

# Comparison of H5N1, H5N8, and H3N2 Using Decision Tree and Apriori Algorithm

S. P. Jang, K. H. Park, Y. L. Kim, H. N. Cho, T. S. Yoon

Natural Science, Hankuk Academy of Foreign Studies, HAFS, Yongin-si, Republic of Korea  
Email: [heresphill@gmail.com](mailto:heresphill@gmail.com)

Received May 2015

---

## Abstract

H3N2, H5N1 and H5N8 virus were wide-spread epidemic in South Korea. Especially in 2014 Korea, the serious outbreak of avian influenza caused by H5N8 took place, effecting not only birds but also dogs. Antibody of H5N8 virus was found on a dog which differentiated the virus from existing H3N2 canine virus. At this point, we wanted to find out why H5N8 was self-medicated in dogs and whether H5N8 would cross species boundaries and be fatal to dogs or other species. While H5N1 is avian influenza like H5N8, many cases of fatal infections among dogs caused by H5N1 have been reported. Another kind of avian influenza, H3N2 is most common type of canine influenza in Asia. With the use of decision tree and apriori algorithm, we could find out characteristics of H5N8 by comparing it with H5N1 and H3N2.

## Keywords

SARS, Neural Network, CoV Mpro, Distorted Key Theory, 8-Mers, 12-Mers

---

## 1. Introduction

From 2003, South Korea is full of anxiety about AI (Avian Influenza), especially those caused by H5N8 virus. A great number of poultry ducks and chickens were buried and it also brought huge economical damage to farmers. However, H5N8 virus is being bandied again about the transmission to mammals in 2014, Cheonan, Chungnam, South Korea. Generally, AI is believed to be infectious only on birds, not mammals. But throughout the whole world, there were some cases that proved dogs are also unsafe from AI. However, it was an unprecedented situation in Korea. A dog, which is being raised in one of those farms where poultry ducks and chickens were culled, was found with an antibody of H5N8 in its nose. So to speak, there were no particular symptoms of avian influenza found on dog, but the evidence of self-medication was left.

Similar to H5N8, H5N1 is an avian influenza. While creation of H5N8 antibodies in dogs was reported recently, there are many cases of fatal infection of H5N2 in dogs and mankind. Also, even though H5N8 infection to dogs in Korea was reported to be self-medicated, since there aren't many similar cases, the result is unreliable. So we wanted to find out the reason why H5N8 was self-medicated in dogs or will it be really safe for dogs.

Also, H3N2 has been major cause of canine flu in Asia, especially in South Korea. Most of the dogs contracted to influenza in South Korea were reported to be infected to H3N2. It seems that H3N2 is becoming

stronger and spreading out to wider range, and it has caused a highly contagious flu outbreak in Chicago recently. By comparing H5N8 with H3N2, we expected to have a deeper understanding about H5N8: whether it would bring deadly outbreak in other parts of Asia and how it is different from the existing canine flu, H3N2.

### 1.1. H5N1

Highly Pathogenic Avian Influenza A (HPAI H5N1) is a subtype of influenza A virus that causes deadly infections among birds in Asia, Europe and Africa. It was initially known to infect only birds, but since the first human infection was reported in 1997, 649 cases have been reported till January, 2014, having a mortality rate of about 60%. While most human infections are caused by direct contact with infected poultry, there's also possibility of transmission between humans. Consequently, there's a high concern that the virus might mutate to become easily transmissible among mankind, resulting in severe consequences.

### 1.2. H5N8

H5N8 is a subtype of H5N1 virus which brings a huge sensation around the world. The most well-known case of H5N8 infection might be an outbreak in Ireland in 1983. About 8000 turkeys, 28,020 chickens, and 270,000 ducks were culled. It was initially known to infect only birds but in Thailand, 2004, a dog with an antibody of H5N8 was found and so did in South Korea, 2014. It was an unprecedented event of H5N8 to be transmitted from birds to mammals in Korea. Still, it is a mystery whether the virus would affect on human-beings or not.

### 1.3. H3N2

H3N2 canine virus is subtype of influenza A which originated in Jiangsu province of China, 2010. Dog flu by H3N2 is highly infectious respiratory disease accompanying severe fever, cough, and decreased appetite. H3N2 virus has been a main cause of dog flu in Asian countries including South Korea, China, and Thailand, but it is spreading to further western countries. Previously, canine influenza in US mostly originated from H3N8, but starting from January 2015, H3N2 canine influenza has been isolated from over 1700 infected dogs in western parts of US-Illinois, Ohio, Indiana, and Wisconsin. Recent dog flu outbreak in Chicago also seems to be caused by H3N2 canine virus and according to CDC it is highly possible to be originated from Asia.

## 2. Method and Experiment

### 2.1. Apriori Algorithm

Apriori algorithm is a usual algorithm that shows the frequency and general rule of the given datasets (Ji Hea Lee *et al.* 2014) [1]. This algorithm works in two steps: in a first step is determining main frequent item. These are flocks that have at least the given minimum support. In the second step is to extract general rule that can show feature of datasets (Borgelt Christian & Rudolf Kruse 2002) [2]. The algorithm uses a "bottom up" approach where frequent subsets are prolonged once, which is called candidate generation. Each set of candidates is tested in many times until no more expanded rules are found. To make it able to use the breadth-first search and a Hash tree structure, this algorithm counts candidate items efficiently by proceeding a few steps. First, it makes candidate dataset of length  $k$  from dataset of length  $k - 1$  that make candidates enclose an infrequent rule. Subsequently, according to the downward closure lemma, the candidate will enclose all frequent dataset of length  $k$ . In the final step, it scans the transaction database to determine frequent item sets among the candidates (Dae Young Kim *et al.* 2015) [3].

### 2.2. Decision Tree

Decision Tree is a common data mining method which is widely used for leading inference, so as to generate a model that can predict the value of a main variable according to several input variables. It poses a series of questions about the features associated with data items. Each questions contained in a node, and every internal node points to one child node for each possible answer to its question (Jae Jun Lee & Taeseon Yoon 2014) [4].

A tree is called "learned" by classifying the data set into subsets generated from branches down recursively. This kind of learning is defined as recursive partitioning. This process of a top-down leading of decision trees is a kind of "greedy algorithm", and is one of the most usual method for learning decision trees (Cheolho Heo &

Taeseon Yoon 2014) [5].

In data mining, decision trees can work as the mixture of mathematical and computational form, utilizing the progress of classification. Usually, the diagram gradually comes in records of the form:  $(X, Y) = (x_1, x_2, x_3, \dots, x_k, \mu)$  (Jooyeol Yun *et al.* 2014) [6].

The dependent variable, Y, is the target variable that we are trying to understand, classify or generalize. The vector  $x$  is composed of the input variables,  $x_1, x_2, x_3$  etc., which are used for testing (Eunby Go *et al.* 2014, Seung Jae Lim *et al.* 2015) [7] [8].

### 2.3. Experiment

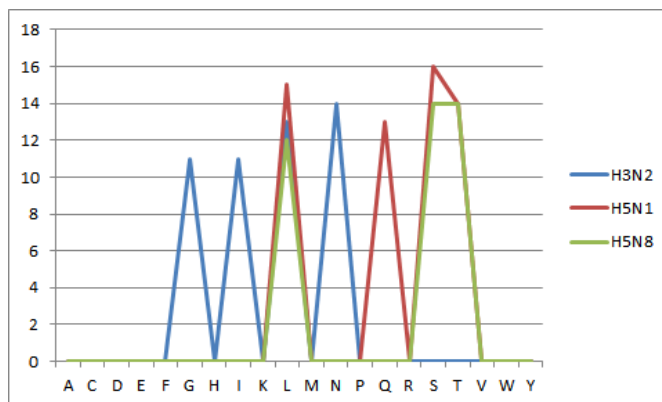
First of all, we gathered nucleotide sequence of H5N1, H5N8, H3N2 viruses from NCBI, which is called “National Center for Biotechnology Information”. We performed experiment through apriori algorithm and decision tree algorithm. For apriori algorithm, we conducted experiment in 5 window, 7 window, and 9 window for each viruses. For decision tree algorithm, we performed experiment for each viruses in class 1 to HA, class 2 to NA.

## 3. Results

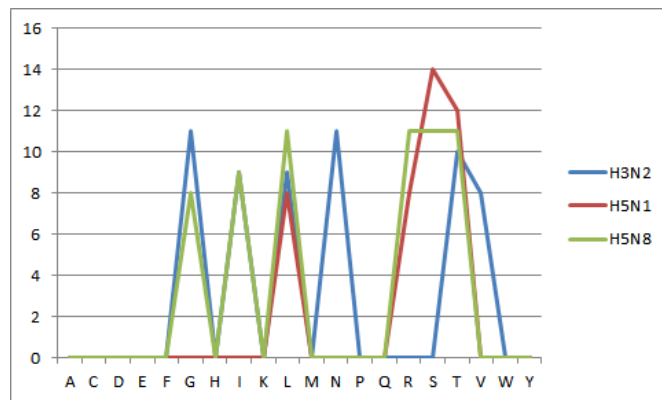
We could find strong rules in apriori algorithm, but there was not clear rule in experiment in decision tree algorithm.

### 3.1. Apriori Algorithm

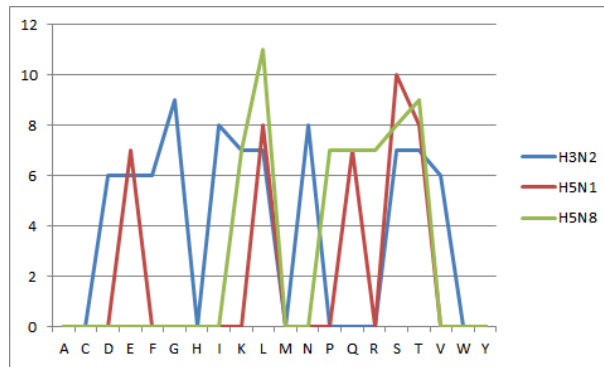
As we can see in **Figures 1-3**, there is clear similarity among three viruses. Also, we noticed leucine is the key factor of hemagglutinin. Also, in **Figures 4-6**, we found quite similar rules among three viruses. Moreover, we found that isoleucine is the key factor of three viruses’ neuraminidase.



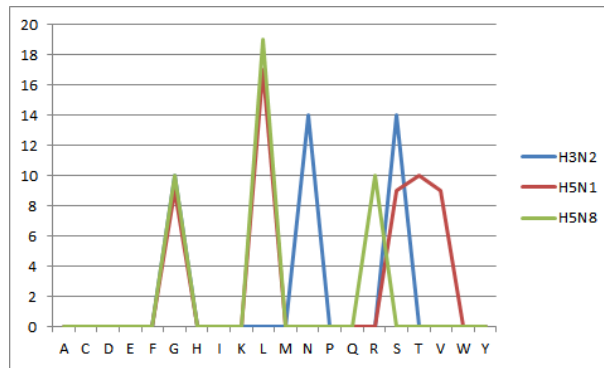
**Figure 1.** Haemagglutinin in 5 window.



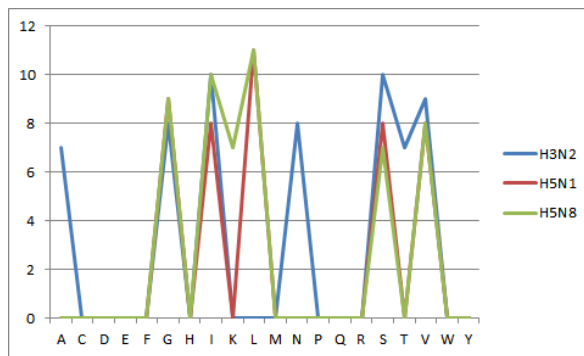
**Figure 2.** Haemagglutinin in 7 window.



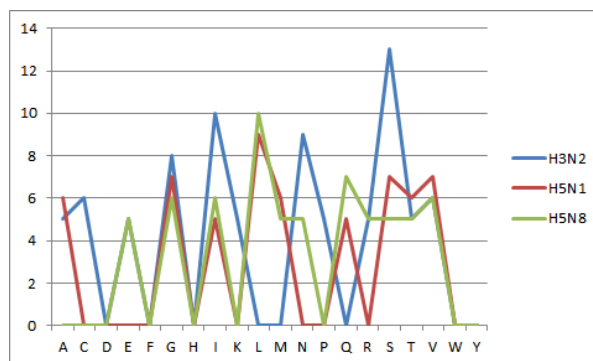
**Figure 3.** Haemagglutinin in 9 window.



**Figure 4.** Neuraminidase in 5 window.



**Figure 5.** Neuraminidase in 7 window.



**Figure 6.** Neuraminidase in 9 window.

### 3.2. Decision Tree

Not as we thought, no key rule was found in the performance involving the decision tree. The presence of different amino acids among all sites did not have differences that are huge enough to consider as key factor. In decision tree experiment, most of dataset shows to be converged to heamagglutinin feature.

### 4. Conclusion

With apriori algorithm, we noticed that three viruses seem high similarity. Particularly, apriori algorithm showed that amino acid leucine is the key factor of performing heamagglutinin, and isoleucine is the key factor of performing Neuraminidase. As a result, the experiment showed that three of them can infect same animal such as dog. However, we couldn't find strong rule through decision tree algorithm. The reason might be high similarity of heamagglutinin and Neuraminidase of each virus. In conclusion, all viruses can infect dog. As we find difference between H5N8, and H3N2, asparagine may be the cause of impossible of making anti-body of the flu virus.

### References

- [1] Leea, J.H., Ahna, S.H., Pyuna, S.M., Janga, E.J. and Yoon, T. (2014) Analysis of Malaria Inducing *P. falciparum*, *P. ovale*, and *P. vivax* through Apriori Algorithm and Decision Trees.
- [2] Borgelt, C. and Kruse, R. (2002) Induction of Association Rules: Apriori Implementation. Compstat. Physica-Verlag HD.
- [3] Kim, D.Y., Kim, H.-J., Bae, J. and Yoon, T. (2015) Examining the Probability of the Critical Mutation of H5N8 by Comparing with H7N9 and H5N1 Using Apriori Algorithm and Support Vector Machine. *International Journal of Computer Theory and Engineering*, **7**.
- [4] Lee, J.J. and Yoon, T. (2014) The New Approach on Fuzzy Decision Trees. *International Journal of Fuzzy Logic Systems (IJFLS)*, **4**.
- [5] Heo, C. and Yoon, T. (2014) Deeper Understanding about Attributes of HIV Employing Support Vector Machine. *International Journal of Bioscience, Biochemistry and Bioinformatics*, **4**, 336-339. <http://dx.doi.org/10.7763/IJBBB.2014.V4.366>
- [6] Yun, J., Seo, J.W. and Yoon, T. (2014) The New Approach on Fuzzzy Decision Trees. *International Journal of Fuzzy Logic Systems (IJFLS)*, **4**.
- [7] Go, E., Lee, S. and Yoon, T. (2014) Analysis of Ebolavirus with Decision Tree and Apriori Algorithm. *International Journal of Machine Learning and Computing*, **4**.
- [8] Lim, S.J., Heo, C., Hwang, Y. and Yoon, T. (2015) Analyzing Patterns of Various Avian Influenza Virus by Decision Tree. *International Journal of Computer Theory and Engineering*, **7**.