

Treatment of various avian influenza virus based on comparison using decision tree algorithm

Jongchan Lee¹, Seojun Kim² and Taeseon Yoon³

¹Natural Science Course, Hankuk Academy of Foreign Studies, Yongin, Republic of Korea

²Natural Science Course, Hankuk Academy of Foreign Studies, Yongin, Republic of Korea

³Computer Science and Engineering, Korea University, Seoul, Republic of Korea

Abstract. Recently, the world is full of anxiety about AI(Avian Influenza). Among Avian Influenza virus subtypes, H5N1 is considered the most threatening to not only birds, but also humans as numerous human cases with high mortality have been reported. Unlike H7N9, which has not been reported infect humans, H5N8 also became infectious to humans due to dramatic mutation. As human infection cases of AI have increased, numerous researchers have been trying to develop an effective treatment against them. Thus, our project group decided to analyze the similarity and difference of H5N1, H5N8, and H7N9, since it would be useful for finding effective treatment of AI, using Decision Tree Algorithm which figures out distinctive factors of given dataset for comparing protein sequences of each viruses. The comparison using Decision Tree Algorithm, which indicates correlation among H5N1, H5N8, and H7N9, will be effective on narrowing the range of attempts on developing treatment for Avian Influenza virus.

1 Introduction

Recently, the world is full of anxiety about AI(Avian Influenza), as many Avian Influenza viruses subtypes have killed not only a number of birds but also other species and even humans. Among them, the highly pathogenic influenza A virus subtype H5N1 is considered the most threatening to both birds and humans.

H5N1 has killed millions of poultry in numerous countries throughout Asia, Europe, and Africa, causing enormous damage. Generally, AI is believed to be infectious only to birds. However, since the first transmission of H5N1 to human was reported in Hong Kong in 1997, 844 cases have been reported until December, 2015 in Cambodia, Egypt, Indonesia, Viet Nam and so on, resulting a high mortality rate of about 50% [1, 2].

Similar to H5N1, H5N8 is an avian influenza virus. As H5N8 devastated South Korea in 2014, about 200 million 3 thousand poultry were killed [3]. It has not been reported to infect humans unlike H7N9, a kind of AI, which became infectious to humans starting from 2013 due to dramatic mutation [4]. However, it seems that H5N8 is fairly capable of infecting humans in that H7N9 went through a dramatic mutation in a relatively short period.

As human infection cases of AI, which has a considerably high mortality rate, have increased, numerous researchers have been trying to develop an effective treatment against H5N1, H7N9. Actually, there

are various suggested vaccines that mainly target the glycoprotein of influenza [5].

In this research, we analyzed the similarity and difference among the respective Protein HA, NA, NP, M of H5N1, H5N8, and H7N9 using Decision Tree Algorithm, since it would be useful for finding effective treatment of AI.

2 Materials and methods

2.1 H5N1

H5N1 virus is a subtype of Influenza A virus which triggers highly pathogenic avian influenza (HPAI). One of the most disastrous infections of H5N1 occurred mid-December 2003 to January 2004 [6]. Avian influenza H5N1 has been found in wide areas; its human susceptibility has been reported in South Korea, Japan, Vietnam, Thailand, Cambodia, Indonesia, Pakistan, Laos, and China [7], threatening human health and resulting a plunge in agricultural production [8]. Direct contact with pathogenic materials such as excreta or secretion of poultry seems to cause infection of H5N1 to human. H5N1 has been reported to cause human respiratory problems, fever, pneumonia, and relatively severe cytokine releases [9].

2.2 H5N8

H5N8 virus is a subtype of Influenza A virus. It was first reported from turkey in 1983 Ireland. In January 2014 South Korea, avian influenza that occurred among poultry and caused enormous damage was confirmed as a H5N8 virus [10]. H5N8 viruses cause fever, respiratory problems, and conjunctivitis among poultry.

2.3 H7N9

H7N9 virus is a subtype of Influenza A virus, but was newly revealed as a human infectious virus since the outbreak from China in 2013 [11]. Several H7 viruses such as H7N2, H7N3, and H7N7 had been previously verified as human-infectious. Its infection pathways also include direct and indirect exposure with pathogenic materials. Generally, H7N9 brings deadly pneumonia and fever on human while it causes relatively slight illness among poultry. As a result, it requires researches to distinguish infected poultry from others [12].

2.4 Glycoprotein of Influenza A virus; Hemagglutinin (HA) and Neuraminidase (NA)

Hemagglutinin (HA) and neuraminidase (NA) are among the essential proteins for invasion of virus into host cells; there are 11 type of Influenza A viral proteins – PB1, PB1-F2, PB2, PA, HA, NA, M1, M2, NS1, NEP, and NP. Hemagglutinin is a surface glycoprotein that enables virus to initiate chemical reactions to get into plasmic membranes of infected cell. It combines with monosaccharide sialic acid, accelerating attach on the surface of cell [13]. Neuraminidase is also viral surface glycoprotein that is, on the other hand, responsible for the escape of virus from host cell. It decomposes sialic acid from host cell, facilitating transmission to other cells [14]. The type of hemagglutinin and neuraminidase from virus has been considered as distinguishable elements for classification in influenza viruses. As a result, the name of Influenza A virus reflects the type of HA and NA, resulting H1N1, H5N8, H7N9, etc.

2.5 Decision Tree Algorithm

Decision Tree is a common tool of data mining using tree diagrams. It contains 'nodes' that require decisions for each questions about the relative degree of influence on specific results. Decision Tree Algorithm(ID3 algorithm; Iterative Dichotomiser 3 algorithm) utilizes Decision Tree in order to deal with relevance between given data($x_1, x_2, x_3, \dots, x_k$) with major component(Y). Decision Tree algorithm indicates the relationship as

$$(X, Y) = (x_1, x_2, x_3, \dots, x_k, Y)$$

Decision Tree Algorithm calculates entropy(H) of dataset.

$$H(S) = \sum(-P_i \log(P_i)) \quad (P: \text{given dataset})$$

It extracts rules that show certain dataset which has lowest H; in other words, which has highest relevance with result(Y) sets. This process originates from decision trees judging which data to accept; based on entropy(H) measure [15].

In our research, Decision Tree Algorithm was used for showing interrelation of H5N1, H7N9, and H5N8 viruses in molecular level; amino acidic sequences of each viruses. From extracted rules that indicates the degree of differences among viruses, it is possible to infer distinguishable characteristics of each objects [16]

2.6 Dataset

For the experiments on Influenza A virus; H5N1, H5N8, and H7N9, nucleotide sequences of viral proteins; HA, NA, NP, M derived from NCBI data were used in our research. KF182741.1(HA), KF182743.1(NA), KF182742.1 (NP), KF182744.1(M) are viral protein sequences of H5N1 from Vietnam, KJ476657.1(HA), KJ476659.1(NA), KJ476658.1(NP), KJ476660.1(M) are viral protein sequences of H5N8 from Shanghai, KJ549786.1(HA), KJ549788.1(NA), KJ549789.1(NP), KJ549789.1(M) are viral protein sequences of H7N9 from Shanghai. These sequences of viral protein HA, NA, NP, M were used to compare the corresponding protein sequences of each viruses.

3 Result

We basically extracted data with frequency rates 0.750 or higher in HA. However, we extracted data with frequency rates 0.800 or higher in H7N9 HA under 5 window, 9 window due to observation of relatively excessive rules than other viruses. Also, we extracted data with frequency rates higher than 0.750 in NA.

Table 1. HA Rule Extraction Under 5 Window

virus	Rules	Frequency
H5N1	pos2 = R pos3 = S	0.750
	pos3 = R pos5 = Q	0.750
H5N8	pos3 = L pos4 = R	0.750
	pos3 = L pos4 = T	0.750
H7N9	pos3 = D pos4 = L	0.800
	pos3 = T pos4 = E	0.800
	pos3 = M	0.800
	pos1 = L pos3 = F	0.800
	pos3 = S pos5 = G	0.833
	pos2 = G pos3 = N	0.800

According to Table 1, It is noticeable that all viruses show their rules extracted. However, rules extracted from H7N9 were distinctively numerous than rules from H5N1 and H5N8. This result is due to relative lack of unique amino acidic features of H5N1 and H5N8 compared to H7N9. It also supports that H7N9 has different amino acid features of HA compared with H5N1, H5N8, and that H5N1 and H5N7 have similarities in some amino acidic

features in HA. Also, it can be assumed that position 3 is the important factors that differentiate each other since position 3 is observed most frequently. Position 1 and 4 also seem to be important factors since they are also observed frequently.

Table 2. HA Rule Extraction Under 7 Window

virus	Rules	Frequency
H5N1	not extracted	
H5N8	pos5 = P	0.750
H7N9	pos5 = H	0.800
	pos1 = D pos 5 = M	0.750
	pos5 = R	0.750

Table 3. HA Rule Extraction Under 9 Window

virus	Rules	Frequency
H5N1	not extracted	
H5N8	pos4 = G pos6 = F	0.750
H7N9	pos9 = W	0.800
	pos9 = F	0.800
	pos4 = F	0.800
	pos4 = R	0.909
	pos3 = Q	0.833

According to Table 2, Table 3, similar to 5 window, rules extracted from H7N9 were distinctively numerous than rules from the H5N1, H5N8. Furthermore, there were no extracted rules of H5N1. This result also supports that H5N1, H5N8 have similar amino acidic features in HA, and H7N9 has distinctive amino acidic features compared to others. In Table 2, it can be assumed that position 5 is the important factors which differentiate each other. In Table 3, position 4 and position 9 are the important factors which differentiate each other. Due to altered window condition, although positions which are considered as important factors changed, the fact that H7N9 has excessive rules compared with H5N1 and H5N8 did not change.

Table 4. NA Rule Extraction Under 5 Window

virus	Rules	Frequency
H5N1	pos3 = P pos4 = D	0.800
	pos3 = T pos4 = D	0.800
H5N8	pos3 = H	0.800
H7N9	pos3 = G pos4 = A	0.800
	pos1 = A pos4 = I	0.800

According to Table 4, the results under 5 window, it is noticeable that rules of amino acids at position 3 and position 4 were shared between H5N1, H5N8, and H7N9. Compared to Table 2 and Table 3, NA sequences generally showed equal frequency of extracted rules. It showed identical positions but different corresponding amino

acids. We assume that these viruses showed different amino acidic features in NA with each other probably due to the difference in types of viral neuraminidase.

Table 5. NA Rule Extraction Under 7 Window

virus	Rules	Frequency
H5N1	pos2 = P pos5 = S	0.833
	pos1 = G pos3 = V	0.800
H5N8	pos1 = T pos4 = V	0.800
	pos2 = R pos7 = T	0.800
H7N9	pos3 = N	0.857

According to Table 5, the results under 7 window, rules of amino acids at position 1, position 2 and position 3 were shown as important factors that differentiates viruses each other. This result that shows existence of relatively equal number of unique rules also supports that H5N1, H5N8, and H7N9 have distinguishable amino acidic features from each other in NA.

Table 6. NA Rule Extraction Under 9 Window

virus	Rules	Frequency
H5N1	pos5 = S pos6 = G	0.800
	pos3 = D pos7 = I	0.800
H5N8	pos6 = V	0.786
H7N9	pos6 = D	0.833
	pos6 = S pos8 = S	0.800
	pos7 = V	0.800

According to Table 6, the results under window 9, we could assume that position 6 works as the important factors for differentiating three viruses because it is observed it is observed most frequently.

Not as our thought, no rule was found in NP and M of H5N1, H5N8, and H7N9. This result is due to the absence of distinguishable amino acidic differences that are huge enough to be considered as key factor. Thus, we could assume that these viruses may indicate similar amino acidic features in NP and M proteins.

4 Conclusion

With the Decision Tree algorithm, we could see the similarities and differences between three typical avian influenza viruses; H5N1, H5N8, and H7N9. Particularly, Decision Tree Algorithm showed that NP and M of H5N1, H5N8, H7N9 have significantly similar amino acidic features. However, NA of three viruses have considerable differences amongst themselves. Also, in HA, H7N9 has distinctively different amino acidic features compared with H5N1 and H5N8, which have similar amino acidic features.

As a result, we could conclude that H5N1 virus treatments targeting NP and M would be effective to H5N8 and H7N9, which also means we can use common treatments to cure all of H5N1, H5N8, and H7N9. Also, we made a conclusion that H5N1 virus treatments targeting HA will be effective to H5N8, not to H7N9,

which indicates we can use common treatments to attack both H5N1 and H5N8.

The outbreaks of Influenza A virus have been causing fatal diseases on both human and poultry. Researchers generally search for vaccines and its infection pathways to prevent its spread when new type of Influenza A virus occurs. Thus, if they lacked data of such virus, the treatment would require numerous attempts on finding appropriate viral vaccine. Hopefully, from our research on Influenza A virus (H5N1, H5N8, and H7N9), we investigated the similarity and difference between viruses using data mining Decision Tree Algorithm. As we mentioned above, it suggests the possibility of effective methods for developing treatments on each viruses in terms of time and economy.

Based on its similarity among viruses, we can narrow the range of attempts on developing treatments. Even when a new type of Influenza A virus emerges, it is possible to infer its overall characteristics with indirect bioinformatic research on correlation between known viruses. We hope that our research on Influenza A virus will be widely used for epidemiologic research of various viruses and contribute to finding effective treatment of them.

References

1. Peiris, J. S. M, M. D. De Jong, and Y. Guan. "Avian Influenza Virus (H5N1): A Threat to Human Health." *Clinical Microbiology Reviews* 20.2 (2007): 243-67.
2. "Cumulative Number of Confirmed Human Cases of Avian Influenza A(H5N1) Reported to WHO." World Health Organization. N.p., n.d. Web. 17 Jan. 2016.
3. Herriman, Robert. "South Korea H5N8 Avian Flu Update, Virus Spreads to Chicken Farm." *The Global Dispatch RSS*. 26 Jan. 2014.
4. Young, Kim Dae, Kim Hye-Jun, Bae Junhyeok, and Yoon Taeseon. "Examining the Probability of the Critical Mutation of H5N8 by Comparing with H7N9 and H5N1 Using Apriori Algorithm and Support Vector Machine." *IJCTE International Journal of Computer Theory and Engineering* 7.2 (2015): 145-48.
5. Jang, S. P., K. H. Park, Y. L. Kim, H. N. Cho, and T. S. Yoon. "Comparison of H5N1, H5N8, and H3N2 Using Decision Tree and Apriori Algorithm." *Journal of Biosciences and Medicines JBM* 03.06 (2015): 49-53.
6. "Bird Flu: Global Response Sought," CNN.com, Jan. 27, 2004
7. David P. Fidler, "Global Outbreak of Avian Influenza A (H5N1) and International Law," *American Society of International Law*, January 25, 2004
8. WHO, Avian Influenza A (H5N1) Infection in Humans: Urgent Need to Eliminate the Animal Reservoir-Update 5, Jan. 22, 2004
9. Chan MC, Cheung CY, Chui WH, Tsao SW, Nicholls JM, Chan YO, Chan RW, Long HT, Poon LL, Guan Y, Peiris JS. (2005). "Proinflammatory cytokine responses induced by influenza A (H5N1) viruses in primary human alveolar and bronchial epithelial cells". *Respir. Res.* 6 (1): 135
10. Kang HM, Lee EK, Song BM, Jeong J, Choi JG, et al, "Novel reassortant influenza A(H5N8) viruses among domestic and wild ducks, South Korea, 2014," *Emerg Infect Dis.* 2015 Feb
11. "Frequently Asked Questions on human infection with A(H7N9) avian influenza virus, China," WHO, 2013
12. "Avian Influenza A (H7N9) Virus," Centers for Disease Control and Prevention, 2013
13. White JM, Hoffman LR, Arevalo JH, et al. (1997). "Attachment and entry of influenza virus into host cells. Pivotal roles of hemagglutinin". In Chiu W, Burnett RM, Garcea RL. *Structural Biology of Viruses. Oxford University Press.* pp. 80–104.
14. von Itzstein M (December 2007). "The war against influenza: discovery and development of sialidase inhibitors". *Nature Reviews. Drug Discovery* 6 (12): 967–74
15. Go, Eunby, Seungmin Lee, and Taeseon Yoon. "Analysis of Ebolavirus with Decision Tree and Apriori Algorithm." *International Journal of Machine Learning and Computing IJMLC* 4.6 (2014): 543-46.
16. Jang, S. P., K. H. Park, Y. L. Kim, H. N. Cho, and T. S. Yoon. "Comparison of H5N1, H5N8, and H3N2 Using Decision Tree and Apriori Algorithm." *Journal of Biosciences and Medicines JBM* 03.06 (2015): 49-53.