Forecasting symptomatic influenza A infection based on pre-exposure gene expression using a Deep Learning approach

Short title: Symptomatic influenza infection forecast and deep learning

Anna Zan[1], Zhong-Ru Xie[2], Yi-Chen Hsu[1], Yu-Hao Chen[1], Tsung-Hsien Lin[1], Yong-Shan Chang[1], and Kuan Y. Chang[1]*

[1] Computational Biology Laboratory, Department of Computer Science and Engineering, National Taiwan Ocean University, Keelung, Taiwan

[2] Computational Drug Discovery Laboratory, School of Electrical and Computer Engineering, College of Engineering, University of Georgia, Athens, GA, USA

* Corresponding author

E-mail: kchang@ntou.edu.tw (KC)

Abstract

Not everyone gets sick after an exposure to influenza A viruses (IAV). It was unclear whether forecasting symptomatic flu infection based on pre-exposure host gene expression might be possible. We utilized the human gene expression data infected with IAV subtype H1N1 or H3N2 viruses to examine this problem using the state-of-the-art deep learning approach. The results indicate that such forecast is possible and, in other words, gene expression could reflect the strength of host immune systems. Here we have developed DeepFlu—based on deep neural networks—to predict who would catch the flu prior to an exposure to IAV. We validated DeepFlu both on a leave-one-person-out internal cross-validation and an independent external cohort. DeepFlu performed better than the models using convolutional neural network, random forest, or support vector machine, achieving 70.0% accuracy, 0.787 AUROC, and 0.758 AUPR for H1N1 and 73.8% accuracy, 0.847 AUROC, and 0.901 AUPR for H3N2. Besides, DeepFlu which was trained only by pre-exposure data worked the best than by other time spans and mixed training data of H1N1 and H3N2 did not necessarily enhance prediction. These findings suggest that deep learning models could moderately recognize individuals susceptible to flu to prevent the spread of IAV.

Author summary

Influenza A, one of the leading infectious diseases, causes major health and socioeconomic burdens in the world. Hence, it is important to develop methods to accurately forecast who will succumb to influenza A virus (IAV) in advance. To demonstrate that forecasting influenza susceptibility is possible, we develop DeepFlu, based on the powerful deep learning method, to predict who will catch the flu based on 'healthy' pre-viral-exposure host gene expression. DeepFlu, which outperforms other machine learning methods, can be used to forecast susceptibility for two IAV subtypes H1N1 and H3N2. The results suggest that we could identify people at risk for the flu and alarm them in advance to lessen the spread of IAV and reduce the death tolls caused by the flu and its complications.

Keywords: Deep learning, Gene Expression, Influenza A, Susceptibility

Introduction

Influenza is a highly contagious and deadly viral respiratory disease second only to coronavirus disease 2019 (covid-19). There are about 3 to 5 million severe cases of influenza worldwide every year, resulting in approximately 291,000 to 646,000 deaths annually [1, 2]. Influenza A caused by influenza A (genus *influenzavirus A*) virus, also known as IAV, accounts for about 75% of all influenza infections [3]. IAV is also responsible for 1918, 1957, 1968, and 2009 influenza pandemics [4]. About 2.3% of the adults have contracted the IAV [5, 6], but not everyone gets sick after an exposure to IAV [7]. Slightly more than half of young adults would get infected and develop symptoms of acute upper respiratory tract infection after a controlled IAV exposure; the rest remain either asymptomatic or uninfected [8, 9].

Due to the huge socioeconomic and health burden of influenza [10-12], it is important to identify and anticipate who will contract influenza and become ill. Delaying the recognition of infected persons with IAV will increase the viral transmission [13] and sometimes worsen the infection, thereby increasing the death toll. Although vaccination is recommended to prevent the influenza and improve public health in schools, hospitals, communities, and societies [14], flu vaccines are not necessarily working for everyone [15]. Instead, foreseeing who will succumb to IAV infection prior to viral exposure could be an alternative to prevent the spread of IAV and reduce the death toll from the IAV and its related complications.

However, it remains unclear whether host susceptibility to IAV is predictable based on pre-viral-exposure host gene expression. Forecasting the onset of viral infection prior to viral exposure has been only demonstrated on respiratory syncytial virus (RSV) [9]. Two prediction models have shown promising results for forecasting host susceptibility to RSV. One is an epsilon support vector regression model [16] with a radial basis function kernel [9] learning directly from host gene expression. The other is a LASSO regularized regression model [9] based on modulation of biological pathways, deriving from the Hallmark gene sets [17] of the Molecular Signature Database (MSigDB). No one yet has demonstrated whether forecasting symptomatic IAV infection based on pre-viral-exposure host gene expression is feasible and, if so, how accurate the forecast is.

Forecasting host susceptibility to IAV differs from detecting IAV infection during the onset of influenza symptoms based on host gene expression. Although both host gene expression responses are influenced by the same factors such as host genetics [18], immunological memory [19], seasonal and circadian changes [20], age [21], and sex [22, 23], they attend to different times. Unlike Forecasting host susceptibility to IAV centers on the time prior to an IAV exposure, but IAV-infected hosts can be detectable only after exposure to the IAV [24-26].

It has been a success to detect IAV infection based on host gene expression. The peripheral blood gene expression signatures of post IAV infection different from those of bacterial and other respiratory viral infections can be used to detect symptomatic IAV infection [8, 22], which have been applied to the real-world cases of 2009 H1N1 pandemic with over 90% accuracy. In fact, comparing top 50 discriminative genes [25], the infection signatures across different IAV sub-strains are so similar [25] that some have exhibited cross-strain detectability [27]. Remarkably, a multi-cohort analysis further demonstrates that detecting symptomatic subjects infected with IAVs can count on only 11 genes derived from an influenza meta-signature [22]. Whether such success can occur in forecasting host susceptibility to IAV is waiting to unravel.

To our knowledge, this is the first study to attempt the-state-of-the-arts deep learning to forecast host susceptibility to IAV prior to a viral exposure. In this study, we report the development of a deep learning model called DeepFlu.

Results

Best Learning Method for the Gene Expression Data

By randomly dividing the influenza gene expression data into 80% training and 20% testing data, Table 1 shows the performances of four different predictive modeling approaches using the full 22,277 array probes. The setting of the training and testing data used here, which were not exclusive to those prior to exposure, differed from most of the experiments throughout this study. All the four models, CNN, DNN, RF, and SVM, demonstrate that the special gene expression signals of the flu infected subjects could be learned. Among the four models, DNN generally performed the best, reaching averagely 96.1% accuracy, 0.957 AUROC, and 0.963 AUPR for H1N1; 95.4% accuracy, 0.954 AUROC, and 0.965 AUPR for H3N2. Thus, DNN was chosen as the backbone of DeepFlu for the following development.

Table 1. Model comparison for distinguishing symptomatic flu-infected subjects.

Method	Accuracy	Sensitivity	Specificity	Precision	AUROC	AUPR
H1N1						
CNN	0.955	0.**951**	0.960	0.929	0.955	0.949
DNN	**0.961**	0.936	0.**978**	0.**964**	**0.957**	0.963
RF	0.911	0.835	0.960	0.927	0.898	**0.969**
SVM	0.841	0.639	0.891	0.881	0.789	0.859
H3N2						
CNN	0.937	0.962	0.914	0.940	0.946	0.951
DNN	0.**954**	0.**975**	0.**932**	0.**942**	**0.954**	0.965
RF	0.924	0.930	0.921	0.927	0.925	**0.984**
SVM	0.867	0.851	0.888	0.898	0.866	0.913

AUROC: area under receiver operating characteristic curve; AUPR: area under precision-recall curve; CNN: convolutional neural network; DNN: deep neural network; RF: random forest; SVM: support vector machine.

Time Spans on Prediction Performance

Based on the leave-one-person-out (L1PO) cross-validation, Table 2

compares the prediction performances of DeepFlu trained under different time spans. T_0, T_{24}, T_{48} and T_{all} represent the expression data taking from the beginning (one day before exposure) to the moment of exposure, one day after exposure, two days after exposure, and one week after exposure, respectively (Fig 1). In other words, T_0 contains only the expression data without viral exposure. The results show that training by the data prior to viral exposure (T_0) was the best for DeepFlu to predict whether one would catch the flu and develop symptoms after an exposure, achieving 70.0% accuracy, 0.787 AUROC, and 0.758 AUPR for H1N1 and 73.8% accuracy, 0.847 AUROC, and 0.901 AUPR for H3N2.

Table 2. Performance comparison of DeepFlu trained under different time spans by L1PO cross-validation.

Span	Accuracy	Sensitivity	Specificity	Precision	AUROC	AUPR
H1N1						
T_{All}	0.637	0.556	0.668	0.618	0.661	0.638
T_{48}	0.658	0.602	0.689	0.662	0.713	0.701
T_{24}	0.669	0.613	0.715	0.679	0.725	0.712
T_0	**0.700**	**0.616**	**0.822**	**0.718**	**0.787**	**0.758**
H3N2						
T_{All}	0.638	0.707	0.548	0.649	0.668	0.688
T_{48}	0.661	0.683	0.635	0.688	0.679	0.744
T_{24}	0.689	0.689	0.723	0.759	0.759	0.806
T_0	**0.738**	**0.722**	**0.756**	**0.770**	**0.847**	**0.901**

AUROC: area under receiver operating characteristic curve; AUPR: area under precision-recall curve; L1PO: leave one person out.

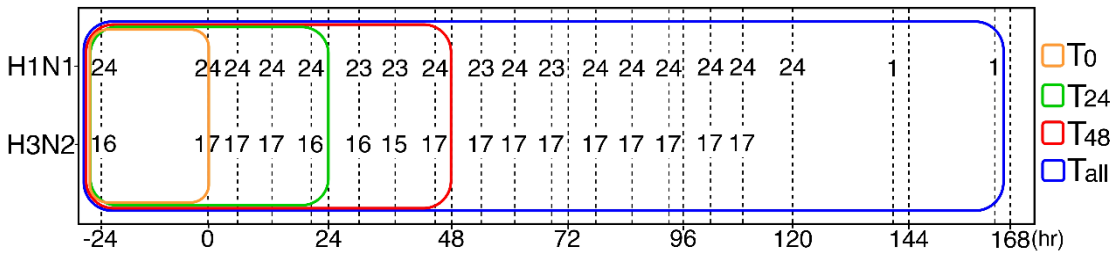

Fig 1. Datasets in four different time spans. T_0, T_{24}, T_{48}, and T_{all} represent the data taking from 1 day before influenza exposure up to immediately before exposure, 24h, 48h, and 168h after exposure, respectively. Each number in the dash line is the number of subjects at that time point.

157 **Feature Sets on Prediction Performance**

158 Based on the L1PO cross-validation test focusing on host gene expression
159 prior to exposure (T_0), Table 3 compares the performances of the four predictive
160 models using three different feature sets for forecasting symptomatic flu-
161 infected subjects. Besides DeepFlu, the three other predictive models were
162 CNN, RF, and SVM; the three feature sets were the complete probe set of the
163 microarray (22,277 features), the Hallmark genes [17] of MSigDB (4,164
164 corresponding features), and the top discriminative genes for distinguishing
165 IAV post infection (64 and 61 corresponding features for H1N1 and H3N2
166 respectively) [16]. The results indicate that the prediction models using more
167 features generally performed better and DeepFlu achieved the overall best
168 performance using the full 22,277 features for both H1N1 and H3N2. For H1N1,
169 DeepFlu had 70.0% accuracy, 0.787 AUROC, and 0.758 AUPR; for H3N2, it
170 reached 73.8% accuracy, 0.849 AUROC, and 0.901 AUPR.

171 Compared to using the full 22,277 features, DeepFlu gave a comparable but
172 weaker performance by using 4,164 features. Still it outperformed the other
173 three models.

174 For using the top discriminative genes of IAV post infection, all the four
175 models performed poorly. All the four models using the limited gene set gave
176 the worst AUROC and AUPR values for H1N1. For H3N2, the models
177 performed slightly better than a random guess.

178

179 **Table 3. Performance comparison using different feature sets trained under**
180 **T_0 by L1PO cross-validation.**

Feature #	Method	Accuracy	Sensitivity	Specificity	Precision	AUROC	AUPR
H1N1							
64	**CNN**	**0.450**	**0.647**	0.269	**0.443**	**0.495**	**0.432**
64	**DeepFlu**	0.336	0.430	0.245	0.341	0.298	0.364
64	**RF**	0.293	0.268	**0.319**	0.281	0.165	0.337
64	**SVM**	0.135	0.120	0.151	0.120	0.083	0.319
*4225	**CNN**	0.651	0.629	**0.665**	**0.646**	0.723	0.633
4164	**DeepFlu**	0.**691**	**0.835**	0.573	0.642	**0.762**	**0.732**
4164	**RF**	0.513	0.498	0.529	0.514	0.517	0.466
4164	**SVM**	0.620	0.640	0.600	0.615	0.631	0.564

*22500	**CNN**	0.588	0.470	0.664	0.486	0.636	0.624
22277	**DeepFlu**	**0.700**	**0.616**	**0.822**	**0.718**	**0.787**	**0.758**
22277	**RF**	0.519	0.495	0.543	0.520	0.545	0.482
22277	**SVM**	0.585	0.586	0.585	0.617	0.627	0.604
H3N2							
64	**CNN**	0.548	0.628	0.468	0.572	**0.563**	**0.611**
61	**DeepFlu**	0.486	0.461	0.494	0.565	0.422	0.492
61	**RF**	0.**615**	0.**689**	**0.506**	0.**670**	0.553	0.546
61	**SVM**	0.528	0.574	0.476	0.552	0.522	0.524
*4225	**CNN**	0.550	0.550	0.550	0.581	0.556	0.534
4164	**DeepFlu**	0.653	0.615	0.696	0.696	**0.766**	**0.821**
4164	**RF**	0.**690**	0.602	0.**818**	0.**828**	0.712	0.719
4164	**SVM**	0.624	0.**794**	0.432	0.612	0.723	0.760
*22500	**CNN**	0.705	**0.739**	0.667	0.715	0.812	0.863
22277	**DeepFlu**	**0.738**	0.722	**0.756**	0.770	**0.847**	**0.901**
22277	**RF**	0.730	0.723	0.742	0.**803**	0.782	0.808
22277	**SVM**	0.702	0.655	0.755	0.750	0.739	0.757

AUROC: area under receiver operating characteristic curve; AUPR: area under precision-recall curve; CNN: convolutional neural network; L1PO: leave one person out; RF: random forest; SVM: support vector machine. *Zero paddings are utilized to satisfy the requirement of CNN.

Mixed Training Data on Prediction Performance

Because the post infection signatures of H1N1 and H3N2 were similar, sharing up to 90% of top 50 discriminative genes [25], and that of H3N2 could be used to detect H1N1 infection [27], it was speculated that extending the training set by mixed data of H1N1 and H3N2 might enhance prediction performance. Instead of training by the expression data for the same IAV strain as seen in Table 3, Table 4 compares the performance of the L1PO cross validation of the predictive models trained by the mixed T_0 data of H1N1 and H3N2. Table 4 shows that despite a drop in performance, DeepFlu was still the best for forecasting symptomatic influenza A-infected subjects prior to viral exposure. For H1N1, DeepFlu using the full 22,277 features of the mixed training data achieved 72.1% accuracy, 0.752 AUROC, and 0.676 AUPR, not better than that trained only by the H1N1 data with 70.0% accuracy, 0.787 AUROC, and 0.758 AUPR. Similarly, as for H3N2, DeepFlu using the mixed data dropped performance than using the H3N2 data alone, but retained a high

performance, keeping more than 0.790 AUROC and AUPR. The results indicate that the mixed H1N1 and H3N2 training data neither improved identifying who were susceptible to H1N1 nor those to H3N2 infection.

Table 4. Performance comparison of DeepFlu trained using different feature sets by mixed H1N1 and H3N2 T_0 data by L1PO cross-validation.

Feature #	Accuracy	Sensitivity	Specificity	Precision	AUROC	AUPR
H1N1						
64	0.529	0.545	0.517	0.427	0.488	0.389
4164	0.682	**0.787**	0.610	0.574	0.751	0.660
22277	0.**721**	0.625	**0.782**	**0.656**	**0.752**	**0.676**
H3N2						
61	0.688	0.611	**0.775**	0.759	0.522	0.652
4164	0.696	**0.785**	0.596	0.687	0.748	0.772
22277	**0.707**	0.650	0.771	**0.764**	**0.796**	**0.855**

AUROC: area under receiver operating characteristic curve; AUPR: area under precision-recall curve; L1PO: leave one person out.

External Validation on DeepFlu

To examine the robustness of DeepFlu, an independent validation set of 38 H1N1 and 21 H3N2 records from 19 H1N1 and 21 H3N2 subjects were used. Table 5 shows the performance of DeepFlu on the external validation data. Here DeepFlu achieved 71.4% accuracy, 0.700 AUROC, and 0.723 AUPR for H1N1; 73.5% accuracy, 0.732 AUROC, and 0.749 AUPR for H3N2.

Table 5. External validation of DeepFlu.

	Accuracy	Sensitivity	Specificity	Precision	AUROC	AUPR
H1N1	0.714	0.733	0.767	0.700	0.700	0.723
H3N2	0.735	0.714	0.750	0.667	0.732	0.749

AUROC: area under receiver operating characteristic curve; AUPR: area under precision-recall curve.

Discussion

According to the L1PO cross-validation results, we have demonstrated that it

220 is possible based on pre-IAV-exposure host gene expression to forecast whether
221 an individual would catch the flu and develop symptoms immediately after an
222 exposure to IAV. The results are consistent with the previous finding on the
223 predictability of RSV susceptibility [9]—both studies demonstrate the
224 possibility of forecasting susceptibility of respiratory viral infections. This is
225 also the first study to demonstrate that deep learning can work for such forecast.
226 Feature selection can be important for predicting IAV susceptibility. Our
227 results generally show that the larger the feature size, the better the prediction
228 performance. In fact, DeepFlu using the complete 22,274 features is the optimal
229 model. The DNN model which used the Hallmark genes of MSigDB, about only
230 one-fifth of the complete features, could still reach a comparable performance,
231 indicating that there might exist limited candidate genes closely relating to
232 one's susceptibility to IAV. However, the underlying biological mechanism
233 might be too complex to learn well with the selected features. Instead entire
234 features became a better alternative.
235 The top 50 discriminative post-IAV-infection genes was inadequate to be used
236 as features for this problem. The models using the limited IAV infection genes
237 all consistently performed poorly compared with the others. This poor result
238 suggests that the IAV infection genes, which were stimulated after viral
239 exposure, may significantly differ from the genes relating to one's susceptibility
240 to IAV, which may regulate and control immune responses. In other words, the
241 biological mechanisms behind IAV post infection, such as inflammatory
242 responses [28], might differ from those behind IAV immunity.
243 The results also indicate that DeepFlu performs better for recognizing gene
244 expression patterns of IAV immunity than the other three models, including
245 the CNN models. DeepFlu outperformed the CNN model in both random
246 setting (80% training and 20% test data) and L1PO setting, perhaps because
247 gene expression profiles did not conform to the CNN properties—neither
248 subsampling gene chips would hold the original meaning nor the same
249 expression patterns could happen in different regions of gene chips.
250 Why forecasting H3N2 susceptibility could work better than forecasting
251 H1N1 susceptibility is unsettled. We suspect the length of latency period may
252 be a factor, for the maximal symptoms of H3N2 and H1N1 patients peak at day
253 two and day four, respectively [9]. The H3N2 models foresee clearer, perhaps
254 because T_0 is closer to the symptom onset of H3N2 than that of H1N1.
255 Moreover, using which time span to train is also important. The results
256 support that the best outcome happened when both the training and test data
257 came from the same time span—prior to exposure (T_0). Having other time

spans might bring in other signals or noises, thus lowering the performance.

The performance of forecasting may still have room to improve. However, by mixing H1N1 and H3N2 records to expend the training data may not be the way to enhance prediction performance. Using the mixed data, the performance of DeepFlu dropped for identifying the susceptible to H1N1 or H3N2, suggesting there may exist distinct differences in susceptibility to different IAV strains.

There are some limitations in this study. First, limited influenza gene expression data are available. Although DeepFlu excels even with limited gene expression data, the deep learning model may improve further with more training data. Second, it is unknown if the subjects were immunologically naïve to IAV. Such conditions may need to be controlled for a more appropriate comparison or validation. Third, the biological mechanisms behind IAV susceptibility or immunity remain unclear. Although the development of RSV infection has linked the presence of heme metabolism of hosts using an enrichment analysis of the target gene set [9], DeepFlu with full features, which does not rely on specific gene sets, may require more subtle approaches to study the mechanisms behind IAV susceptibility or immunity.

Conclusion

Gene expression can portend IAV susceptibility and deep learning can catch the expression signals to identify people at risk for the flu and alarm them in advance to reduce hospitalization and mortality caused by IAV.

Materials and Methods

Datasets Description

The datasets used in this study were the human peripheral blood gene expression data obtained from Gene Expression Omnibus (GEO) database. RMA normalization [29] was performed on each of the datasets. GEO GSE52428 [25] which was done by the Affymetrix Human Genome U133A 2.0 array with 22,277 probes was the primary dataset for training and internal testing. 24 and 17 human subjects who aged 19 to 41 were inoculated intranasally with influenza A H1N1 (A/Brisbane/59/2007) or H3N2 (A/Wisconsin/67/2005),

respectively. All subjects were healthy prior to the inoculation and monitored about every 8 hours over the course of 7 days. Consequently, 9 of the 24 H1N1 subjects and 8 of the 17 H3N2 subjects became symptomatic IAV-infected. 3 H1N1 subjects who developed symptoms but not infected by the H1N1 viruses were not classified as symptomatic IAV-infected. The rest were either asymptomatic or uninfected. Consequently, a total of 382 H1N1 and 267 H3N2 expression records were collected (Fig 1).

Being symptomatic IAV-infected is defined as the presence of both clinical flu symptoms and microbiological infection. Technically speaking, a symptomatic infected subject should hold the modified Jackson score $\geqq 6$ for at least 5 consecutive days and a positive viral culture or RT-PCR viral test [30] for at least 2 consecutive days. The modified Jackson score [31] assesses each of eight symptoms on a severity scale of 0 to 3, including sneeze, cough, malaise, runny nose, stuffy nose, sore throat, myalgia, and headache, to determine the presence or absence of flu symptoms. The viral culture or RT-PCR test is used to confirm the existence of IAV infection. Once the subject was identified as symptomatic IAV-infected, all of the subject's records were marked as symptomatic infected.

The external validation dataset came from GEO GSE73072 [32]. The experiments of DEE2 and DEE3 in this dataset were the same as those of GSE52428, but they were done by the Affymetrix Human Genome U133A 2.0 array with 12,023 genes instead. Therefore, the validation set were DEE4 and DEE5, which challenged another 19 and 21 subjects with the H1N1 and H3N2 viruses respectively.

Leave-one-person-out cross-validation

Due to the limited gene expression data, a leave-one-person-out (L1PO) cross-validation strategy on all subjects was applied to the primary dataset to evaluate how well the models perform for predicting whether a previously unknown subject would develop symptoms and become IAV-infected upon to viral exposure. This performance test involves a rotation of all individuals partitioning into one-person validation set and the remaining set. Models which were trained using the observations of the remaining set were tested on the validation set. Because of stochastic nature of the prediction models, we repeated the procedure 100 times to estimate the overall model performance.

DeepFlu Layout

Fig 2 illustrates the architecture of DeepFlu designed for predicting symptomatic influenza infection. DeepFlu is based on Deep Neural Network (DNN), also known as multilayer perceptron, which is a feedforward neural network that can distinguish linear non-separable patterns [33]. DeepFlu is a 6-layer DNN, consisting four fully connected hidden layers each with 100 nodes, one input layer to extract the gene expression information, and one output layer to predict the presence of symptomatic flu infection.

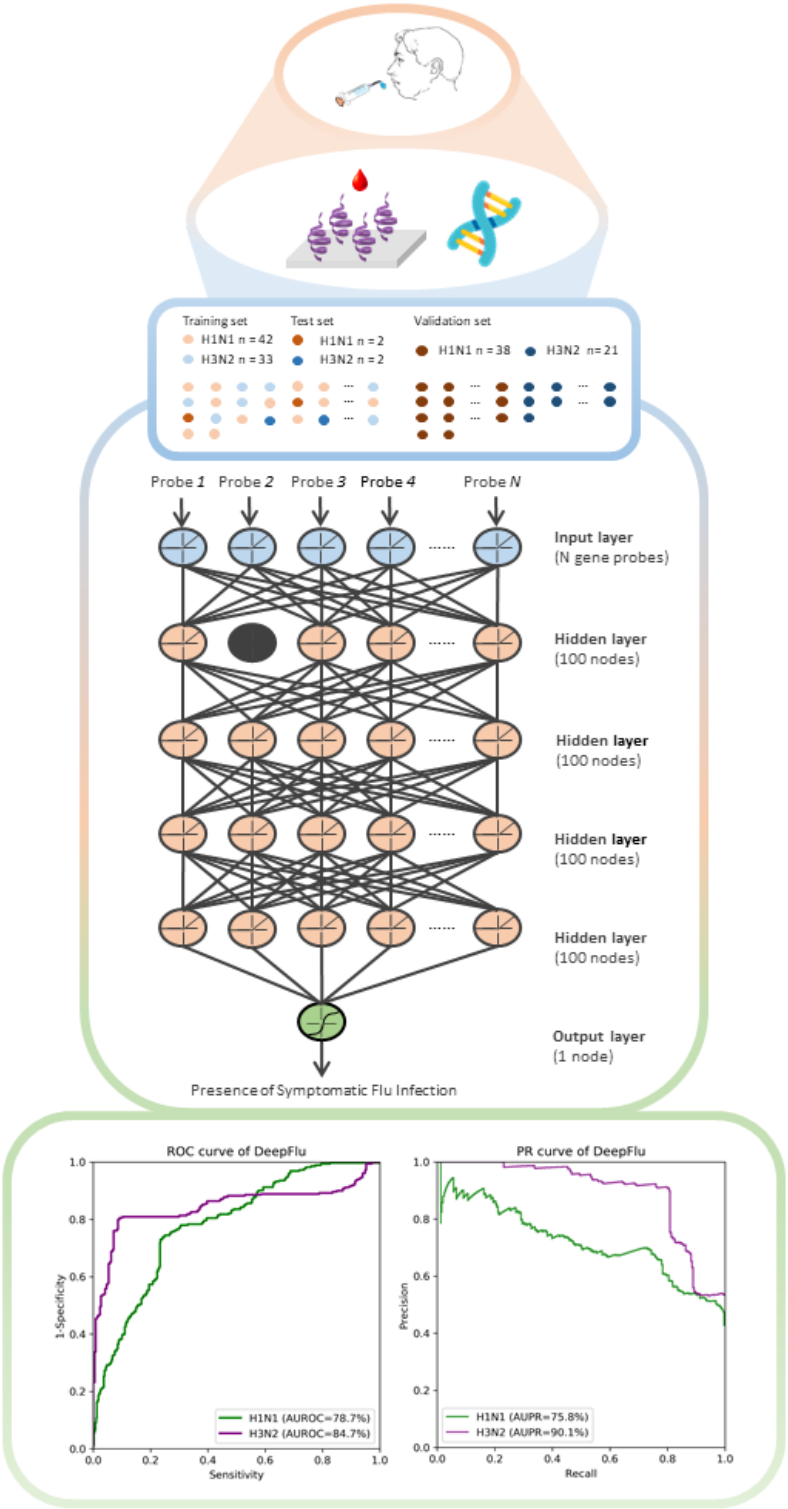

Fig 2. DeepFlu Architecture. DeepFlu is a 6-layer DNN. The input layer consists of N nodes corresponding to the number of given gene expression probes. The four fully connected hidden layers consist of 100, 100, 100, and 100 nodes respectively. The black node in the hidden layer represents the dropout. The output layer utilizes single node to generate a binary outcome.

The following parameters and methods were used in DeepFlu. Both the input and hidden layers used a rectified linear unit (ReLU) function [34] as the activation function, for it could overcome the vanishing gradient for nonnegative inputs. The output layer adopted a sigmoid function for a binary decision, whose outcome was strictly increasing and [0, 1]. To minimize the loss measured by binary cross-entropy, adaptive moment estimation (ADAM) optimization function [35], which used the moving averages of the past gradient and squared gradient to update model weights, was applied during backpropagation, for ADAM could work effectively for high-dimensional parameter spaces. In addition, 150 training epochs were run with a learning rate of 0.001 without weight decay and 10% dropout as regularization was applied to the first hidden layer to avoid overfitting.

The performance of DeepFlu was evaluated by both L1PO cross-valiation on the primary dataset and the external validation.

Convolutional Neural Network

Convolutional neural networks (CNNs) [36] are one of the most popular deep learning models, especially for image recognition. Many well-known image recognition models such as AlexNet [37], GoogLeNet [38], and ResNet [39] are based on CNNs. However, it remains unclear how CNNs perform in gene expression data.

We built a 10-layer CNN model, extending from the 6-layer DeepFlu by adding two convolutional layers and two pooling layers. CNNs, which can acquire spatial relations of data, can be divided into two sections. The first section is to extract features through alternate convolutional filters and pooling layers, and the second section is to learn through fully connected layers. Moreover, the activation function of the convolutional layers was a ReLU function [34] and the filter of the pooling layers was a max pooling function. The rest of parameters and methods used in the CNN model were adopted from those of DeepFlu. Because the CNN model had more layers and parameters, the CNN model would have more epochs than DeepFlu.

The CNN models can be described simply as follows: the input is a tensor X $\in \mathbb{R}^{w \times h \times d}$, where w, h, and d are the image width, height, and depth, respectively, and the depth refers to the number of channels. The convolution transforms the input X using m filter tensors K $\in \mathbb{R}^{n \times n \times d}$ with q zero-

paddings and s strides into m feature maps $M \in \mathbb{R}^{\left(\frac{w-n+2q}{s}+1\right) \times \left(\frac{h-n+2q}{s}+1\right) \times m}$, where n is the order of filter kernels, d keeps the same number of channels as the input, and m which determines the number of the output channels is the number of filter kernels. To effectively reduce the dimensions of feature maps, the pooling layer adopts a filter kernel. The pooling layer transforms the feature map using a pooling filter tensor $P \in \mathbb{R}^{q \times q \times m}$ into a pooled feature map, where q is the order of pooling kernels and m is the channel number of feature maps. Excluding the convolutional and pooling layers, the makeup of the CNN models used in this study is the same as that of DeepFlu.

Random Forest

Random Forest (RF) is a classification method using an ensemble of bagged decision trees with randomly selected features [40]. The RF algorithm embeds the concept of random subset feature selection into bagged decision trees which are trained on random samples with replacement to avoid overfitting. To determine the class of the predicted case, each decision tree which is best split by a small fraction of randomly selected features casts a vote independently under majority rule. In this work, the parameters and methods used for the RF models were adopted from the Chang study [41]. Each RF classifier had 1,000 decision trees. Gini impurity [42] was used to measure the quality of a split by a selected feature and when looking for a best split, the number of tried features was set to be square root of the total features [43].

Support Vector Machine

SVM is also a classification method that searches an optimal hyperplane to separate two classes with maximum margin [44]. In this work, a radial basis function kernel [45] was used for our SVM models, whose cost and gamma parameters [46] for the kernel were set as 1 and one over the number of features.

Performance Comparison

The performances of the predictive models were measured by sensitivity, specificity, precision, accuracy, area under receiver operating characteristic curve (ROC), and area under precision recall curve (PRC). These evaluation measurements were calculated as follows:

$$\text{Sensitivity (recall)} = \frac{TP}{TP + FN} \quad (1)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (2)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

where TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negative, respectively. Sensitivity, specificity, precision, and accuracy are the percentage of the correct predictions on positive data, negative data, predicted positive data, and all the data, respectively.

Like accuracy, ROC and PRC are used to evaluate the overall performance of a binary classifier [47, 48]. ROC and PRC are based on dynamic positive-negative prediction thresholds instead of a static threshold like accuracy. ROC is the area under the curve which plots true positive rates (sensitivity) against false positive rates using different prediction thresholds; similar to ROC, PRC is the area under the curve by plotting precision against recall rates (sensitivity). ROC and PRC are in the range [0,1]. The higher ROC or PRC, the better the classifier. A perfect classifier has ROC or PRC equal to 1. However, unlike ROC, PRC only focuses on positive cases. The ROC value tends to be overly optimistic for an imbalance dataset, but PRC does not [49].

Acknowledgments

This study was supported by Ministry of Science and Technology, R.O.C. under MOST-108-2221-E-019-052 (https://www.most.gov.tw). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Contributions

Conceptualization: Zhong-Ru Xie, Kuan Y. Chang.
Formal analysis: Anna Zan, Yi-Chen Hsu, Kuan Y. Chang.
Investigation: Anna Zan, Yi-Chen Hsu, Yu-Hao Chen, Tsung-Hsien Lin,

433 Yong-Shan Chang, Kuan Y. Chang.
434 **Methodology:** Anna Zan, Yi-Chen Hsu, Yu-Hao Chen, Tsung-Hsien Lin,
435 Kuan Y. Chang.
436 **Writing – original draft:** Yi-Chen Hsu, Yu-Hao Chen, Zhong-Ru Xie, Kuan Y.
437 Chang.
438 **Writing – review & editing:** Anna Zan, Zhong-Ru Xie, Kuan Y. Chang.

439 References

440 1. Iuliano AD, Roguski KM, Chang HH, Muscatello DJ, Palekar R, Tempia S, et al.
441 Estimates of global seasonal influenza-associated respiratory mortality: a modelling
442 study. Lancet. 2018;391(10127):1285-300. Epub 2017/12/19. doi: 10.1016/S0140-
443 6736(17)33293-2. PubMed PMID: 29248255; PubMed Central PMCID:
444 PMCPMC5935243.
445 2. WHO. Influenza (Seasonal) Facts: World Health Organization; 2018 [updated 6
446 November 2018; cited 2020 1 May 2020]. Available from:
447 http://www.who.int/news-room/fact-sheets/detail/influenza-(seasonal).
448 3. Heikkinen T, Ikonen N, Ziegler T. Impact of influenza B lineage-level mismatch
449 between trivalent seasonal influenza vaccines and circulating viruses, 1999-2012.
450 Clin Infect Dis. 2014;59(11):1519-24. Epub 2014/08/21. doi: 10.1093/cid/ciu664.
451 PubMed PMID: 25139969.
452 4. Paules C, Subbarao K. Influenza. Lancet. 2017;390(10095):697-708. Epub
453 2017/03/18. doi: 10.1016/S0140-6736(17)30129-0. PubMed PMID: 28302313.
454 5. Park JE, Ryu Y. Transmissibility and severity of influenza virus by subtype. Infect
455 Genet Evol. 2018;65:288-92. Epub 2018/08/14. doi: 10.1016/j.meegid.2018.08.007.
456 PubMed PMID: 30103034.
457 6. Jayasundara K, Soobiah C, Thommes E, Tricco AC, Chit A. Natural attack rate of
458 influenza in unvaccinated children and adults: a meta-regression analysis. BMC Infect
459 Dis. 2014;14:670. Epub 2014/12/17. doi: 10.1186/s12879-014-0670-5. PubMed
460 PMID: 25495228; PubMed Central PMCID: PMCPMC4272519.
461 7. Iwasaki A, Pillai PS. Innate immunity to influenza virus infection. Nature Reviews
462 Immunology. 2014;14:315-28. doi: 10.1038/nri3665.
463 8. Zaas AK, Chen M, Varkey J, Veldman T, Hero AO, 3rd, Lucas J, et al. Gene
464 expression signatures diagnose influenza and other symptomatic respiratory viral
465 infections in humans. Cell Host Microbe. 2009;6(3):207-17. Epub 2009/08/12. doi:
466 10.1016/j.chom.2009.07.006. PubMed PMID: 19664979; PubMed Central PMCID:
467 PMCPMC2852511.

468 9. Fourati S, Talla A, Mahmoudian M, Burkhart JG, Klén R, Henao R, et al. A
469 crowdsourced analysis to identify ab initio molecular signatures predictive of
470 susceptibility to viral infection. Nature Communications. 2018;9(1):4418-28. doi:
471 10.1038/s41467-018-06735-8.
472 10. Cassini A, Colzani E, Pini A, Mangen MJ, Plass D, McDonald SA, et al. Impact of
473 infectious diseases on population health using incidence-based disability-adjusted
474 life years (DALYs): results from the Burden of Communicable Diseases in Europe
475 study, European Union and European Economic Area countries, 2009 to 2013. Euro
476 Surveill. 2018;23(16). Epub 2018/04/26. doi: 10.2807/1560-7917.ES.2018.23.16.17-
477 00454. PubMed PMID: 29692315; PubMed Central PMCID: PMCPMC5915974.
478 11. Putri W, Muscatello DJ, Stockwell MS, Newall AT. Economic burden of seasonal
479 influenza in the United States. Vaccine. 2018;36(27):3960-6. Epub 2018/05/29. doi:
480 10.1016/j.vaccine.2018.05.057. PubMed PMID: 29801998.
481 12. Gordon A, Reingold A. The Burden of Influenza: a Complex Problem. Curr
482 Epidemiol Rep. 2018;5(1):1-9. Epub 2018/03/06. doi: 10.1007/s40471-018-0136-1.
483 PubMed PMID: 29503792; PubMed Central PMCID: PMCPMC5829127.
484 13. Poon LL, Song T, Rosenfeld R, Lin X, Rogers MB, Zhou B, et al. Quantifying
485 influenza virus diversity and transmission in humans. Nat Genet. 2016;48(2):195-200.
486 Epub 2016/01/05. doi: 10.1038/ng.3479. PubMed PMID: 26727660; PubMed Central
487 PMCID: PMCPMC4731279.
488 14. Houser K, Subbarao K. Influenza vaccines: challenges and solutions. Cell Host
489 Microbe. 2015;17(3):295-300. Epub 2015/03/15. doi: 10.1016/j.chom.2015.02.012.
490 PubMed PMID: 25766291; PubMed Central PMCID: PMCPMC4362519.
491 15. Nakaya HI, Wrammert J, Lee EK, Racioppi L, Marie-Kunze S, Haining WN, et al.
492 Systems biology of vaccination for seasonal influenza in humans. Nat Immunol.
493 2011;12(8):786-95. Epub 2011/07/12. doi: 10.1038/ni.2067. PubMed PMID:
494 21743478; PubMed Central PMCID: PMCPMC3140559.
495 16. Chang C-C, Lin C-J. LIBSVM: a library for support vector machines. ACM
496 transactions on intelligent systems and technology (TIST). 2011;2(3):27.
497 17. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The
498 Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst.
499 2015;1(6):417-25. Epub 2016/01/16. doi: 10.1016/j.cels.2015.12.004. PubMed PMID:
500 26771021; PubMed Central PMCID: PMCPMC4707969.
501 18. Casanova JL, Abel L. The human model: a genetic dissection of immunity to
502 infection in natural conditions. Nat Rev Immunol. 2004;4(1):55-66. Epub 2004/01/06.
503 doi: 10.1038/nri1264. PubMed PMID: 14704768.
504 19. Zhang A, Stacey HD, Mullarkey CE, Miller MS. Original Antigenic Sin: How First
505 Exposure Shapes Lifelong Anti-Influenza Virus Immune Responses. J Immunol.

506 2019;202(2):335-40. Epub 2019/01/09. doi: 10.4049/jimmunol.1801149. PubMed
507 PMID: 30617114.
508 20. Dopico XC, Evangelou M, Ferreira RC, Guo H, Pekalski ML, Smyth DJ, et al.
509 Widespread seasonal gene expression reveals annual differences in human immunity
510 and physiology. Nat Commun. 2015;6:7000. Epub 2015/05/13. doi:
511 10.1038/ncomms8000. PubMed PMID: 25965853; PubMed Central PMCID:
512 PMCPMC4432600.
513 21. Qi Q, Liu Y, Cheng Y, Glanville J, Zhang D, Lee JY, et al. Diversity and clonal
514 selection in the human T-cell repertoire. Proc Natl Acad Sci U S A.
515 2014;111(36):13139-44. Epub 2014/08/27. doi: 10.1073/pnas.1409155111. PubMed
516 PMID: 25157137; PubMed Central PMCID: PMCPMC4246948.
517 22. Andres-Terre M, McGuire HM, Pouliot Y, Bongen E, Sweeney TE, Tato CM, et al.
518 Integrated, Multi-cohort Analysis Identifies Conserved Transcriptional Signatures
519 across Multiple Respiratory Viruses. Immunity. 2015;43(6):1199-211. Epub
520 2015/12/20. doi: 10.1016/j.immuni.2015.11.003. PubMed PMID: 26682989; PubMed
521 Central PMCID: PMCPMC4684904.
522 23. Cutolo M, Sulli A, Capellino S, Villaggio B, Montagna P, Seriolo B, et al. Sex
523 hormones influence on the immune system: basic and clinical aspects in
524 autoimmunity. Lupus. 2004;13(9):635-8. Epub 2004/10/16. doi:
525 10.1191/0961203304lu1094oa. PubMed PMID: 15485092.
526 24. Zaas AK, Chen M, Varkey J, Veldman T, Hero AO, Lucas J, et al. Gene Expression
527 Signatures Diagnose Influenza and Other Symptomatic Respiratory Viral Infections in
528 Humans. Cell Host and Microbe. 2009;6(3):207-17. doi:
529 https://doi.org/10.1016/j.chom.2009.07.006.
530 25. Woods CW, McClain MT, Chen M, Zaas AK, Nicholson BP, Varkey J, et al. A host
531 transcriptional signature for presymptomatic detection of infection in humans
532 exposed to influenza H1N1 or H3N2. PLoS One. 2013;8(1):e52198. Epub 2013/01/18.
533 doi: 10.1371/journal.pone.0052198. PubMed PMID: 23326326; PubMed Central
534 PMCID: PMCPMC3541408.
535 26. Andres-Terre M, McGuire H, Pouliot Y, Bongen E, Sweeney T, Tato C, et al.
536 Integrated, Multi-cohort Analysis Identifies Conserved Transcriptional Signatures
537 across Multiple Respiratory Viruses. 2015;43:1199-211. doi:
538 10.1016/j.immuni.2015.11.003.
539 27. Chen M, Carlson D, Zaas A, Woods CW, Ginsburg GS, Hero A, 3rd, et al.
540 Detection of viruses via statistical gene expression analysis. IEEE Trans Biomed Eng.
541 2011;58(3):468-79. Epub 2010/07/21. doi: 10.1109/TBME.2010.2059702. PubMed
542 PMID: 20643599.
543 28. Iwasaki A, Pillai PS. Innate immunity to influenza virus infection. Nat Rev

544 Immunol. 2014;14(5):315-28. Epub 2014/04/26. doi: 10.1038/nri3665. PubMed
545 PMID: 24762827; PubMed Central PMCID: PMCPMC4104278.
546 29. Bullard JH, Purdom E, Hansen KD, Dudoit S. Evaluation of statistical methods for
547 normalization and differential expression in mRNA-Seq experiments. BMC
548 Bioinformatics. 2010;11(1):94-106. doi: 10.1186/1471-2105-11-94.
549 30. Boggild AK, McGeer AJ. Laboratory diagnosis of 2009 H1N1 influenza A virus.
550 Critical care medicine. 2010;38:e38-e42.
551 31. Jackson GG, Dowling HF. Transmission of the common cold to volunteers under
552 controlled conditions.IV.spcific immunity to the common cold. Journal of Clinical
553 Investigation. 1959;38(5):762-9. PubMed PMID: PMC293221.
554 32. Liu TY, Burke T, Park LP, Woods CW, Zaas AK, Ginsburg GS, et al. An
555 individualized predictor of health and disease using paired reference and target
556 samples. BMC Bioinformatics. 2016;17:47. Epub 2016/01/24. doi: 10.1186/s12859-
557 016-0889-9. PubMed PMID: 26801061; PubMed Central PMCID: PMCPMC4722633.
558 33. Cybenko G. Approximation by superpositions of a sigmoidal function.
559 Mathematics of control, signals and systems. 1989;2(4):303-14.
560 34. Nair V, Hinton GE. Rectified linear units improve restricted boltzmann machines.
561 Proceedings of the 27th international conference on machine learning (ICML-
562 10)2010. p. 807-14.
563 35. Kingma DP, Ba LJ. Adam: A Method for Stochastic Optimization. International
564 Conference on Learning Representations (ICLR); 2015; San Diego, California, USA:
565 arXiv.org; 2015.
566 36. Zhang W. Shift-invariant pattern recognition neural network and its optical
567 architecture. Proceedings of annual conference of the Japan Society of Applied
568 Physics1988.
569 37. Iandola FN, Han S, Moskewicz MW, Ashraf K, Dally WJ, Keutzer K. SqueezeNet:
570 AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size.
571 International Conference on Learning Representations(ICLR); Palais des Congrès
572 Neptune, Toulon, France2017.
573 38. Szegedy C, Wei L, Yangqing J, Sermanet P, Reed S, Anguelov D, et al. Going
574 deeper with convolutions. IEEE Conference on Computer Vision and Pattern
575 Recognition (CVPR); 7-12 June 2015; Boston, Massachusetts, USA2015. p. 1-9.
576 39. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition.
577 Proceedings of the IEEE conference on computer vision and pattern recognition2016.
578 p. 770-8.
579 40. Ho TK. Random decision forests. Proceedings of 3rd international conference
580 on document analysis and recognition: IEEE; 1995. p. 278-82.
581 41. Chang KY, Yang JR. Analysis and prediction of highly effective antiviral peptides

based on random forests. PLoS One. 2013;8(8):e70166. Epub 2013/08/14. doi: 10.1371/journal.pone.0070166. PubMed PMID: 23940542; PubMed Central PMCID: PMCPMC3734225.

42. Ellerman D. Logical entropy: Introduction to classical and quantum logical information theory. Entropy. 2018;20(9):679.

43. Breiman L. Random Forests. Machine Learning. 2001;45(1):5-32. doi: 10.1023/A:1010933404324.

44. Cortes C, Vapnik V. Support-vector networks. Machine Learning. 1995;20(3):273-97. doi: 10.1007/bf00994018.

45. Vert J-P, Tsuda K, Schölkopf B. A primer on kernel methods. Kernel methods in computational biology. 2004;47:35-70.

46. Chang C-C, Lin C-J. LIBSVM: A library for support vector machines. ACM Trans Intell Syst Technol. 2011;2(3):1-27. doi: 10.1145/1961189.1961199.

47. Fawcett T. An introduction to ROC analysis. Pattern recognition letters. 2006;27(8):861-74.

48. Davis J, Goadrich M. The relationship between Precision-Recall and ROC curves. Proceedings of the 23rd international conference on Machine learning: ACM; 2006. p. 233-40.

49. Yu W, Park T. AucPR: an AUC-based approach using penalized regression for disease prediction with high-dimensional omics data. BMC genomics. 2014;15 Suppl 10(Suppl 10):S1. doi: 10.1186/1471-2164-15-S10-S1. PubMed PMID: 25559769.

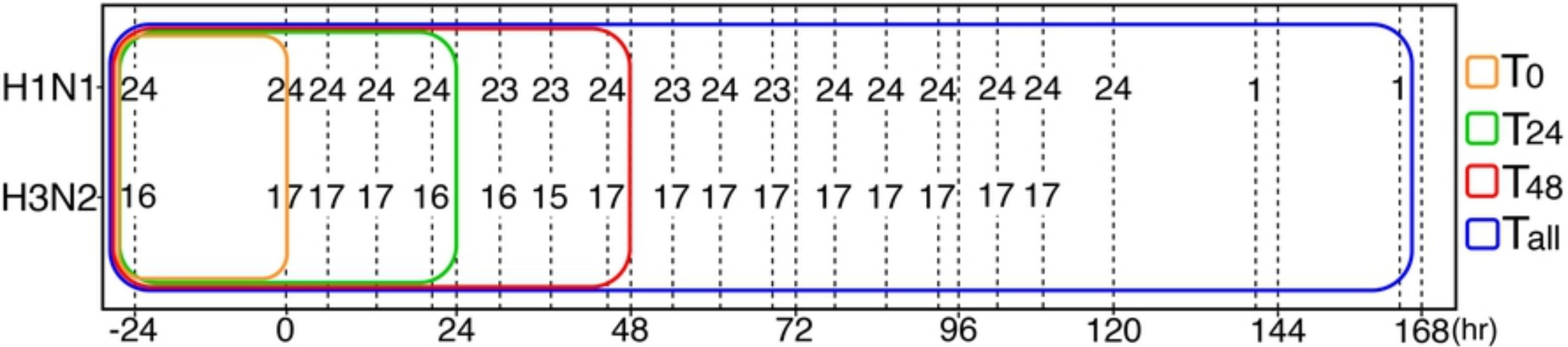

Figure 1. Data Frames

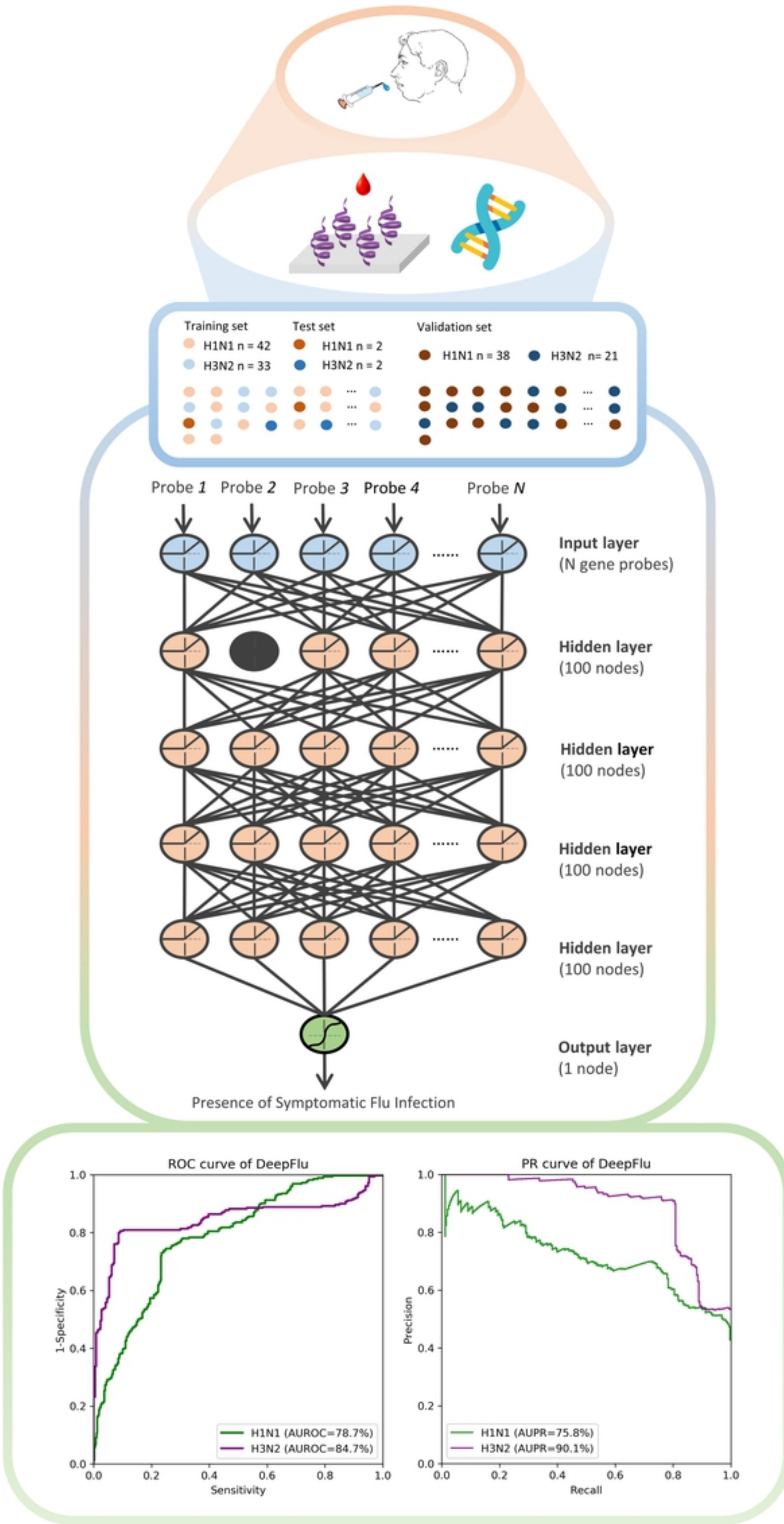

Figure 2. DeepFlu Layout