Design of Genetic Circuit for Monitoring and Treatment of Stroke

Xiang Wang\textsuperscript{a}, Qianqian Yu\textsuperscript{b}, Yu Si, Xiaocui Wang, Wei Shi, Tongsheng Xia and Hongge Li

School of Electronic and Information Engineering, Beihang University, Beijing 100191, China
\textsuperscript{a}wxiang@buaa.edu.cn, \textsuperscript{b}qianqianyu@buaa.edu.cn

Abstract. Nowadays, stroke has become a huge threat to human health and has become the first one of the top ten causes for death in the world. In the detection and diagnosis of stroke, optimization of markers and design of corresponding gene circuits have important research and application prospects. In this paper, we propose an improved Support Vector Model (SVM) by combining Principle Component Analysis (PCA) with Mutual Information (MI) to effectively figure out the proper stroke biomarkers with high specificity and sensitivity. The design of the single-index detection circuits is implemented. And the input circuit and critical circuit of the Single-index Detection Circuit are analysed. On the basis of the single-index detection circuit, a multi-marker detection circuit is built using OR GATE and AND GATE logic circuits to achieve the detection of multi-markers.

1. Introduction

Stroke, also known as “cerebral vascular accident”, is a group of diseases caused by sudden rupture of blood vessels in the brain or blood circulation disorder caused by vascular obstruction. Stroke has a serious impact on human life. One in six people in the world will have a stroke, one person dies from stroke every six seconds, and more than one in five survivors will have a stroke in 5 years [1]. At present, there are some difficulties in the methods of stroke detection, which make them unable to fully adapt to the current clinical tests [2].

The ideal regulatory protein, that is, the biomarkers should be characterized by high specificity and high sensitivity, and its expression level or blood level is positively correlated with the probability of disease. In order to improve the detection rate of the marker, it is necessary to select and combine multiple joint detection methods by means of joint detection of multiple markers, and select the “best marker detection group” applicable to a certain disease to improve the diagnosis efficiency [3].

The genetic circuit, in simple terms, is an artificially designed gene regulatory network. Regulatory protein can control the transcription and translation of genes, and the product of gene transcription and translation is protein. If this protein is also a kind of regulatory protein, it can continue to regulate other genes and eventually form a gene regulatory network [4]. By changing and adding regulatory proteins, structural genes, and adjusting their network structure, we can design genetic circuits that meet our expectations. Genetic circuit has made a great advance in the direction of unit circuits and cell communication. Yokobayashi et al. [5,6] conducted a preliminary study on the construction of cellular logic gates and successfully established a cell inverter, an AND gate and two other logic gates. Basu et al. [7] used a boundary detection circuit constructed by five non-gates and an AND gate to detect the AHL concentration in the environment. Attila Becskei et al. [8,9] made a special study on the stability of positive feedback loops and negative feedback loops in self-regulating systems in gene circuits.
Stability in gene expression refers to the tendency of the system to stay in a stable state (balance of protein production and degradation rate) in the event of external disturbances. Timothy et al. [10] focused on the study of some of the module functions in gene networks to synthesize the properties of the entire circuit. Some of these modules act as switches and oscillators, and others have the characteristics of logic gates. Tetsuya Kobayashi [11] also used E. coli as the experimental object, and used a positive feedback loop to synthesize the toggle switch, and further proved that the toggle switch circuit composed of the positive feedback loop has some very good properties.

The application of genetic circuits to stroke therapy fundamentally updates existing treatment mode and provides a powerful tool for stroke detection by accurately manipulating cellular behaviour. We can build specific gene circuits based on the biological characteristics of cells and download them to engineered cells to identify and track abnormal cells for early diagnosis of stroke. Compared with conventional methods, genetic circuits have the advantages of early detection, accurate diagnosis and precise treatment[12-15].

2. Optimization of Stroke Markers

2.1. Application of SVM in Detection of Stroke Markers

Support Vector Machine (SVM) has many unique advantages in solving small sample, nonlinear and high-dimensional pattern recognition, and can be applied to other machine learning problems such as function fitting [16]. This chapter proposes a method of combining principal component and mutual information analysis. In order to improve the SVM, firstly, the principle component analysis is used to eliminate the correlation between the input variables, and then the mutual information analysis is used to obtain the input variables most relevant to the output variation as the input variables of the model.

We collected 23 types of data from hospitalized patients with neurological treatment in Xuanwu Hospital from January 2015 to August 2015. The SVM was used to process the clinical data to obtain the best combination of markers to achieve identification and early diagnosis of stroke.

2.2. Mutual Information

The mutual information definition between the variables X and Y is the equation (1). When the variables X and Y are independent or completely independent of each other, the mutual information I(X; Y) equals to 0, that is, there is no commonly owned information between the two variables; when the variables X and Y are highly dependent on each other, the mutual information I(X; Y) will also be large, that is, there is much common information between the two variables.

\[
I(X;Y) = \sum_{i=1}^{v} \sum_{x,y} p(x,y) \log \left( \frac{p(x,y)}{p(x)p(y)} \right)
\]  

(1)

Multiple biomarkers \(x_1, x_2, x_3, \ldots\) are used as input variables and whether a person has stroke or not as variable Y whose value is 0 and 1. Then we calculate the mutual information between \(x_1, x_2, x_3, \ldots\) and Y. The results of mutual information analysis based on these 500 sets of artificial data are shown in Fig. 1. We chose 0.02 as the threshold for mutual information, and screened 12 biomarkers with high correlation with stroke.

![Figure 1. Mutual information between 23 biomarkers and stroke](image-url)
2.3. Principal Component Analysis

There is usually a complex interrelationship between 12 high-relevance factors screened out. Principal component analysis is a method that can convert multiple original variables into several linearly independent comprehensive variables.

Here are 12 random variables of 12-dimensional population variable \( x \). And their linear combination can constitute up to 12 comprehensive variables:

\[
\begin{align*}
    f_1 &= a_{11}x_1 + a_{12}x_2 + \cdots + a_{112}x_{12} = a_1^T x \\
    f_2 &= a_{21}x_1 + a_{22}x_2 + \cdots + a_{212}x_{12} = a_2^T x \\
    \vdots \\
    f_{12} &= a_{121}x_1 + a_{122}x_2 + \cdots + a_{1212}x_{12} = a_{12}^T x
\end{align*}
\]  

For each \( i (i = 1, 2, \ldots 12) \), the following normalization conditions should be met:

\[ a_i^T a_i = a_{i1}^2 + a_{i2}^2 + \cdots + a_{i12}^2 = 1 \]  

Mutually unrelated conditions, that is the covariance of different composite variables equals to 0:

\[ \text{cov}(f_j, f_i) = 0 (j < i, j = 1, 2, \ldots 11) \]  

Under the premise of satisfying these two conditions, the variance \( \text{var}(f_i) \) reaches the maximum, then the integrated variable \( f_i \) is the ith principal component of the population \( x \). The ratio of information used to measure the original independent variables contained in the principal component is:

\[ \gamma_i = \text{var}(f_i) / \sum_{k=1}^{12} \text{var}(f_k) = \lambda_i / \sum_{k=1}^{12} \lambda_k \]  

The variance contribution rate or contribution rate of the principal component is:

\[ \eta_m = \sum_{i=m}^{12} \gamma_i = \sum_{i=m}^{12} \lambda_i / \sum_{k=1}^{12} \lambda_k \]  

Often the size of \( \eta_m \) determines how many principal components should be used instead of the original variables. Usually, when the value of \( \eta_m \) reaches 70%-90%, the remaining principal components can be omitted, thus completing the function of dimensionality reduction. Table 1 shows the variance contribution rates for these twelve variables.

**Table 1. Cumulative variance contribution of each marker**

| Component number | Eigen value | Cumulative variance contribution/% |
|------------------|-------------|------------------------------------|
| 1                | 2.986       | 24.882                             |
| 2                | 2.253       | 43.653                             |
| 3                | 1.634       | 55.543                             |
| 4                | 1.354       | 65.978                             |
| 5                | 1.212       | 77.197                             |
| 6                | 1.010       | 86.828                             |
| 7                | 0.886       | 89.075                             |
| 8                | 0.730       | 91.133                             |
| 9                | 0.487       | 94.838                             |
| 10               | 0.281       | 97.179                             |
| 11               | 0.148       | 98.996                             |
| 12               | 0.120       | 100.000                            |
2.4. Support Vector Machine Model

In this chapter, we have taken a total of 200 sets of experimental data. The specific solution process is as follows:

1) We first define the training set and validation set.

**Training set:**

\[ S = \{(x_1, y_1), ..., (x_l, y_l)\} \in (X \times Y)^l \]

**Validation set:**

\[ T = \{(x_{l+1}, y_{l+1}), ..., (x_{l+n}, y_{l+n})\} \in (X \times Y)^n \]

\( x_i \in X \in \mathbb{R}^n, y_i \in Y = \{0, 1\} (i=1, 2, ..., n), n \) is the total medical case, and \( l \) is the number of cases as the verification set, so the number of cases as the verification set is \( n-l+1 \);

\( R^n \) represents the multidimensional space of the biomarker, \( x_i \) represents each of these biomarkers. \( y_i \) is a label to indicate whether or not a disease is present, 0 means no disease, and 1 means disease. In distinguishing between hemorrhagic stroke and ischemic stroke, \( y_i \) is converted to \( \{2, 3\} \), 2 represents hemorrhagic stroke while 3 represents ischemic stroke.

2) The input variables are then mapped into the high-dimensional Hilbert space \( H \) by nonlinear mapping \( \Phi: \mathbb{R}^n \rightarrow H \) using the RBF kernel function.

\[ K(x_i, x'_j) = \phi(x_i) \cdot \phi(x'_j) \]  

Define:

\[ K(x_i, x'_j) = e^{-\frac{|x_i - x'_j|^2}{2\sigma^2}} \] (7)

Then the objective function of the "maximum interval" nonlinear support vector machine becomes:

\[ w^* = \arg \min_{w} \left\{ \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{l+n} \xi_i \right\} \] (10)

\[ s.t. \begin{cases} \phi(x_i) + b \geq 1 - \xi_i \geq 0, \xi_i \geq 0, i = 1, 2, ..., l \end{cases} \]

Thus, the nonlinear support vector machine is the following optimization problem:

\[ w^* = \arg \min_{w} \left\{ \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{l+n} \xi_i \right\} \] (11)

\[ s.t. \begin{cases} \phi(x_i) + b \geq 1 - \xi_i \geq 0, \xi_i \geq 0, i = 1, 2, ..., l \end{cases} \]

The SVM model has two very important parameters \( C \) and \( \gamma \), \( C \) is the penalty parameter. The higher the \( C \) value, the more the error cannot be tolerated, so it is easy to over-fitting, but the smaller the \( C \) value, the easier it is to fit. Therefore, if the \( C \) value is too large or too small, the generalization ability will be deteriorated. \( \gamma \) is a parameter that comes with the function after selecting the RBF function as a kernel function (Kernel). Implicitly determines the distribution of multidimensional data after mapping to a new feature space. The larger its value, the less the support vector; the smaller the \( \gamma \) value, the more support vectors. The number of support vectors affects the speed of training and prediction.

Therefore, for the accuracy of prediction and the efficiency of time, we have to choose these two parameters. Regarding the optimization of SVM parameters, we use K-fold cross-validation to let \( c \) and \( g \) take values within a certain range, use the training set as the original data set and use the K-CV method to obtain the training set under different combinations of \( c \) and \( g \). Verify the classification
accuracy rate, and finally make the training set verify the group with the highest accuracy and the group c and g as the best parameters.

In K-fold cross-validation, the initial samples are segmented into K sub-samples, a single sub-sample is retained as data for the validation model, and the other K-1 samples are used for training. Cross-validation is repeated K times, each sub-sample is verified once, and the average K-time results or other combinations are used to finally obtain a single estimate. The advantage of this method is that it repeatedly uses randomly generated sub-samples for training and verification, and each time the results are verified once.

The first step in predicting stroke using SVM is to distinguish between stroke patients and normal people. The experimental results show that the PCA-MI-SVM can accurately distinguish between stroke patients and normal people.

![Figure 2. Distinguish between normal people and patients using SVM](image)

Once the first step of the test confirms the disease, SVM needs to further determine the type of stroke (ischemic stroke or hemorrhagic stroke). The results show that when judging the specific type of stroke, it is not guaranteed to be completely correct, but the accuracy rate remains above 80%.

![Figure 3. Determine the type of stroke using SVM](image)

3. Single-Index Detection Circuit

3.1. Design of Single-Index Detection Circuit

3.1.1 Structure of the inverter

Inverter is the core component of genetic circuits. This is mainly because the concept of "suppressor" in molecular biology has almost a one-to-one correspondence with the concept of "inverter" in steady state [17]. This is why the inverter can be one of the core components in the genetic circuit. Figure 4 describes its reaction process. The functional components of the inverter include transcription, regulatory binding, translation, degradation of mRNA and protein molecules [18].

![Figure 4. Schematic diagram of the biological process of the inverter](image)
3.1.2 Biochemical model and ibiosim model of single-index detection circuit

Biochemical reaction process is described by the chemical reaction equations, which represent the relationship between the formation and transformation of various substances. This is the basis for modelling with iBioSim. The biochemical reactions in living organisms include four types: transcription and translation, regulation of protein multimerization, regulation of protein binding, and deconstruction of various proteins and mRNAs. These four processes are reflected in the single-index detection module.

According to the structure of the inverter, the chemical reaction equations of the single-index detection circuit are established. There are several basic biochemical substances, which are input substance (mRNA\(_A\)), representative marker (rRNA, RNA\(_P\)), output substance (mRNA\(_Z\)) and protein (A, A\(_2\), A\(_4\), Pz, PzA\(_2\), PzA\(_4\)). Table 3 is the biochemical reaction equations of the single-index detection module.

As shown in Table 2, the left of the chemical equation represents the reactant, the right represents the product, the arrow represents the reaction process, and the reaction conditions and kinetic parameters are above the arrow. Different kinetic parameters lead to different reaction rates.

### Table 2. Biochemical reaction model of single-index detection circuit

| Reaction Equation | Kinetic Parameter |
|-------------------|------------------|
| \( P_Z A_4 \xrightarrow{k_{is}} P_Z A_2 + A_2 \) | (12) |
| \( mRNA_A \xrightarrow{k_{ec}} \) | (13) |
| \( A + A \xrightarrow{k_{sngl}} A_2 \) | (14) |
| \( A_2 \xrightarrow{k_{sngl}} A + A \) | (15) |
| \( A \xrightarrow{k_{ec}} \) | (16) |
| \( A_4 \xrightarrow{k_{ec}} \) | (17) |
| \( P_Z + A_2 \xrightarrow{k_{prs}} P_Z A_2 \) | (18) |
| \( P_Z A_2 \xrightarrow{k_{is}} P_Z + A_2 \) | (19) |
| \( P_Z A_2 + A_2 \xrightarrow{k_{prs}} P_Z A_4 \) | (20) |
| \( P_Z A_4 \xrightarrow{k_{is}} P_Z A_2 + A_2 \) | (21) |
| \( P_Z + RNA_p \xrightarrow{k_{scribe}} P_Z + RNA_p + mRNA_Z \) | (22) |
| \( mRNA_Z \xrightarrow{k_{ec}} \) | (23) |

Figure 5 shows a schematic diagram of the structure of the single-index detection circuit on the iBioSim platform.

Figure 5. IBioSim model of single-index detection circuit

It can be seen from Figure 6 that the output of the single-index detection circuit is consistent with the results of the reaction equation: when the input concentration reaches a high level, it will respond and the output mRNA\(_Z\) concentration will become low; when the input returns to a low level, the inverter output will eventually return to a high level.
3.2. Analysis of Single-Index Detection Circuit Module

3.2.1 Analysis of input module design

The input module is used for detecting the input signal and then transmitting the signal to other modules. In Figure 7, the dot R1 indicates that rRNA reacts with the input signal to generate protein A, R2 indicates the dimerization process of protein A and its reverse process, R3 and R4 indicate the degradation process of protein. The concentration level of rRNA directly affects the threshold level of the detection signal so that the steady-state response of the detection circuit would be influenced, as shown in Figure 8. Therefore, it is reasonable to select the suitable concentration level of rRNA according to the threshold requirement. If the rRNA concentration level is too high, the circuit will be too sensitive and susceptible to interference. On the contrary, too low rRNA concentration levels will cause the circuit to fail to properly detect the input signal.

3.2.2 Analysis of critical branch

The portion shown in Figure 9 is the core part of the single-index detection circuit. Its design idea is also derived from the inverter. It is divided into two branches: the suppression branch and the recovery branch. As shown in Figure 9, the red arrow indicates the main part of the suppression branch, and the green arrow indicates the main part of the recovery branch. The suppression branch reduces the Pz concentration mainly by the combination of A2 and Pz to generate PzA2. When the concentration of the input substance exceeds the threshold, A2 will be produced in large quantities, and Pz will greatly decrease, thereby suppressing the production of the output protein. Then, PzA2 will continue to react...
with A₂ to generate PzA₄. When the concentration of the input substance is lower than the threshold concentration, the concentration of A₂ will decrease. At this time, the recovery branch dominates position instead of the suppression branch. PzA₂ generates PzA₄ by decomposition reaction and reverts to Pz by PzA₂, and the output protein concentration increases. Thus, the signal is restored.

**Figure 9.** Schematic diagram of the critical path (red indicates the suppression path, green indicates the recovery path)

The dot R5 indicates the core reaction of the suppression branch. PzA₂ is generated by the combination of A₂ and Pz. When the concentration of A₂ is increased, the concentration of Pz is directly decreased. The production of the output protein is inhibited. It is worth noting that PzA₂ reacts with A₂ to form PzA₄ through R6 reaction. However, this reaction actually has little effect on Pz, because the Pz concentration is close to 0 when the PzA₂ concentration reaches a certain level. Therefore, the R6 reaction mainly affects the concentration level of PzA₂, and prepares for the next signal recovery.

**Figure 10.** Results of the change of the main substances in the branch (K of R5 = 66.67)

Under the condition that the input material concentration is constant, the simulation results are consistent with the results discussed above. The decrease in the concentration of Pz is directly affected by the A₂ protein. PZ react with A₂ to generates PzA₂ and PzA₂ continues to react to generate PzA₄, which results in a low level of PzA₂ and a high level of PzA₄ shown in Figure 10.

**Figure 11.** Comparison of effects of R7 on the output

The recovery branch has the key reactions includes R6, R7 and R8. Since the reaction paths of R7 and R8 dominate the recovery branch, this paper mainly discuss the influence of the dynamic
parameters K of R7 and R8 on the recovery branch, and the affect of the output signal recovery time of the entire detection circuit. The result is shown in Figure 11.

Recovery time is defined as the time that the signal to rise from a low level to a steady state value. After comparative analysis, adjust the kinetic parameters of the PzA4 to PzA2 reaction (R7 in the Figure 9): 1) when k=0.025, recovery time is $\Delta t=6200s$; 2) when k=0.25, recovery time is $\Delta t=3800s$; 3) when k=2.5, recovery time is $\Delta t=3000s$. A comparative analysis of the changes in these three sampling points leads to preliminary conclusions: 1) the change of the k value does not affect the amplitude of the output, only affecting the recovery time; 2) an increase in k causes the inverter recovery time to decrease; 3) recovery time decreases with the increase of K.

The result indicates the kinetic parameters K of reaction R8: 1) k=2.887, $\Delta t=3700s$, steady state value is $1.2e-6\mu$Mole; 2) k=28.87, $\Delta t=3000s$, steady state value is $1.5e-6\mu$Mole. We can conclude that: 1) the change of the value of K affects the magnitude of the output, but the magnitude of the change is small; 2) the K value also affects the recovery time, when the value of K increases, the recovery time decreases; 3) the amplitude and recovery time will decrease with the increase of K value.

![Figure 12](image.png)

**Figure 12.** Comparison of the effects of R8 on the output

4. Design of Multi-index Detection Circuit

4.1. Design Ideas

The overall idea of multi-index detection circuit design is to design a single-index detection circuit for each brain stroke marker, and then connect each single-index detection circuit through a logic fusion circuit, so that the circuit tested the multiple brain stroke indicators.

Due to the complexity of stroke, "parallel detection method" and "joint detection method" are used to design the multi-index detection circuit. "Parallel detection method" means that when any one of the concentrations of the substances detected is greater than or equal to a threshold value, an early warning signal will output. This method can improve the sensitivity of the detection. "Joint detection method" means that a positive warning signal will output when concentrations of all the substances detected are greater than or equal to the threshold value. And this method improves the specificity. In this paper, two kinds of logic OR Gates are constructed to achieve parallel detection, and logic AND Gate is constructed to achieve joint detection.

4.2. Construction of Logic OR Gate

4.2.1 Construction methods of OR Gate

The first construction method of OR Gate is:

$$\text{OR}(X, Y)=\text{NAND}(\text{NOT}(X), \text{NOT}(Y))$$  \hspace{1cm} (24)

Figure 13 shows the mechanism of action of the first method: protein X and Y bind to their respective promoters $P_X$ and $P_Y$, thereby inhibiting the expression of subsequent proteins NOT (X) and NOT (Y). We named NOT(X) as Z1 and NOT(Y) as Z2.
Any one of Z1 or Z2 does not act, it cannot bind to Pg and inhibit the transcription process of the fluorescent protein gene. The production of two key NOT proteins is regulated by proteins X and Y. When the concentration of any of the protein X or Y rises, the concentration of NOT(X) or NOT(Y) will decrease and eventually leads to the expression of fluorescent proteins and high concentrations of output.

Figure 13. Composition principle of OR Gate-1

The second construction method of OR Gate is:

\[ \text{OR}(X, Y) = \text{NOT} (\text{NOT} (\text{OR}(X, Y))) \]  

(25)

Unlike the first method, either X or Y can bind to the promoter Pxy to achieve the purpose of inhibiting the transcription process of the corresponding gene fragment.

Figure 14. Composition principle of OR Gate-2

### 4.2.2 Biochemical reaction model of OR Gate

The biochemical reaction models and equations of OR GATE-1 and OR GATE-2 are shown in Tables 3 and 4. The K above the equal sign of the chemical reaction equation represents different kinetic parameters; obviously, the chemical rates of the different reactions are different, and the lower corners of the representative K values are also different.

**Table 3. Biochemical reaction model of OR GATE-1**

| Reaction Equation | (Equation) |
|-------------------|------------|
| \( X + P_x \xrightarrow{k_{x\text{prom}}^{+}} P_x X \) | (26) |
| \( Y + P_y \xrightarrow{k_{y\text{prom}}^{+}} P_y Y \) | (27) |
| \( P_x + \text{RNA}_{\text{Ap}} \xrightarrow{k_{\text{prom}}^{x}} P_x + \text{RNA}_{\text{Ap}} + m\text{RNA}_1 \) | (28) |
| \( m\text{RNA}_1 + r\text{RNA} \xrightarrow{k_{s\text{prom}}^{x}} m\text{RNA}_2 + r\text{RNA} + Z_1 \) | (29) |
| \( P_y + \text{RNA}_{\text{Ap}} \xrightarrow{k_{\text{prom}}^{y}} P_y + \text{RNA}_{\text{Ap}} + m\text{RNA}_2 \) | (30) |
| \( m\text{RNA}_2 + r\text{RNA} \xrightarrow{k_{s\text{prom}}^{y}} m\text{RNA}_3 + r\text{RNA} + Z_2 \) | (31) |
| \( Z_1 + Z_2 + P_y \xrightarrow{k_{s\text{prom}}^{z}} P_y Z \) | (32) |
| \( P_y + \text{RNA}_{\text{Ap}} \xrightarrow{k_{\text{prom}}^{y}} P_y + \text{RNA}_{\text{Ap}} + m\text{RNA}_y \) | (33) |
| \( m\text{RNA}_y + r\text{RNA} \xrightarrow{k_{s\text{prom}}^{y}} m\text{RNA}_y + r\text{RNA} + XFP \) | (34) |
Table 4. Biochemical reaction model of OR Gate-2

\[
\begin{align*}
X + P_{XY} & \xrightarrow{k_{reps}} P_{XY} X \\
Y + P_{XY} & \xrightarrow{k_{reps}} P_{XY} Y \\
P_{XY} + RNAp & \xrightarrow{k_{ts}} P_{XY} + RNAp + mRNA_z \\
mRNA_x + rRNA & \xrightarrow{k_{ts}} mRNA_x + rRNA + Z \\
Z + P_t & \xrightarrow{k_{reps}} PZ \\
P_t + RNAp & \xrightarrow{k_{ts}} P_t + RNAp + mRNA_f \\
mRNA_f + rRNA & \xrightarrow{k_{ts}} mRNA_f + rRNA + XFP 
\end{align*}
\]

(35) \quad (36) \quad (37) \quad (38) \quad (39) \quad (40) \quad (41)

4.2.3 Simulation of OR Gate in iBioSim

The iBioSim circuit diagram of OR Gate was built. Since the reaction situation is quite complicated and cannot be completely simulated, here is a simplified schematic diagram in which the concentration of the substance is a relative value and does not indicate the actual situation.

1) Simulation of OR Gate-1

Figure 15 is a schematic diagram of OR Gate-1. The simulation results are shown in Figure 16. It can be seen that during the period of 20-80, at least one of X and Y maintain at high concentration; although there is a certain lag, the output is maintained at a high level during the period of 40-100. It should be pointed out that there is a pulse in the output curve at the beginning of the simulation, because the initial output is less suppressed, resulting in the expression of fluorescent protein, but it will rapidly decrease to a low level as the reaction occurs.

![IBioSim model of OR Gate-1](image)

**Figure 15.** IBioSim model of OR Gate-1

![Simulation results of OR Gate-1](image)

**Figure 16.** Simulation results of OR Gate-1

a) Input of OR Gate-1 b) Output of Fluorescent protein

2) Simulation of OR Gate-2

Figure 17 is a schematic diagram of OR Gate-2. The simulation results are shown in Figure 18. During the period of 20-80, at least one of the two input variables X and Y keeps at a high concentration, and the output rises to a high level within 30-90.
3) Comparison of simulation results
Comparing the two simulation images of Figure 16 and Figure 18 above can lead to the following conclusions: 1) The simulation curve of OR Gate-2 does not have the kind of pulse that appears in the simulation image of OR Gate-1. 2) The lag time of OR Gate-2 is less than the lag time of OR Gate-1. Therefore, it is recommended to use the second method to construct an “OR Gate” logic circuit.

4.3. Construction of Logic AND Gate

4.3.1 Construction methods of AND Gate
The construction method of AND Gate is as shown in Equation 40, and Figure 19 illustrates the specific structure composition.

\[
\text{AND}(X, Y) = \text{NOT} (\text{NAND}(X, Y))
\] (42)

Any one of protein X and Y can bind to the promoter to inhibit the production of protein R; R can bind to the promoter Pz to suppress the expression of Z. Only when X and Y are all at high concentration, protein R will be at a low concentration level, thereby reducing the inhibition of the expression of the output protein and increasing the concentration of protein Z. At this point, the logical function of "AND gate" is realized.

4.3.2 Biochemical reaction model of AND Gate
The simplified biochemical reaction model for constructing AND GATE is shown in Table 5. The equation includes processes in which the protein binds to promoters to generate inhibitory products, and the corresponding fluorescent protein is transcribed and translated.
Table 5. Biochemical reaction model of AND Gate

\[
X + P_x \xrightarrow{\text{lysP}} P_x X \\
Y + P_y \xrightarrow{\text{lysP}} P_y Y \\
P_x + \text{RNAp} \xrightarrow{\text{ktx}} P_x + \text{RNAp} + \text{mRNA}_1 \\
\text{mRNA}_1 + r\text{RNA} \xrightarrow{\text{ktx}} \text{mRNA}_2 + r\text{RNA} + Z_1 \\
P_y + \text{RNAp} \xrightarrow{\text{ktx}} P_y + \text{RNAp} + \text{mRNA}_2 \\
\text{mRNA}_2 + r\text{RNA} \xrightarrow{\text{ktx}} \text{mRNA}_3 + r\text{RNA} + Z_2 \\
Z_1 + P_y \xrightarrow{\text{lysP}} P/Z_3 \\
Z_2 + P_y \xrightarrow{\text{lysP}} P_y Z_2 \\
P_y + \text{RNAp} \xrightarrow{\text{ktx}} P_y + \text{RNAp} + \text{mRNA}_4 \\
\text{mRNA}_4 + r\text{RNA} \xrightarrow{\text{ktx}} \text{mRNA}_5 + r\text{RNA} + XFP
\]

(43)  (44)  (45)  (46)  (47)  (48)  (49)  (50)  (51)  (52)

4.3.3 Simulation of AND Gate in iBioSim

Figure 20 shows the circuit diagram of AND Gate. The simulation results are shown in Figure 21. During the period of 40-90, both the two input X and Y maintain at a high concentration; and the output rises to a high level. The function of logic AND Gate is realized.

Figure 20. iBioSim model of AND Gate

Figure 21. Simulation results of AND Gate

5. Conclusion

Results show that the PCA-MI-SVM model can achieve the optimal selection of stroke markers and disease prediction. Analysis of the principal components and mutual information can filter the principal components that have higher correlation with stroke, and then SVM can realize the prediction of stroke. Single-index detection module can be designed by setting up the biochemical
reaction model and the differential equation model for the single index detection circuit. IBioSim is used to model and simulate the circuit structure. The multi-marker detection circuit can achieve the detection of multi-markers, joint detection can realize by AND GATE while parallel detection can be made by OR GATE.

6. Acknowledgments
This research is supported by the National Science Foundation of China (Grant No. 81571142, and No. 60973106), the Key Project of National Science Foundation of China (Grant No. 61232009), National High-tech R&D Project of China (863 Grant No. 2011AA010404).

7. References
[1] Thrift AG, Cadilhac DA, Thyabaranathan T, Howard G, Howard VJ, Rothwell PM, Donnan GA. Global stroke statistics. International Journal of Stroke, 2014, 9 (1): 6-18
[2] Feigin VL, Wang W, Fu H, et al. Primary stroke prevention in China – a new approach. Neurological Research, 2015, 1743132815Y. 0000000025.
[3] Wang Likun, Wang Fan, Wu Guofeng, Early-stage minimally invasive procedures decrease perihematomal endothelin–I levels and improve
[4] Juan Li, Bo Yao, Huang Huang, Zhao Wang, Changhong Sun, Yu Fan, Qing Chang, Shaolu Li, Xiang Wang, Jianzhong Xi, Real-Time PCR MicroRNAs Detection Based on Enzymatic double stem-loops Ligation, Analytical Chemistry, JUL 1, 2009, 81(13): 5446-5451.
[5] Yokobayashi Y, Weiss R, Basu S, et al. Evolutionary design of genetic circuits and cell-cell communication [J]. Adv Comp Systems, 2003, 9(3): 25-32.
[6] Basu S, Karig D, Weiss R., Engineering Signal Processing in Cells: Towards Molecular Concentration Band Detection, Natural Computing, 2003, 2: 463-478.
[7] Arnold FH, Yokobayashi Y, Weiss R. Directed evolution of a genetic circuit[J]. Proceedings of the National Academy of Science of the United States of America, 2002, 99(26): 16587-16591.
[8] Attila Becskei, Luis Serrano. Engineering stability in gene networks by autoregulation [J]. Nature, 2000, 405(6786): 590-593.
[9] Timothy S. Gardner, Charles R. Cantor, James J. Collins. Construction of a genetic toggle switch in Escherichia coli [J]. Nature, 2000, 403(6767): 339-342.
[10] Tetsuya Kobayashi, Luonan Chen. Gene toggle switches only with positive feedback loops [C]. Genome Informatics, 2001, 12: 294-295.
[11] Adams HPJ, Bendixen BH, Kappelle LJ, Biller J, Love BB, Gordon DL, Marsh EE 3rd. Classification of subtype of acute ischemic stroke: definitions for use in a multicenter clinical trial, TOAST. Trial of Org 10172 in Acute Stroke Treatment, Stroke, 1993, 24: 35-41.
[12] WANG Xiang, YUAN GuangQian, WANG Xun, ZHAO ZeXi, WANG Kan. Communication and Monitor of Breast Cancer Signal in the Pulse-Output Genetic Circuit Network[J], SCIENCE CHINA (Information Sciences), March, 2014, 57(3): (032105)1-10.
[13] Bonnet J, Yin P, Ortiz ME, Subsoontorn P, Endy D, Amplifying Genetic Logic Gates, Science, May 3, 2013, 340: 599-603.
[14] Wang X, Yuan G Q, Wang X, Zhao Z X. Pulse-Output Monitor Genetic Circuit of Breast Cancer Testing, 2013 IEEE International Conference on GreenCom-iThings-CPSCom., Beijing, China, Aug 20-23, 2013, 317: 1722-1728.
[15] Padilla P, Lopez M, Gorriz J M, et al. NMV-SVM based CAD tool applied to functional brain images for the diagnosis of Alzheimer's disease [J]. IEEE Transactions on Medical Imaging, 2012, 31(2):207-16.
[16] Wang, X., Shi, W., Wang, X. C., & Wang, T., Optimised Selection of Stroke Biomarker Based on SVM and Information Theory. In ITM Web of Conferences (ISSN2271-2097, Vol. 12, 05013). EDP Sciences. ITA 2017, 712-717.
[17] Zhao W, Chen H, Xiang L, et al. Passive acoustic detection of diver based on SVM[C], IEEE International Conference on Mechatronics and Automation. 2016.
[18] Ron Weiss, Sabbayu Basu, et al. Genetic Circuit Building Blocks for Cellular Computational [J]. Natural Computing, 2003: 1-40.