Journal of Bioinformatics and Neuroscience (JBINS), 7 (1): 308-314, April 9, 2021
e-ISSN: 2432-5422, p-ISSN: 2188-8116
Applied Science and Computer Science Publications

Comparing Two Feature Selection Methods for Influenza-A Antivral Resistance Determination.

Nermin Shaltout, Ahmed Rafea, Mohamed Moustafa, Ahmed Moustafa, Mahmoud ElHefnawi, Mohamed ElHefnawi

Abstract: The paper thoroughly analyzes the use of Principal Component Analysis (PCA) in comparison to Information Gain (IG) as a feature selection method for improving the classification of Influenza-A antiviral resistance. Neural networks were used as the classification method of choice with PCA, while decision trees were the classification of choice with IG. The experiment was conducted on cDNA viral segments of Influenza-A belonging to the H1N1 strain. The 7 Infleunza-A segments generating the best results were used for comparison. Sequences from each segment were further divided into Adamantane-resistant, & non-Adamantane-resistant. Accuracy, sensitivity, specificity precision & time were used as performance measures. Using PCA for feature selection increased preprocessing speeds from an average processing time of 1.5 hours to 5 minutes, as opposed to IG. IG had higher accuracy. The best accuracy generated by PCA & NNs on the M1/M2 was 96.5%, while that of IG & DTs was 98.2% Using PCA features & DTs also generated a comparable accuracy to that of IG features & DT at 97.6% on the M1/M2 segment. There was a 88% increase in feature selection processing speed when using PCA compared to IG on the M1/M2 segment alone

Keywords: Influenza-A, Principle component analysis (PCA), Machine learning, Information gain, DNA classification, decision trees, neural networks.

I. INTRODUCTION

The Influenza-A virus is known for its high mutation rate. Due to the resulting high morbidity & mortality rate, antivirals are sparsely used as the virus may increase its resistance to the available antiviral brands via mutation. e.g. In *2009* Adamantane eliminated the virus with a *50*% success rate compared to Osetamivir which had a *90*% success rate.

In addition to infecting human hosts, some Influenza-A viral strains can infect multiple hosts simultaneously causing dangerous outbreaks that can propagate rapidly across the population. Without proper knowledge of the drugs that can control the outbreaks, the situation could lead to high mortality & morbidity rates. The research's objective is examining efficient ways for determining the antiviral resistance of viral strains across several viral segments. By measuring the consistency of the results across all the viral segments, the appropriate drug for virus elimination can be correctly predicted with a higher confidence during pandemics. The segment generating the most accurate results can also be determined.

A. Summary of Influenza-A's Structure

Influenza-A's genetic data is divided into eight RNA segments [1], that can be interchanged between different virus strains to produce new viruses.

Corresponding Authors: Nermin Shaltout, Mohamed Moustafa, Ahmed Rafea, Ahmed Moustafa. Department of Computer Science; Biology Department. The American University of Cairo, New Cairo, Cairo 11835, Egypt. E-mail: nermeena@gmail, m.moustafa@aucegypt.edu, rafea@aucegypt.edu, amoustafa@aucegypt.edu.
Mahmoud ElHefnawi. Informatics & System Department and Biomedical Informatics & Chemoinformatics group, Division of Engineering Research and Centre of Excellence for Advanced Sciences. National Research Center Tahreer Street, Cairo 12311, Egypt. Email: mahef111@gmail.com

Hemagglutinnin (HA) & Neuraminidase (NA) are the most important segments since they code for the virus's surface antigens.

The target host's antibodies identify & eliminate the invading virus via these surface antigens. Outbreaks result when virus's mutation leads to the creation of HA & NA proteins that are unrecognized by the host's antibodies. Influenza-A's rapid mutation rate has thus led to the emergence of at least 16 HA and 9 NA known viral subtypes [2]. The virus's strain is determined by the different combination of HA (H) & NA (N) subtypes. E.g. H1N1 is the strain responsible for the swine flu pandemic post mutation. Fig. 1 summarizes the most important features of the Influenza-A virus.

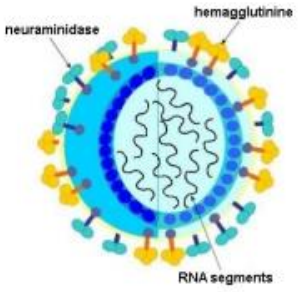

Fig. 1 Summary of Influenza-A's main features

Influenza-A's severity during outbreaks is determined by its virulence or mortality rate [3]. Furthermore Influenza-A's virulence is increased by its capacity to infect various hosts [3]. Based on previous pandemic & epidemic history, human, avian & swine hosts were the most problematic. Certain strains pertaining to the aforementioned hosts have gained the capability of infecting multiple hosts simultaneously via mutation as well as the ability to propagate from one host to another [3]. To facilitate the virus's analysis, decoded Influenza-A sequences are stored in online databases in the form of RNA/cDNA & Protein sequences. These sequences can aid in predicting the virus's antiviral resistance. Since the H1N1 strain caused a severe outbreak in 2009, it will be used for antiviral resistance prediction in this research's scope. The most important viral hosts will also be used to investigate the problem.

B. **Motivation**

Previous studies by Shaltout [4] revealed that online records of the H1N1 Infleunza-A virus show a higher resistance to Adamantane as an antiviral in comparison to Oseltamivir. Following that logic, Oseltamivir would be the more appropriate choice of antiviral during a pandemics caused by the H1N1 strain. However, during drug shortages, determining if the viral strain is susceptible to Adamantane can increase the drug pool available for virus elimination.

Feature selection should also be more thoroughly analyzed due to Influenza-A's cDNA sequences reaching around 2000-3000 units per cDNA sequence. Since there are thousands of Influenza-A viral sequences available on the database, finding a time efficient data reduction technique that preserves accuracy is necessary. Feature extraction on cDNA sequences may also lead to more fine-grained classification, that detects minor changes that are otherwise undetectable in protein sequences.

II. LITERATURE REVIEW

In Bioinformatics, prior to classification, feature selection is used for reducing or compressing high dimensionality datasets to enhance machine learning efficiency. The following researches were conducted in feature selection: Saeys et al. [5] surveyed the pros & cons of three feature selection methods in the scope of Bioinformatics: filter, embedded & wrapper methods. They pointed out that univariate filter methods are favored when reducing high dimension data due to their speed, scalability & classifier independence. Leung et al.'s [6] utilized a filter method known as Information Gain (IG) to pinpoint the RNA biomarkers in the Hepatitis-B virus responsible for hepatocellular carcinoma in the infected host. Rule Learning & Nonlinear Integrals were utilized for classification after the feature selection step. Shaltout & Shaltout et al. [4,7] further researched IG as a feature selection technique, by applying IG to aligned Influenza-A viruses in order to detect the most informative positions thus enhancing hosts classification. Although IG improved the classification results, features had an average length of 100 units. Shaltout [4] further tested the effects of reducing the features to 10 units, but it yielded varying results based on viral subtype: viruses pertaining to the H1 subtype yielded more accurate results. The results were conducted in more than one segment in both [4,7]. In Shaltout et. al principle component analysis (PCA) was utilized as a feature selection technique for antiviral resistance of H1N1 but only one viral segment was used. [8]

Encoding schemes are usually required in order for classifiers to utilize bioinformatics data which are usually nominal in nature. The following researches were conducted on encoding schemes: Attaluri [9] analyzed the effect of different neural network (ANN) encoding schemes on the accuracy of Influenza-A

classification. The experiment revealed that including gaps generated by multiple sequence alignment (MSA) increased classification accuracy. Attaluri additionally uncovered the k-frequencies at which indirect encoding schemes generated optimal results.

The following key researches on Influenza-A analysis using machine learning were conducted by Attaluri, ElHefnawi et al. & Shaltout et al. [4,7,8,9,10,11]. ElHefnawi et al. [10] classified Influenza-A hosts & subtypes using hidden Markov models (HMMs). Protein sequences from the HA segment pertaining to subtypes H1, H2, H3, H4 & H5 were used. To classify hosts, the virus sequences were further divided into human & nonhuman hosts. The research yielded an overall subtype classification accuracy of 100% & a host classification accuracy ranging from 50% to 100%, depending on subtype. The method didn't analyze RNA classification performance or the effect of applying feature selection prior to classification. In another experiment, ElHefnawi et al. [11] used HMMs & decision trees (DTs) on extracted host associated protein signatures to increase host identification accuracy. The experiments were conducted on the HA protein of various subtypes. DTs yielded higher host classification accuracies, ranging from 92%-100%, as opposed to HMMs. The research did not explore extracting host associated signatures at the RNA level.

Attaluri [9] analyzed the use of ANNs, DTs & support vector machines (SVMs) for Influenza-A hosts & subtypes identification. He conducted the experiments using cDNA & protein sequences. A subset of virus sequences belonging to the H1N1 strain was used for host classification. Sequences belonging to the H1, H2, H3, N1 & N2 subtypes were used for subtype classification. The overall classification accuracies for both subtype & host classification were 96.5%, 96.2% & 95.1% when using DTs, SVMs & ANNs respectively. Attaluri [9] additionally integrated DTs & HMMs in a singular model to classify Influenza-A hosts & subtypes. After using DTs to extract informative positions from the cDNA sequences he converted them into their protein equivalent. The obtained protein sequences were then fed as input to a HMM classifier. Both viral host & subtype classification were analyzed. The technique yielded an overall accuracy of 97%. Although the technique yielded accurate results, the capacity of feature selection to improve classification performance in [9] was not analyzed fully at the cDNA/RNA level. This is due to the conversion of the features to protien sequences prior to classification. Additionally means of measuring the efficiency of the classification process were not devised.

The previous problems were later addressed in Shaltout et. al [4,7]. They studied the effect of feature selection in terms of accuracy & efficiency on the cDNA classification of Influenza-A. Although the research showed promising results, the process of feature selection itself was lengthy & still yielded a relatively large amount of features. Shaltout [8] addressed this issue but demonstrated it only using one viral segment. In this research we seek to expand on the feature selection method developed by Shaltout et al. in [8] & measure the respective performance on Influenza-A classification compared to using Information Gain as a feature selection technique for more viral segments to determine the consistency of the results & the extent of accuracy of other viral segments in comparison to the M1/M2 segment.

III. RESEARCH AIM

The research seeks to analyze the improvement in the efficiency & performance of classifying Influenza-A antiviral resistance when applying Principal Component Analysis (PCA) as a feature selection technique as opposed to Information Gain (IG). This will be attempted on all the viral segments. The effect of feature selection on the overall classifier's performance will also be recorded. The system will be implemented by reducing the cDNA features of more than one viral segment using PCA, then feeding the reduced features to a classifier. This is done without protein conversion to improve RNA/cDNA analysis of the virus at a more fine grained level. Using PCA is expected to decrease the number of features extracted as opposed to IG without deteriorating accuracy significantly. This will be achieved as follows:

- Running the experiment on all the viral segment belonging to the H1N1 strain.
- Using the minimum amount of PCA features to extract the cDNA positions showing greatest variation.
- Feeding the extracted PCA features from each virus segment to a neural network (NN) & analyzing classification performance.
- Comparing the classification performance attained when using information gain (IG) & decision trees (DTs) to that attained when using

PCA & NNs. Additionally comparing PCA & DTs to IG & DTs for the segment generating the best results.

- Recording the time taken to construct the classifier & select feature to measure both classifier efficiency & performance for the best segment.

IV. cDNA COMPRESSION WITH PCA

The PCA is a data compression method. It compresses the data by projection unto vectors in the direction of highest variability, thus detecting even the smallest of differences. It was selected specifically, since the cDNA has only 4 differentiable units, *A*, *G*, *C*, & *T*. The viruses being analyzed are also of the same strain, so subtle differences cannot be detected easily. PCA can reduce the features significantly, while detecting the subtle differences between the cDNA sequences under analysis.

cDNA sequences features are usually represented by the cDNA position due to cDNA sequences being mostly similar except at certain positions. The cDNA sequences are also stored on databases with variable sequence lengths. For the PCA algorithm function properly the cDNA sequences must be aligned to unify their length & bring similar sections together. Furthermore cDNA sequences are nominal. Since the PCA algorithm is a numerical technique, the cDNA units have to be first encoded to numerical values before use with the PCA. The encoding scheme is discussed further in the methodology section. The derivation of the PCA algorithm is detailed in [12].

To apply the PCA algorithm to a dataset, the following steps must be performed. After encoding the DNA sequences, convert the dataset to mean centered points by subtracting the dataset values from the mean: the mean attained in (1) is a row vector containing the mean per nucleotide position; the per position mean is subtracted from every single sequence.

$$J(x_o) = \sum_{k=1}^{n} \|x_o - x_k\|^2 \quad (1)$$

Calculate the scatter/covariance matrix of the sequences: this is done by multiplying the transpose of each mean projected cDNA sequence by itself for every sequence & summing the results. Determine the eigenvectors, *e*, of the dataset arranged using highest eigenvalues, *λ*, of the resulting scatter matrix as depicted in (3).

$$S = (x - m)' * (x - m) \quad (2)$$

$$Se = \lambda e \quad (3)$$

After selecting best *n* eigenvectors with highest eigenvalues, reduce the data to *n* dimensions by multiplying the resulting eigenvectors, each individually represented by *e*, with the mean projected sequences as shown in (4).

$$a_k = e^T (x_k - m) \quad (4)$$

This is done once per sequence; each sequence will then be represented by *n* features instead of approximately 1500 features. The process will be repeated for all viral segments to ensure consistency.

V. METHODOLOGY

The following chapter details the experimentation steps. The main steps are: data collection, sequence alignment, feature selection using PCA, classifier construction & classifier evaluation.

--Data Collection: Collect cDNA data from online Influenza databases; http://www.fludb.org/ was used. Select viral H1N1, Influenza-A viral segments. The H1N1 strain was selected as it was known to cause an outbreak in 2009 by infecting multiple hosts. Additionally, half the viral sequences were Adamantane-susceptible while the other half was Adamantane-resistant. Viruses infecting the most common hosts for H1N1 were used: Avian, Human & Swine hosts. Data annotated with Adamantane-antiviral resistance & susceptibility was chosen. Duplicate & nonannotated sequences were excluded.

--Data Alignment: The PCA algorithm requires input data of equal length to work. cDNA sequences are not stored in identical lengths online, thus multiple sequence alignment (MSA) is a required step. To unify the sequences' length, align the data using a MSA program, *Mafft*. *Mafft* was used due to its ability to align high dimensionality data swiftly.

--Data Preprocessing:

(a) Divide the data to two classes using the tags Adamantane-Resistant & Adamantane-Susceptible. The sequence headers are analyzed for the aforementioned tags & the sequences are divided accordingly.

(b) Convert the data into numerical values. This is done since PCA is a statistical method & the values of cDNA are nominal: A, C, G, & T. The encoding scheme used is summarized in Table-1. The scheme includes the gaps produced during MSA.

Table.1 Encoding Scheme

Nucleotide Value	Encoding Scheme
A	-20
G	-10
-	0
C	10
T	20

(c) Divide the training data further into 70% training & 30% testing sets .

--Feature Selection:

For the training dataset:

(a)Find the mean, *m*, of the training dataset for both classes, across all nucleotide positions. Obtain mean centered data by subtracting the mean from all the sequences.

(b) Apply the PCA algorithm for the set by: finding the scatter matrix of the mean centered points. Obtain the best *n* eigenvectors corresponding to the *n* highest eigenvalues from the scatter matrix; Project the mean normalized data by multiplying it by each of the eigenvectors to get the new feature space.

For the testing dataset:

(a) For each class, obtained the mean centered by subtracting the mean of the training dataset from the values of the testing dataset.

(b) Multiply the mean projected points by the eigenvectors calculated in the training step.

The results were applied to all the viral segments despite the antiviral acting on the M1/M2 [13,14] segment, in order to thoroughly examine the classification potential.

--Classifiers Construction: Build, train & test the binary classifier differentiating between antiviral-resistant strains susceptible strains. This is implemented using NNs. NNs were chosen as it was shown before in [4] to be more robust to extreme conditions despite DTs despite being faster. They were not picked previously in [4] as the main feature selection technique of choice as IG produced more features, so DTs were chosen to shorten the training time. The NN was designed as a binary classifier that determines whether a sequence is antiviral-resistant or antiviral-susceptible:

(a) A three layered feed forward NN was used. The nprtool in Matlab was used to build, train & test the NN.

(b) The network inputs where the 3 compressed features obtained using the PCA algorithm. A dimensionality of *3* as explained in [8].

(c) Ten neurons were used in the hidden layer.

(d) The binary target outputs, *0* & *1*, were used to represent the virus's antiviral-resistance and susceptibility respectively.

(e) The default training algorithm, the scaled conjugate graduate was used. The mean squared error was used for neural network evaluation.

(f) To avoid overfitting the data, the training set was further divided into 70% training, 15% testing & 15% validation sets. The validation set was used to halt the training, when the validation set's accuracy increased compared to the training set.

--Classifier Evaluation: After each experiment, use the testing set to evaluate the results by generating confusion matrices & ROC curves. Compare the results to previous research in the field.

III. RESULTS

Training the classifier with the NNs after reducing the number of features from *1500* to *3* features using PCA yielded the results summarized in Table-2. Table-2 shows the relatively efficient & accurate performance of the technique, via time, accuracy, sensitivity, specificity & precision, when determining Adamantane resistance vs. susceptibility using NNs. Table-2 also shows the time in seconds taken to train the classifier.

Table.2 The performance of Neural Networks on the testing dataset after using PCA on the H1N1 Dataset

Segment	Time (s)	Accuracy	Sensitivity	Specificity	Precision
M1/M2	3	96.5%	99.3%	94.4%	92.9%
PB2	17	96.4%	98.7%	95%	92.1%
NP	9	94.6%	99%	91.8%	88.7%
NS1	8	87.9%	99.8%	79.1%	77.8%
HA	8	93.2%	89.6%	95%	90.1%
NA	13	92.1%	95.9%	89.8%	85.2%
PA	5	93.9%	98.9%	91.2%	85.9%

To ensure the authentic results an unused testing dataset was used to generated the results. The experiment was conducted on all segments to analyze classification accuracy across segments & discover the segment generating the most accurate results. The corresponding ROC curves of the testing dataset are shown in Fig-2. From the results in Table-2 & Fig-2, NNs using PC features did not seem to overlearn & are able to identify unseen data with relatively high accuracy. The PB1 segment was excluded due insufficient data. The M1/M2 segment showed the most promising results while the NS1 segments showed the least accurate results.

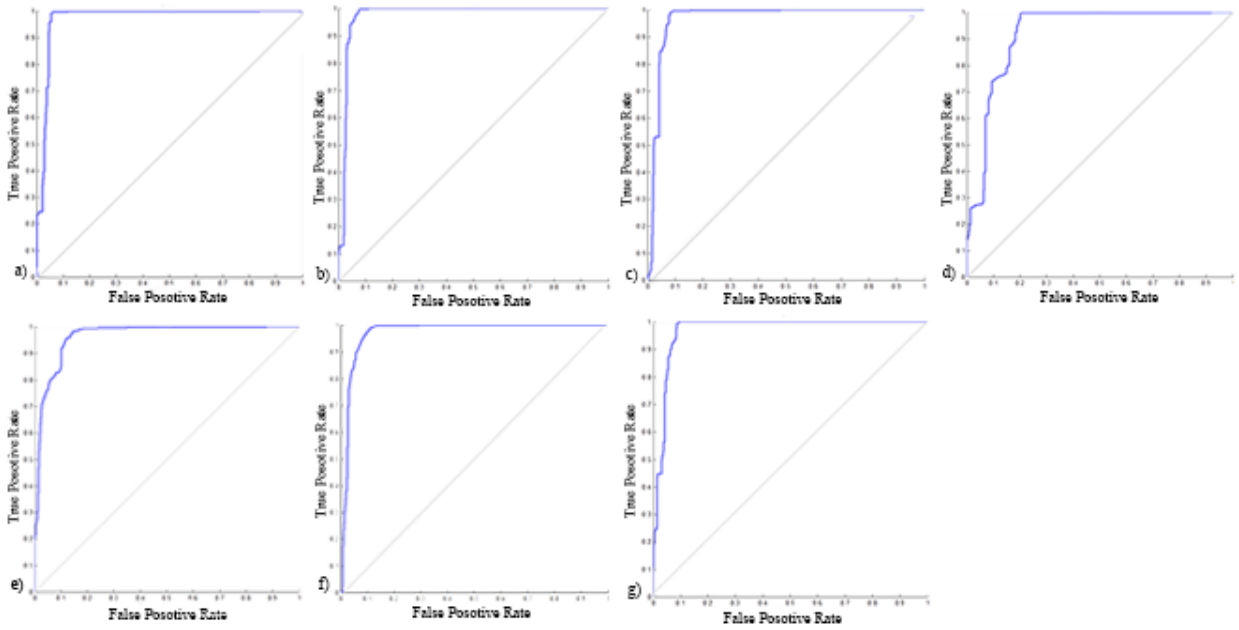

Fig. 2 ROC cures for the 7 of the H1N1 viral segments. The results are obtained by extracting the features with PCA & training NNs. The segments are (a)M1/M2, (b)PB2, (c) NP, (d) NS1, (e)HA, (f)NA, (g)PA.

The results in Table-2 are compared with another machine learning method which uses IG to select the features & J48 DTs to classify. Table-3 shows the previous results attained in the field as detailed in [4], across *7* viral segments, when IG was used to extract the best *100* positions from the sequences to determine antiviral resistance in H1N1.

Table.3 The performance of J48 DTs on the testing dataset after use of Information Gain on the H1N1 Dataset

Segment	Time (s)	Accuracy	Sensitivity	Specificity	Precision
M1/M2	0.3	98.2%	98%	98.6%	98.4%
PB2	0.4	97.4%	97.4%	97.3%	92.1%
NP	0.23	97.6%	97.7%	97.5%	98.3%
NS1	0.31	96.2%	96.7%	95.5%	96.7%
HA	1.54	96.3%	96.8%	95.5%	97.7%
NA	3.08	97.2%	97.4%	96.8%	98.1%
PA	0.64	97.5%	97.9%	96.8%	98.1%

As before the time in Table-3 is the time in seconds used to train the DT. The performance of using IG is slightly higher than PCA. The time used to generate the features selected by IG in Shaltout et al.'s previous method in [4] however ranged from *15* minutes to *3* hours depending on the dataset. This was reduced significantly using PCA to *300* seconds on average, despite a slight accuracy drop. The speed of using PCA & NNs can be useful in emergencies to quickly predict the availability of a drug, as both training & using the network takes five minutes on average.

Table.4 Comparison of PCA & IG Processing Times.

Segment	PCA Processing time (s)	IG Processing time (s)
M1/M2	97.5	855.2

A comparison of the processing time of PCA compared to IG is shown in Table-4. The segment generating the most accurate results M1/M2 is used. PCA shows a vast improvement in performance as it is 88% faster than IG when selecting features.

In attempt to improve the results & allow for a more fair comparison with Table-3, the PCA features of the M1/M2 segment were used to train a J48 decision tree. The M1/M2 as it generated the best results when using both features selection techniques . The PCA & J48 DT results detailed in Table-5 show a rise in classification accuracy with the results being almost comparable to IG & J48 DTs. This proves that with a few tweaks PCA can provide high classification accuracy at 88 times the speed attained by IG. Fig-3 below also shows the corresponding ROC curve attained from the testing dataset. The area under the curve is relatively high

ensuring accurate results.

Table.5 The performance of Decision Trees on the testing dataset after use of PCA on the H1N1 Dataset

Segment	Time (s)	Accuracy	Sensitivity	Specificity	Precision
M1/M2	0.3	97.6%	97%	98.3%	98.8%

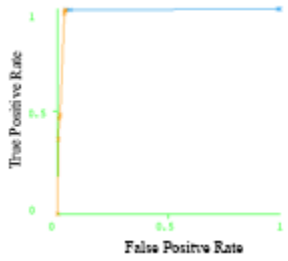

Fig. 3 ROC curve for the M1/M2 segments. The results were obtained by using PCA features & a J48 DT.

III. CONCLUSION

The experiments show that extracting PCA features is significantly faster than extracting IG features. The NN classification results show a slight accuracy drop. Adjusting the NN's learning parameters & the number of PCA features, may increase classification accuracy further in future works. Using PCA features with DTs increases the accuracy, making them comparable to results attained when features are extracted with IG. However, this concept must be tested when data is scarce or imbalanced as DTs do not perform gracefully in these cases. PCA, although a well known technique, can be a powerful feature selection & reduction tool for cDNA classification problems due to its ability to extract features while preserving data variability. The results also show that antiviral resistance can be determined with segments other than M1/M2 at varying accuracies. This should be taken into consideration when the M1/M2 segment is unavailable; a classification model that takes into account more than one segment can be a possible project expansion.

ACKNOWLEDGEMENTS

I would like to acknowledge my corresponding authors who supported my research. I would like to acknowledge Zacharie's support in all the stages.

REFERENCES

[1] N. M. Bouvier, and P. Palese. “The biology of Influenza viruses.” Vaccine, vol. 26, Sep. 2008, D49-53.

[2] E. Ghedin, N. Sengamalay, M. Shumway, J. Zaborsky, T. Feldblyum et al., “Large-scale sequencing of Human Influenza reveals the dynamic nature of viral genome evolution,” Nature, vol. 437, 2005, pp. 1162–6.

[3] T. Fislova and F. Kostolansky, “The factors of virulence of Influenza A virus,” Acta Virologica, vol. 49, 2005, pp.147-157.

[4] N. A. Shaltout, “Improving machine learning techniques for Influenza-A classification,” M.S. thesis, Dept. of CSCE, AUC, Cairo, Egypt, 2014.

[5] Y. Saeys, I. In˜aki I. and L. Pedro ., “A review of feature selection techniques in Bioinformatics,” Briefings in Bioinformatics, vol. 23, 2007, pp. 2507-2517.

[6] K. S. Leung, K.H. Lee, et al., “Data mining on DNA sequences of Hepatitis B virus.” IEEE/ACM Trans Comput Biol Bioinform, vol. 8, 2011, pp.428-40.

[7] N. A. Shaltout, M. ElHefnawi, A. Rafea, and A. Moustafa" Using Information Gain to Compare the Effeciency of Machine Learning Techniques When Classifying Influenza Based on Viral Hosts, Transactions on Engineering Technologies, 2015, pp 707-722

[8] N. A. Shaltout, M. ElHefnawi, A. Rafea, and A. Moustafa, M. Moustafa "Comparing PCA to Information Gain as a Feature Selection Method for Influenza-A Classification", ICIIBMS, 2016.

[9] P. K. Attaluri, Classifying Influenza Subtypes and Hosts using Machine Learning Techniques, ProQuest, UMI Dissertation Publishing, 2012.

[10] M. ElHefnawi, Y. M. Kadah, and F. Sherif, “Influenza A subtyping and host origin classification using Profile Hidden Markov Models,” Journal of Mechanics in Medecine and Biology, vol. 12., 2012.

[11] M. ElHefnawi, Y. M. Kadah, and F. Sherif, “Accurate classification and Hemagglutinin amino acid signatures for Influenza A virus hostorigin association and subtyping,” Virology, vol. 449, 2014.

[12] N. H. Abdi. and L. J. Williams, "Principal component analysis.". Wiley Interdisciplinary Reviews: Computational Statistics, vol. 2 , 2010, pp. 433–459.

[13] C. Wang, K. Takeuchi, L.H. Pinto and R.A. Lamb RA, "Ion channel activity of influenza A virus M2 protein: characterization of the Amantadine block". Journal of Virology 67 vol. 9, 1993, pp5585–94.

[14] X. Jing X, C. Ma, Y. Ohigashi et al. "Functional studies indicate amantadine binds to the pore of the Influenza A virus M2 proton-selective ion channel". Proc. Natl. Acad. Sci. U.S.A. 105 (31): (2008).