

# Deeper understanding about the genetic structure of dengue virus using SVM

Subin Choi, Hyunjin Choi and Taeseon Yoon

*Natural Sciences, Hankuk Academy of Foreign Studies, HAFS, Yongin-si, Republic of Korea.*

**Abstract.** Dengue fever, mainly found in the tropical and subtropical regions, is carried by mosquitoes. With the help of greenhouse effect, places considered to be a Dengue safe-zone are becoming more and more dangerous. Dengue fever shows similar aspects to MERS, which caused heavy casualties in South Korea; Dengue virus does not have clear treatments nor vaccines like MERS. Development of Dengue vaccine is actively investigated lately. However, it is not easy to succeed; the fact that Dengue's 4 serotypes have different properties and that repeated infections worsen the symptoms. This research aims to analyze the 4 serotypes (DENV1, DENV2, DENV3, DENV4) using SVM and ANN algorithms to investigate the constraints in the development of Dengue's vaccines and treatments.

## 1 Introduction

Dengue fever is an acute idiopathic disease caused by the dengue virus which belongs to the Flavivirus family. Dengue virus is native to the Southeast Asia, and is spread all over the tropical and subtropical regions around the world. Dengue becomes prevalent periodically and causes many fatalities. With the help of greenhouse effect, mosquitoes' population has proliferated (dengue is carried by mosquitoes) and it is continuously developing its influence.

In 2008, *Ae. albopictus* which is one of the mosquitoes that communicates Dengue virus was found in Jeju Island, South Korea. Also, 8 out of 35 Korean volunteers were infected at Colombo, Sri Lanka. South Korea is currently at risk since it was considered as a safe zone from Dengue fever. The fact that Dengue doesn't have clear treatments and vaccines classifies Dengue as a high-risk disease. Dengue virus has 4 serotypes (DENV1, DENV2, DENV3, DENV4). Even if DENV1 infections are fully cured, reinfection of a DENV2, DENV3, DENV4 can be fatal [1]. This paper compares and analyzes the 4 serotypes using SVM and Neural Network Algorithm to investigate what delayed the development of the vaccines and to suggest development direction of the vaccine based on the similarities and differences of the 4 serotypes.

Dengue virus has 4 serotypes; DENV1, DENV2, DENV3, DENV4. Like the Dengue virus, Herpesvirus has about 80 different types of serotype. Out of these, eight of them are expressed on humans. Unlike the Dengue virus, where different vaccines are used to prevent the illness for the 4 distinct serotypes, Herpesvirus' serotypes show different symptoms and are expressed in different parts of the body.

## 2 Materials and procedures

### 2.1 Dengue virus

Since Dengue Virus has been found in the 1950s in Philippines and Thailand, cases are constantly reported in most of Asia and Latin America. It is usually carried by a mosquitoes; *Aedes aegypti*, *Ae. albopictus*, *Ae. polynesiensis*, *Ae. pseudoscutellares* [2]. Dengue virus belongs to Flaviviridae, Flavivirus family and has 4 types of different serotypes; DENV1, DENV2, DENV3, DENV4 [3].

The structure of the gene is similar to those of the Flavivirus family. It has about 100 bases at both 5'-end and 3'-end, and noncoding regions (NCR) that contain about 600 bases. Proteins are expressed by a single open reading frame (ORF). 3 structural proteins (capsid (C), premembrane (prM), envelope (E)) and 7 nonstructural proteins (NS1, NS2a, NS2b, NS3, NS4a, NS4b, NS5) are manufactured since the start of the translation stage and during the post-translation processes.

Dengue Virus 1 (DENV1) is most common among the 4 serotypes. Infections by DENV1 have the most serious symptoms [4],[5],[6]. Recently, Dengue virus infections are caused by superinfections or by the genetically mutated viruses rather than mono-infections.

### 2.2 SVM

SVM (Support Vector Machine), a field of Machine learning, is a model of supervised learning used for pattern recognition and data analysis. It is an approximated implementation which minimizes the

structural jeopardy of the previous algorithms such as neural network algorithm [7], [8]. Data of a Test error (its range is decided by the sum of the learning errors' ratio) and a value which relies on the VC-dimension of a learning machine. It is generally used for classification and regression analysis, and is composed of hyperplane or its sets.

SVM divides the data into two groups to analyze with a Maximum margin hyper plane. Each vectors or points created at this stage is connected to form a polygon; convex hull [9]. These lines or polygons serve as a tool to divide the data into multiple groups and to compare the accuracy. The higher the accuracy, the two groups have considerable similarities. The lower the accuracy, the two have quite different properties.

### 2.3 Procedure

Genetic sequences of DENV1, DENV2, DENV3, DENV4 are extracted from the National Center for Biotechnology Information(NCBI). Experiments are carried out 10 times (10 fold cross validation) using SVM's main 4 functions; normal, poly, RBF, and sig. Each sequences of DENV1, DENV2, DENV3, DENV4 is divided into 10 groups and one from each is selected. The unselected 9 groups are put into the SVM algorithm and decide an invariable number representing the 'difference'. With this number, a single pre-selected group is classified using 4 functions, and the accuracy of the classification vector is calculated. This procedure is repeated for 10 groups. In other words, 10 experiments are done per function. 9 window, 13window, 17 window is decided on how many sections the data are divided. The experiments are repeated 3 times per window. For each window, the average of the 4 functions' (sig, RBF, normal. poly) accuracy is found. These values are used to figure out the properties of the classification.

## 3 Results

For each 4 functions (normal, poly, RBF, sig), the experiments were repeated for each window (9 window, 13 window, 17 window). Normal function is involved with the linearity; poly uses polynomial to figure out the nonlinear part; Gaussian function is used for RBF to classify the nonlinears; sig uses sigmoid function to classify the linear sections. DENV1, DENV2, DENV3, DENV4 can be classified using the function that has the highest average of accuracy. The figures below are the average of accuracy for each window.

### 3.1 9 window

We can recognize the similarities between the values of sig, normal and poly. The values of sig, normal and poly are 76.03498, 74.24741, 74.46065. When the sequences were classified with the 3 functions, the accuracy of the classification vectors were relatively high; all functions' data were higher than 74%. Using these data, we could figure out that each serotypes have significant differences in their properties (see Figure 1).

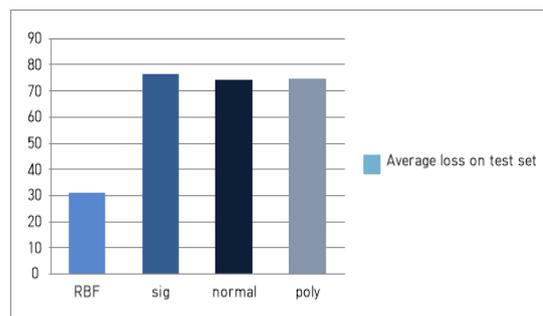


Figure 1. Average of accuracy for 9 window.

### 3.2 13 window

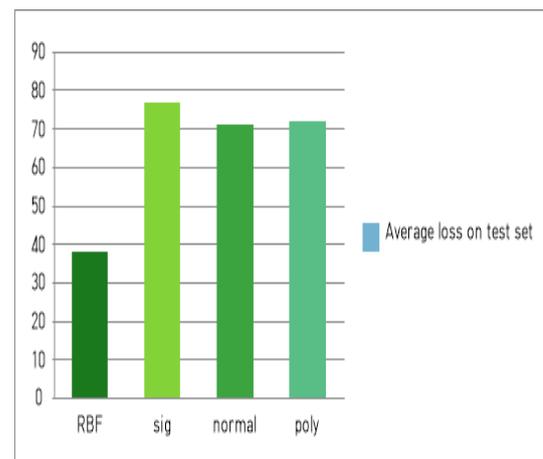


Figure 2. Average of accuracy for 13 window.

According to the results above, it is easy to recognize that sig, normal, poly has higher results than RBF. Therefore, it is obvious that when the sequences of Dengue fever were classified with 13 window, it can be divided nonlinearly and linearly (see Figure 2).

### 3.3 17 window

The result of the 17 window shows similar data with the 2 experiments above. Interestingly, the data of normal form the 3 experiments show values over 70%. Through the last experiment, we could confide that the classification of the Dengue virus is possible using the 4 functions (see Figure 3).

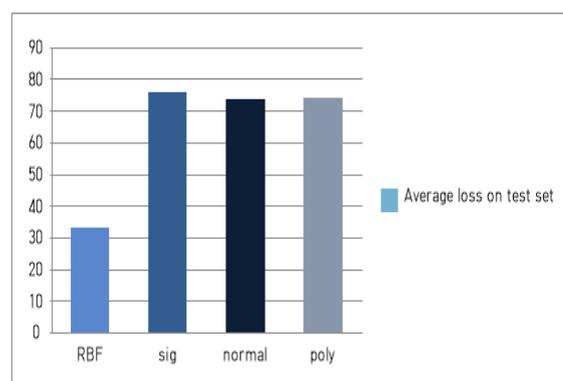


Figure 3. Average of accuracy for 17 window.

## 4 Analysis and discussion

Despite the risks of Dengue virus infections, there are few attempts to develop either the vaccine or the remedy (treatment) for Dengue virus. The most suitable treatment for the Dengue should satisfy the following qualifications. 1) The antibodies for the 4 different serotypes of Dengue should last for a long term. 2) It should stimulate the humoral immunity system and the cell-mediated immunity system fully.

We classified Dengue's 4 serotypes (DENV1, DENV2, DENV3, DENV4) with SVM algorithm using RBF, sig, normal and poly functions. Each result is organized by the number of the windows and the average values were compared. We could notice that the accuracy of sigmoid and polynomial functions were over 70%, and the accuracy of RBF function was relatively low; less than 40%. Therefore, it was obvious that DENV serotypes can be categorized both linearly and nonlinearly. For the normal function, DENV viruses showed accuracy of more than 70%. (However, it is comparatively lower than other viruses', which have the accuracy of over 90%).

As Dengue viruses are divided into 4 different groups, the Dengue vaccine should have the capability of preventing all types of the Dengue [10]. In addition, the antibodies for each serotype should be sustained homogeneously and should irritate both humoral immunity system and the cell-mediated immunity system. For these reasons, vaccine for Dengue is tough to develop.

If the vaccine cannot properly prevent the Dengue for the 4 serotypes, there are 2 things we can think about. 1) The made vaccine doesn't work on certain serotypes (out of the 4 serotypes) 2) The vaccine has side effects. It can cause an additional infection when there are antibodies for other serotypes. It can make the situation worse; it causes Dengue fever and Dengue shock.

Accordingly, Dengue virus can be assorted by their serotypes and have different properties, but they also have few common features, to a certain extent. Although DENV1, DENV2, DENV3, DENV4 show similar characteristics, according to the accuracy of each function, we can conclude that different vaccines are needed for the different serotypes.

It is possible for a Dengue vaccine to induce more serious diseases. Therefore, the vaccine should be carefully developed, and should satisfy all qualifications of the 4 different serotypes of the Dengue fever, which can prevent counterproductive effects of using the vaccine to the patients.

## 5 Conclusion

Dengue virus, which causes the Dengue fever, has four different serotypes; DENV1, DENV2, DENV3, DENV4. According to the classification of each serotypes using SVM, the different serotypes do have similar features. Nevertheless, the sigmoid and polynomial functions' vector accuracy values are over 70% which means that the serotypes have distinct characteristics. Thus, from these results, we could draw the fact that different

vaccines have to be used according to the different serotypes of the Dengue virus, even though they express similar symptoms of Dengue fever. Future vaccine developments have to be focused based on the distinct properties of each serotypes, not on the Dengue virus itself.

However, according to the results of this paper, it turned out that even though the Dengue serotypes show 70% differences, they also have similar properties. The 4 serotypes of Dengue virus (DENV1, DENV2, DENV3, DENV4) have different characteristics and can be classified into distinct groups. With the results of this paper, future research on the common reactions of each antibody on the different serotypes when Dengue is expressed and on the methods of developing a vaccine that efficiently uses the common properties of the 4 serotypes of Dengue virus will present the development course of the vaccines for the Dengue virus.

## References

1. Sessions, October M., Andreas Wilm, Uma Sangumathi Kamaraj, Milly M. Choy, Angelia Chow, Yuwen Chong, Xin Mei Ong, Niranjana Nagarajan, Alex R. Cook, and Eng Eong Ooi. "Analysis of Dengue Virus Genetic Diversity during Human and Mosquito Infection Reveals Genetic Constraints." *PLoS Negl Trop Dis PLOS Neglected Tropical Diseases* 9.9 (2015): n. pag. Web.
2. Mondini, A., A. Terzian, A. Ferraz, F. Oliveira, F. Chiaravalloti Neto, and M. Nogueira. "Surveillance of DENV in a City from São Paulo from 2006 to 2011: The Emergence of DENV-3 and DENV-4 and the Reemergence of DENV-2 and DENV-1." *International Journal of Infectious Diseases* 16 (2012): n. pag. Web.
3. Kim, Hyunseong, Juyoung Yoo, and Taeseon Yoon. "An Analysis of the Genomes of Dengue Virus Using Decision Tree and Apriori Algorithm." *International Journal of Computer and Communication Engineering IJCCE* 5.4 (2015): 294-301. Web.
4. Thu, Hlaing Myat, Kym Lowry, Thein Thein Myint, Than Nu Shwe, Aye Maung Han, Kyu Kyu Khin, Kyaw Zin Thant, Soe Thein, and John Aaskov. "Myanmar Dengue Outbreak Associated with Displacement of Serotypes 2, 3, and 4 by Dengue 1." *Emerg. Infect. Dis. Emerging Infectious Diseases* 10.4 (2004): 593-97. Web.
5. Melo, Fernando Lucas De, Camila Malta Romano, and Paolo Marinho De Andrade Zanutto. "Introduction of Dengue Virus 4 (DENV-4) Genotype I into Brazil from Asia?" *PLoS Negl Trop Dis PLOS Neglected Tropical Diseases* 3.4 (2009): n. pag. Web.
6. Vaughn, David W., Sharone Green, Siripen Kalayanarooj, Bruce L. Innis, Suchitra Nimmannitya, Saroj Suntayakorn, Timothy P. Endy, Boonyos Raengsakulrach, Alan L. Rothman, Francis A. Ennis, and Ananda Nisalak. "Dengue Viremia Titer, Antibody Response Pattern, and Virus Serotype

- Correlate with Disease Severity." *The Journal of Infectious Diseases* 181.1 (2000): 2-9. Web.
7. "Support Vector Machine (SVM)." SpringerReference (n.d.): n. pag. Web.
  8. Law, Ka-Man. "Vaccine Development against the Severe Acute Respiratory Syndrome-coronavirus (SARS-CoV) Using SARS-CoV Spike Protein." (n.d.): n. pag. Web.
  9. Zeng, Jianwu, and Wei Qiao. "Short-term Solar Power Prediction Using a Support Vector Machine." *Renewable Energy* 52 (2013): 118-27. Web.
  10. Gubler, Duane. "Dengue/Dengue Hemorrhagic Fever: The Emergence of a Global Health Problem." *Emerg. Infect. Dis. Emerging Infectious Diseases* 1.2 (1995):55-57.Web.