

Intensified Analysis and Comparison of 5 Flaviviruses with the use of Decision Tree and Support Vector Machine (SVM)

Eujin Yang*, Bokyoung Gu*, Taeseon Yoon**

*Natural Science, Hankuk Academy of Foreign Studies, Young-in, South Korea

** Faculty, Hankuk Academy of Foreign Studies, Young-in, South Korea

eujin0001@hafs.hs.kr, jpj27@naver.com, tsyoon@hafs.hs.kr

Abstract— Flavivirus is spreaded with the help of intermediary, especially mosquitoes. In preceding research, we found out that Leucine has high frequency. Wanting to know specific relationship between 5 flaviviruses ; Yellow fever, West Nile virus, Dengue virus, Tick borne encephalitis, decision tree and support vector machine algorithm were used. Analyzing results of the algorithms, difference or similarity about the viruses and a group as flavivirus were found.

Keywords— Zika virus, Yellow fever, West Nile virus, Dengue virus, Tick borne encephalitis, Flavivirus, Decision tree algorithm, Support Vector Machine(SVM)

I. Introduction

Flavivirus is the virus which is spread mostly with the help of mosquitoes. Flavivirus can be divided into 3 parts. One is the type spread by mosquitoes, another is the type spread by tick, and the other is the type which does not know the intermediary. 25 viruses including Dengue, Yellow fever are spread by mosquitoes and 14 viruses including Russian spring-summer encephalitis are spread by ticks. And 16 viruses are spread by unknown intermediary. And especially patients of Zika virus keep appearing. Having felt the seriousness of the virus and necessity of treatment, we conduct an experiment. Last experiment using apriori algorithm we compared and contrasted flaviviruses, and found out that Glutamine and Leucine show high frequency. But we cannot find out whether differences or similarities exactly exist. So by using decision tree, we expect to find out more specific relationship between 5 types of flavivirus. Also by using SVM, if decision tree's result shows no relationship between the viruses, we are able to find out similarity between viruses as whole. And the result will lead us to the conclusion of which standards make the viruses belonged to flavivirus.

II. Materials and Methods

Materials used in this analysis are flaviviruses ; Zika Virus, Tick Borne Encephalitis, Yellow Fever Virus, Dengue Virus, and West Nile Virus. Their protein sequences were collected from the National Center for Biotechnology

Information (NCBI). And we use decision tree and support vector machine(SVM) to proceed this analysis.

A. Zika Virus

Zika Virus was first discovered in 1947, in Uganda rhesus monkey. This virus is not spread by a routine contact, but by mosquitoes. After three to seven days being infected with it, there are only slightest symptoms like rash, muscle pain, and acute fever. But this virus can lead the infector to have microcephaly. So it is dangerous for women who are pregnant or suspected to be pregnant and many countries are warning seriousness of the virus.[16]

B. West Nile Virus

West Nile virus is shared to human mainly by mosquitoes, but people can also infected by horses, sparrows and crows. People having the virus can experience seizure, feeling stiff, and headache which are appeared after 2~14 days of conducting. Not only these, the virus can harm the brain's state of the central nervous system.[3], [7], [11-12], [16]

C. Tick Borne Encephalitis

Tick Borne Encephalitis(TBE) is transmitted mostly by mite, but unpasteurized milk, goat whole fluid milk, or sheep's milk can transmit it, too. The virus shows symptoms like lack of appetite, headache, vomiting in the incipient stage after incubation period about 7~14 days. And symptoms of the central nervous system appear later. [2], [9], [16]

D. Yellow Fever

Yellow Fever can be caused by mosquitoes. The virus is common in some countries such as North America. After being conducted by three to six days, infector can show signs of fever, cold fit and bleeding in the mouth, eyes, and gastrointestinal tract in toxic cases. [7], [16]

E. Dengue Fever

Dengue virus exists in *Aedes albopictus*' saliva. If the mosquito sucks blood, it can lead the virus to go inside human's body. The virus has incubation period about 5~7 days, and symptoms such as fever, skin rash, severe muscular pain can be shown to infected person. [3], [5], [7], [16]

F. Decision Tree

Decision tree is one of the familiar methods of data mining. It draws leaves, branches and the method is called branching. By that, the method generated model which has its roots in some inputted pronumeral and brings target variable. The algorithm has internal node matched to inputted variables and each branch matches to outcome which can be possible by the inputted ones. Decision tree keeps extending by attaching the input in. This progress continues until subset node is equivalent to variable that is targeted or new value when it cannot be appended because of separation. The algorithm provides successful way to discover gap between comparison target. Therefore, in our experiment, decision tree algorithm is used for the purpose of finding the relationship between 5 types of flavivirus. [3], [5], [7-8], [10], [13], [15-16]

G. Support Vector Machine (SVM)

SVM(Support Vector Machine) is one type of the machine learnings. It is supervised learning model that analyses data used for classification and regression analysis. When given data belongs to a category from 2 other categories, SVM algorithm makes non-probabilistic binary linear classified model based on given data. This classified model's data is expressed as boundary. SVM is algorithm to find a boundary that has largest width. SVM can be used at linear classification and non-linear classification. [1], [3-4], [6], [13-15] In this study, we use 4 functions of SVM: Normal function, Sigmoid function, RBF and Polynomial function. Normal function uses straight line in plane. Sigmoid function uses curve line in plane. RBF is terraced function made in plane. Polynomial function uses space which is raised from one dimension.

III. Result

A. Decision Tree Algorithm

1. Decision Tree Algorithm (9-window)

class (virus)	Dengue virus	TBE	West Nile	Yellow Fever	Zika Virus
Dengue virus	86	78	92	67	54
TBE	81	64	89	82	64
West Nile	96	68	86	70	62
Yellow Fever	111	76	76	70	46

Zika virus	99	87	87	68	40
------------	----	----	----	----	----

Analyzing table 1, Dengue virus has more similar relation between other viruses. In contrast, Zika virus has little relation to others. Dengue virus, TBE, West Nile, and Yellow fever have higher value which shows how much the viruses have individual peculiarities. But Zika virus is different. Looking at Zika virus, there is shortage of characteristics of its own.

2. Decision Tree Algorithm (13-window)

class (virus)	Dengue virus	TBE	West Nile	Yellow Fever	Zika Virus
Dengue virus	59	53	42	42	65
TBE	52	67	43	35	66
West Nile	62	58	39	41	65
Yellow Fever	60	51	49	36	67
Zika virus	54	50	39	50	71

Measured against table 1, the number of Zika virus' rules grows at table 2. Zika virus shows closer relation with other viruses and more own characteristics. Not only Zika virus, but also Dengue virus and Tick Borne Encephalitis have their own peculiarities. On the other hand, viruses like Yellow Fever and West Nile show lack of their own characteristics and have little relationship with others.

3. Decision Tree Algorithm (17-window)

class (virus)	Dengue virus	TBE	West Nile	Yellow Fever	Zika Virus
Dengue virus	38	37	60	35	30
TBE	37	35	42	45	42
West Nile	42	41	36	40	43
Yellow Fever	30	45	45	36	45
Zika virus	36	50	40	35	41

In table 3, especially, West Nile has high rates of similarity toward Dengue virus. Discluding that results, relationship between viruses of 4 types is low. And all 5 viruses have low degree of their own characteristics.

B. Support Vector Machine(SVM)

4. Support Vector Machine (9-window)

	Results(%)	average(%)
Normal	80.00, 77.67, 72.00, 83.67, 82.00, 78.33, 84.00, 78.67, 80.33, 77.67,	74.434
RBF	37.33, 34.00, 36.00, 37.67, 33.67, 34.00, 38.33, 33.33, 34.33, 33.67,	35.233

Sigmoid	80.67, 72.67, 79.00, 78.00, 80.00, 82.67, 80.00, 81.67, 76.00, 78.33,	78.901
poly	78.33, 78.00, 78.00, 78.00, 77.67, 77.33, 76.67, 80.33, 81.67, 77.33,	78.333

9-window SVM algorithm results are shown in table 4. The average error rate of the experiment is highest in Sigmoid and accuracy rate of the experiment is highest in RBF. Analyzing the table, the error percentage is very high which means accuracy is very low about 20 percent. So it can be concluded that the flaviviruses have little similarity.

5. Support Vector Machine (13-window)

	Results(%)	average(%)
normal	80.00, 81.20, 77.60, 78.00, 82.40, 77.60, 81.60, 80.40, 74.00, 84.80,	79.706
RBF	36.40, 38.40, 35.60, 42.40, 36.80, 32.80, 39.20, 38.00, 36.80, 39.20,	37.56
Sigmoid	76.80, 80.40, 82.40, 77.60, 77.20, 77.20, 76.40, 81.20, 82.80, 80.80,	79.28
poly	78.50, 86.50, 79.00, 76.50, 82.00, 78.50, 81.00, 79.00, 82.00, 80.00,	80.30

13-window SVM algorithm results are represented in table 5. The average error rate of the experiment is highest in normal and accuracy rate of the experiment is highest in RBF. Results of 9-window are quite same as 13-window. Also this table shows that flaviviruses have different characteristics.

6. Support Vector Machine (17-window)

	Results(%)	average(%)
normal	79.00, 84.00, 80.00, 79.50, 81.50, 79.00, 77.50, 82.00, 83.50, 81.50,	80.75
RBF	38.50, 35.30, 30.50, 39.00, 39.50, 37.50, 38.00, 37.00, 32.50, 37.00,	36.5
Sigmoid	78.00, 86.00, 77.50, 79.00, 80.50, 83.00, 80.00, 82.00, 76.50, 76.00,	79.85
poly	78.00, 80.80, 80.00, 76.80, 76.00, 76.80, 82.80, 79.20, 77.60, 78.80,	78.68

In table 6, 17-window SVM algorithm results are shown. The average error rate of experiment is highest in normal and accuracy rate of experiment is highest in RBF, same as table 5. According to these results, all flaviviruses (Dengue, Yellow fever, Tick Borne Encephalitis, West Nile, and Zika virus) do not share many characteristics and have properties of their own.

IV. Discussion and Conclusion

Concluding that the further study have to be done from

preceding experiment, we use other algorithms to compare 5 types of flavivirus : Zika virus, Yellow fever, Tick Borne Encephalitis, West Nile virus, and Dengue virus. The algorithms are decision tree and support vector machine(SVM). After sequences of protein were collected, experiments which were split into 3 types : 9-window, 13-window and 17 window were done to compare and contrast the viruses. Decision tree was used to find out exact correlation between the viruses. Looking in 9-window decision tree, Zika virus was different with other viruses. And most of the viruses have their own characteristics. The result means that each virus is very distinctive and has less relationship with other. Wanting to know the existence of commonness that made the viruses belonged to the same category, we use support vector machine (SVM). Analyzing the results from the 4 functions of SVM : normal, sigmoid, poly and RBF, we found that they have less rate of similarity. Values of 9, 13, and 17 window were very alike and accuracy rate of experiment is always highest in. But other 3 functions have very high average error rate, the viruses have different peculiarities and cannot be divided into parts. And these experiments indicate that 5 types of viruses have little or no relationship and they are classified into the same group without similarities like genetic sequences. Further, we cannot treat each of the flavivirus as the same virus and cannot use the same vaccine to cure and prevent the virus. They are just categorized as flavivirus just because of vehicle which is an insect.

REFERENCES

- [1] Chaeyun Jung, Yonghyun Park, Seunghui Han, and Taeseon Yoon, "tion to Hand, Foot and Mouth Disease(HFMD) Using Apriori Algorithm, Decision Tree and Support Vector Machine (SVM)", International Conference on Intelligent ICIC. 2015. (SVM)
- [2] Daniela Amicizia, Alexander Domnich, Donatella Panatto, Piero Luigi Lai, Maria Luisa Cristina, Ulderico Avio, and Roberto Gasparini, "Epidemiology of tick-borne encephalitis (TBE) in Europe and its prevention by available vaccines", U.S. National Library of Medicine. 2013 (tick borne)
- [3] Donghyun Lee, Taeseon Yoon, "Analysis of the Genomes of Chikungunya Virus and Dengue Virus Using Decision Tree, Apriori Algorithm, and Support Vector Machine.", International Conference on Electronics Engineering and Informatics ICEEI. 2016. (dengue, SVM, decision tree)
- [4] Hyorin Park, Yoojin Park, Yerin Moon, and Taeseon Yoon, "Comparison of CIV, SIV and AIV using Decision Tree and SVM", MATEC Web of Conferences MATEC Web Conf. 2016. (SVM)
- [5] Hyunseong Kim, Juyoung Yoo and Taeseon Yoon, "An Analysis of the Genomes of Dengue Virus Using Decision Tree and Apriori Algorithm", International Conference on Future Computer and Communication ICFCC. 2016. (dengue, decision tree)
- [6] Suta Song, Zhichao Zhan, Zhiying Long, Jiakai Zhang, and Li Yao1, "Comparative Study of SVM Methods Combined with Voxel Selection for Object Category Classification on fMRI Data", U.S. National Library of Medicine. 2011. (SVM)
- [7] Seung Hye Song, Yijeong Choi, Taeseon Yoon, "Comparison of episodes of mosquito borne disease: Dengue, Yellow Fever, West Nile, and Filariasis with Decision tree, Apriori Algorithm", International Conference and Advanced Communications Technology ICACTION. 2016. (dengue, west nile, yellow fever, decision tree)
- [8] Jiwon Song and Taeseon Yoon, "Analysis of Mitochondrial Hsp70 Homolog Amino Acid Sequences of Amitochondriate Organisms

- Using Apriori and Decision Tree", Lecture Notes in Computer Science. (decision tree)
- [9] Petra Bogovic and Franc Strle, "Tick-borne encephalitis: A review of epidemiology, clinical characteristics, and management", U.S. National Library of Medicine. 2015. (tick borne)
- [10] [10] Taehwan Kim, Taeseon Yoon, "Artificial Neural Network Hybrid Algorithm Combined with Decision Tree and Table", International Journal of Machine Learning and Computing IJMLC. 2015. (decision tree)
- [11] Tonya M. Colpitts,a, Michael J. Conway,a, Ruth R. Montgomery,b, and Erol Fikrigcorresponding authora,c, "West Nile Virus: Biology, Transmission, and Human Infection", U.S. National Library of Medicine. 2012. (west nile)
- [12] William K. Reisen, "Ecology of West Nile Virus in North America", U.S. National Library of Medicine. 2013. (west nile)
- [13] Yihyun Roh, Seokhyun Yoon, Minyoung Lee, Seongpil Jang, Taeseon Yoon, "Analysis and Comparison of Genomes of HIV-1 and HIV-2, Using Apriori Algorithm, Decision Tree, and Support Vector Machine", International Conference on Intelligent ICIC. 2016. (SVM, decision tree)
- [14] Yi Zhang, Jinchang Ren, and Jianmin Jiang, "Combining MLC and SVM Classifiers for Learning Based Decision Making: Analysis and Evaluations", U.S. National Library of Medicine. 2015. (SVM)
- [15] Younghoon Cho, Seungwon Burm, Nayoung Choi, Taeseon Yoon, "Analysis of Human Papillomavirus Using Data Mining - Apriori, Decision Tree, and Support Vector Machine (SVM) and its Application Field", MATEC Web of Conferences. 2016. (SVM, decision tree)
- [16] Youjin Yang, Bokyung Gu, and Taeseon Yoon, "Deeper understanding of Flaviviruses including Zika virus by using Apriori Algorithm and Decision Tree", MATEC Web of Conferences MATEC Web Conf. 2016.



You Jin Yang was born in Gyeonggi, South Korea in 1999. She is now in Hankuk Academy of Foreign Studies. She feels an interest in flavivirus especially zika virus and bio informatics. So based on a paper of analyzing 5 types of flavivirus using apriori algorithm which she wrote she writes another monograph. And it is about 5 types of flavivirus compared by decision tree and support vector machine



Bokyung Gu was born in Seoul, South Korea in 1999. She majors in science at Hankuk Academy of Foreign Studies. She is interested in viruses and bio-informatics. So in this research, she analysed 5 flavivirus by using decision tree algorithm and SVM algorithm.



Taeseon Yoon was born in Seoul, Korea, in 1972. Hereceived the Ph.D. candidate degree in computereducation from the Korea University, Seoul, Korea, in2003. From 1998 to 2003, he was with EJB analystand SCJP. From 2003 to 2004, he joined theDepartment of Computer Education, University ofKorea, as a lecturer and Ansan University, as anadjunct professor. Since December 2004, he has beenwith the Hankuk Academy of Foreign Studies, where he was a computerscience and statistics teacher.