:121-136.
- [13] Xia S, Liu Q, Wang Q, Sun Z, Su S, Du L, Ying T, Lu L, Jiang S. "Middle East respiratory syndrome coronavirus (MERS-CoV) entry inhibitors targeting spike protein." Virus Res. 2014;194:200-210.
- [14] Cortes, C.; Vapnik, V. (1995). "Support-vector networks". Machine Learning. 20 (3): 273-297. doi:10.1007/BF00994018



Soyoung Hong was born in Anyang, Republic of Korea, in 1999 and currently attends the Natural Science Department of the Hankuk Academy of Foreign Studies. She has a strong interest in biology and biochemistry. She has been recently interested in bio-informatics and analyzing DNA sequences of viruses and started her first research in 2015.



Sungwoo Choi was born in Suwon, Republic of Korea, on March 13, 1999. He is currently a student in a natural science major of Hankuk Academy of Foreign Studies, Yongin, Republic of Korea. He is mostly interested in neurobiology, and recently studying bioinformatics, researching various topics of immunology utilizing bioinformatics techniques.



Donghyun Kim was born in Andong, Republic of Korea, on September 12th, 1999 and currently attends the Natural Science Department of the Hankuk Academy of Foreign Studies. He's major interest is in biology, neuroscience, and biomimicry. He recently studied immunology and bioinformatics and has written a few theses about the topic including this one.



Taeseon Yoon was born in Seoul, Korea, in 1972. He was Ph.D. Candidate degree in Computer education from the Korea University, Seoul, Korea, in 2003. From 1998 to 2003, he was with EJB analyst and SCJP. From 2003 to 2004, he joined the Department of Computer Education, University of Korea, as a Lecturer and Ansan University, as an Adjunct professor. Since December 2004, he has been with the Hankuk Academy of Foreign Studies in Yongin, Gyeonggi, where he is a Computer Science and Statistics Teacher. He was the recipient of the Best Teacher Award of the Science Conference, Gyeonggi-do, Korea, 2013.

APPENDIX

TABLE 2. PROVISIONAL RULE EXTRACTION OF MERS COV IN CLASS C

pos	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
9	S 72	L 79	L 84	L 100	S 84	L 86	S 76	L 90	L 82								
13	L 58	L 66	L 61	L 65	L 67	L 69	S 55	L 56	L 61	L 59	L 67	S 61	S 64				
17	S 61	L 54	L 51	S 47	L 54	L 44	LS 48	L 51	L 52	L 57	L 55	L 49	L 50	S 49	S 46	S 55	S 55