In Silico Analysis Using Hybrid Support Vector Machine and Second Order of Markov Chain for Multiple Sequence Alignment to Identify the Types of Leukaemia

Mohammad Isa Irawan 2Mohammad Hamim Zajuli Al Faroby and 3Awik Pudji Dyah Nurhayati

1,2 Department of Mathematics, Faculty of Mathematics, Computing and Data Science, Institut Teknologi Sepuluh Nopember. 3 Department of Biology, Faculty of Science, Institut Teknologi Sepuluh Nopember

mii@its.ac.id 2alfa.crash@gmail.com, 3awikpudji@biology.its.ac.id

Abstract. Cancer is a disease that causes an abnormal growth of cells and can attack every part of the body, which is occurred because of a damage in deoxyribonucleic acid (DNA) that leads to a mutation in a vital gene that controls cell division. The biomarker technology that used in clinical practice still used a high cost and need a long time to detect the cancer signs. As the former studies about cancer, the biomarker has been detected in the microarray data. In this paper, we used a support vector machine (SVM) to classify 4 type of leukaemia. Begin with extracting the data feature of sequence DNA from a string into numeric using Second order of Markov chain, SVM classified DNA using 40 data for the training step and 25 data for testing step. In this paper, SVM used 3 types of the kernel, which are linear, Gaussian radial basis function, and polynomial. The results showed that the Gaussian kernel has the best accuracy then other kernel.

1. Introduction
Cancer is terms for a big group of disease that can influence every part of the body. One of the features to define cancer as abnormal growth of the new cells that exceeding normal growth and they can attack parts of the body and spread to other organ, this process called metastasis. Cancer grows if gens in a normal cell mutate or have a genetic problem. Uncontrolled growth that caused by the damage in deoxyribonucleic acid (DNA) chain code, leads to a mutation in vital genes that controls cell division [1].

Technology that used for biomarker in clinic practice didn’t strong enough to detect the cancer signs, because of the high cost and need a long time for detecting the signs. That’s why microarray detections were needed to detect the signs [2]. Traditional analytic techniques, like hierarchical clustering, assume the biological linearity and use a statistical approach to conclude the connection for each gen (feature selection), but these techniques didn’t work well when the datasets contaminated with noises. In order the noise-contaminated datasets do not significantly affect the results of feature selection, we used a machine learning method. As part of the machine learning methods support vector machine (SVM) and artificial neural networks (ANN) has been succeed to analysis gene expression [3]. The success in analysing needs an identification of a gene that related to the cancer class and removal of unused variables. A bad analysis will cause data overfitting and unidentified result [4].
Another machine learning as well probabilistic neural network (PNN) with the first order and second order Markov chain to classify the DNA sequence of a bacterium, found that using second-order Markov chain, gave a better result in classification [6]. Therefore, this paper used SVM to classify the DNA sequence of human that mutated because of cancer, focused in Acute Lymphoblastic Leukaemia (ALL), Acute Myeloid Leukaemia (AML), Chronic Lymphocytic Leukaemia (CLL) and Chronic Myeloid Leukaemia (CML). This paper created an application that detects the type of blood cancer by classifying the DNA into 4 type of blood cancer [5].

2. Materials and Methods

2.1. Blood Cancer

Cancer is a disease caused by an abnormal growth of cells in body tissues that mutate into a cancer cell. Cancer cells can spread to other parts of the body and can lead to death. Leukaemia is a blood cancer that starts from bone marrow, the place where blood cell produced. Human blood contains red blood cells, white blood cells, and platelets. In Leukaemia case, the bone marrow produces leukemic cells, white blood cells that have not fully developed yet. These leukemic cells do not work like white blood cells, but it swarms the healthy cells. Leukaemia can be acute (deteriorate rapidly) or chronic (deteriorate slowly). Leukaemia divided into 4 types, i.e. Acute Lymphoblastic Leukaemia (ALL), Acute Myeloid Leukaemia (AML), Chronic Myeloid Leukaemia (CML), and Chronic Lymphocytic Leukaemia (CLL). Basically, there is others type beside of that four, but it rarely occurred. These four types of leukaemia are so common in the world. Therefore, the researchers are focused on that four types.

2.2. Deoxyribonucleic Acid

Deoxyribonucleic acid also has known as DNA, a kind of bio molecule that encodes genetic instructions in every organism. Generally, DNA consists of two biopolymer strands that twisting, creating a double helix form. The two strands are known as polynucleotide because they consist of individually nucleotide molecules. Every nucleotide consisting of one nitrogenous base, it can be Adenine (A), Cytosine (C), Guanine (G), and Thymine (T). Two strands of DNA are anti-parallel, which means that both are paired in opposite way. Every sugar cluster bonded one of 4 types of nitrogenous base. Sequences of the nitrogenous base in the back of DNA that save the biologic information.

2.3. Markov Chain

Markov chain is a stochastic method for determining the corresponding description state so that the integrated process will have Markovian property, a knowledge of a state to predict the future stochastic. If Markov state is a limited or countable set then the Markov chain is called discrete Markov chain. The value of state in first-order Markov chain only depends on the one previous period, and in $m – th$ order Markov chain, the state value is depending on $m$ previous periods. The first order Markov chain can be formulated as below:

$$P(X_i(i + 1) | X_i) \quad i = 1,2,3, ..., n$$

where,

- $X_{i+1}$ = state in $(i + 1)$-th position
- $X_i$ = state in $i$-th position that affect the value of $X_{i+1}$
- $P(X_{i+1}|X_i)$ = Probability of $X_{i+1}$ with $X_i$ as the state before.
The m-th order Markov chain is formulated as in equation (1) below:
\[ P[X_{n+1} = j | X_{(n+1)−m} = i_1, X_{(n+1)−(m+1)} = i_2, \ldots, X_n = i_n] \quad n, m = 1, 2, 3, \ldots \] (2)
the transition probability matrix \( P \), of which each element is \( p_{ij} \), as in the form of the matrix below, in equation (2):
\[
P = \begin{bmatrix}
p_{11} & p_{12} & \cdots & p_{1n} \\
p_{21} & p_{22} & \cdots & p_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
p_{n1} & p_{n2} & \cdots & p_{nn}
\end{bmatrix}
\] (3)

The transition probability matrix is formed from the probability of the previous occurrences of a particular base with a particular base. The bases are referred to as DNA bases, i.e. Adenine (A), Thymine (T), Guanine (G), and Cytosine (C). Therefore, the transition matrix of first-order Markov Chain consists of \( P(A|A) \), the probability of the occurrence of Adenine with the previous state is Adenine, \( P(C|A) \), the probability of the occurrence of Cytosine with the previous state is Adenine, and so on. Thus, the transition matrix of the first order Markov Chain consists of 16 elements [7], written as follow:
\[
P = \begin{bmatrix}
P(A|A) & P(C|A) & P(G|A) & P(T|A) \\
P(A|C) & P(C|C) & P(G|C) & P(T|C) \\
P(A|G) & P(C|G) & P(G|G) & P(T|G) \\
P(A|T) & P(C|T) & P(G|T) & P(T|T)
\end{bmatrix}
\] (4)

2.4 Support Vector Machine

Support vector machine was first introduced by Vapnik in 1992 as superior concepts in the field of pattern recognition. As a method of pattern recognition, SVM is still relatively new. However, the evaluation ability of SVM in various applications places him as state of the art in pattern recognition. The advantages of the support vector machine are effective in high-dimensional space, it is still effective if the number of dimensions is greater than the number of samples, using a subset of the training points on the decision function, so the memory used is more efficient and versatile, because different kernel functions can be decisive for decision-making functions. But, the downsides of the support vector machine are if the number of features is greater than the number of samples, it cannot implement kernel function and it is very important to use regularization, it does not directly provide estimate probability, the probability is calculated using validation.

Support vector classification (SVC) is an SVM that used for classification based on mathematical equations. The SVC concept can be explained simply as a search for the best hyperplane that serves as a separator of two classes in the input space. Figure 1(a) shows several patterns that which are the member of two classes: +1 and −1 and shows various alternatives of discrimination boundaries. Measuring the hyperplane margins and finding the maximum point will lead to finding the best dividing hyperplane. The margin is the distance between the hyperplane and the nearest pattern of each class, and the support vector is the closest pattern. The solid line in Figure 1(b) shows the best hyperplane, which is located right in the middle of the two classes, while the support vectors are the red and yellow dots that are in the black circle. The effort to locate the hyperplane is the main subject in the learning process in SVC.
Figure 1. SVM’s try to find the best hyperplane which divides data into two class, -1 and +1.

In hard-margin support vector machines, the training data are linear separable. In real problem the training data is linearly inseparable. So, hard-margin SVM is unsolvable. To allow inseparability, let $\xi_i \geq 0$ is slack variable, Figure 2 [8]:

$$y_i(w^T x_i + b) \geq 1 - \xi_i \quad \text{for} \quad i = 1, 2, ..., M \quad (5)$$

Figure 2. Inseparable case in a two-dimensional space

By the slack variable $\xi_i$, feasible solution always existing for the training data $x_i$, if $0 < \xi_i < 1$ (fig 2), the data do not have maximum margin but still correctly classified. But if $\xi_i \geq 1$ (fig 2), the data are misclassified by the optimal hyperplane. To obtain optimal hyperplane in which the number of training data that do not have maximum margin is minimizing,

$$Q(w) = \sum_{i=1}^{m} \theta(\xi_i) \quad (6)$$

where

$$\theta(\xi_i) = \begin{cases} 1, & \xi_i > 0 \\ 0, & \xi_i = 0 \end{cases}$$
that equation is a combinatorial optimization problem and difficult to solve. So, we consider minimizing,

$$Q(w, b, \xi) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{m} \xi_i^p$$  \hspace{1cm} (7)$$

subject to the constrains

$$y_i(w^T x_i + b) \geq 1 - \xi_i \quad \text{for } i = 1, 2, ..., m$$

Where $\xi = (\xi_1, ..., \xi_m)^T$ and $C$ is the margin parameter that determines the trade-off between the maximization of the margin and minimization of classification error. If $p = 1$ called L1 support vector machine (L1SVM) and when $p = 2$ called L2 support vector machine (L2SVM).

The optimal hyperplane in support vector machine is determined to maximize the generalization ability. The training data are not linearly separable, the obtained classifier may not have high generalization ability although hyperplanes are determined optimally. Thus to enhance linear separability, the original input space is mapped into high-dimensional dot-product space called kernel trick (Hsu & Lin, 2002).

**Table 1. Kernels in SVM**

| Kernel Type            | Definition                                      |
|------------------------|-------------------------------------------------|
| Linear                 | $K(x_i, x_j) = (x_i, x_j)$                      |
| Polynomial             | $K(x_i, x_j) = (y(x_i, x_j) + n)^d$             |
| Gaussian Radial Basic Function | $K(x_i, x_j) = e^{-\gamma \|x_i - x_j\|^2}$ |
| Sigmoid                | $K(x_i, x_j) = \tanh (y(x_i, x_j) + \beta)$   |

Using the kernel, the dual problem in the feature space is given as follows.

Maximize

$$Q(a) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} \alpha_i \alpha_j y_i y_j K(x_i, x_j)$$ \hspace{1cm} (8)$$

Subject to constrain,

$$\sum_{i=1}^{m} y_i \alpha_i = 0, \quad 0 \leq \alpha_i \leq C, \quad \text{for } i = 1, 2, ..., m$$

Because $K(x_i, x_j)$ is a positive semidefinite kernel, the optimization problem is a concave quadratic programming problem. And because $a = 0$ is a feasible solution, the problem has the global optimum solution. So, the decision function is,
\[ \sum_{i=1}^{m} y_i \alpha_i K(x_i, x_j) + b \]  

(9)

The performance of the Support Vector Machine based classifier was evaluated in term of three parameters namely accuracy, precision and recall ability.

\[ \text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \]  

(10)

\[ \text{Precision} = \frac{TP}{TP + FP} \]  

(11)

\[ \text{Recall} = \frac{TP}{TP + FN} \]  

(12)

where, \(TP\) is true positive that mean the data in positive class was classified by SVM on correct class. \(TN\) is true negative that mean the data in negative class is classified by SVM on correct class. \(FP\) is false positive, the data in positive class was classified on incorrect class. And the last one, \(FN\) is false negative, the data in negative class was classified on incorrect class.

### 3. Result and Discussion

#### 3.1. Feature Extraction

Feature Extraction will create the feature or mathematical model from the sequence, which will generate a matrix or vector for classification method. Feature Extraction in this paper used Second Order Markov Chain, where the feature extraction was taken from the probability of 2 nitrogenous bases that come after a single nitrogenous base. The matrix used in the second-order Markov chain is:

\[
M = \begin{bmatrix}
P(AA|A) & P(AA|C) & P(AA|G) & P(AA|T) \\
P(AC|A) & P(AC|C) & P(AC|G) & P(AC|T) \\
P(AG|A) & P(AG|C) & P(AG|G) & P(AG|T) \\
P(AT|A) & P(AT|C) & P(AT|G) & P(AT|T) \\
P(CA|A) & P(CA|C) & P(CA|G) & P(CA|T) \\
P(CC|A) & P(CC|C) & P(CC|G) & P(CC|T) \\
P(CG|A) & P(CG|C) & P(CG|G) & P(CG|T) \\
P(CT|A) & P(CT|C) & P(CT|G) & P(CT|T) \\
P(GA|A) & P(GA|C) & P(GA|G) & P(GA|T) \\
P(GC|A) & P(GC|C) & P(GC|G) & P(GC|T) \\
P(GG|A) & P(GG|C) & P(GG|G) & P(GG|T) \\
P(GT|A) & P(GT|C) & P(GT|G) & P(GT|T) \\
P(TA|A) & P(TA|C) & P(TA|G) & P(TA|T) \\
P(TC|A) & P(TC|C) & P(TC|G) & P(TC|T) \\
P(TG|A) & P(TG|C) & P(TG|G) & P(TG|T) \\
P(TT|A) & P(TT|C) & P(TT|G) & P(TT|T)
\end{bmatrix}^T
\]  

(13)
Sum of each row is 1 due to the Markovian properties. $P(AA|A)$ is the probability of nitrogenous bases AA occur after A and so on Fig. 3 shows the output of a matrix transition of Markovian.

$$
\begin{bmatrix}
0.61169591 & 0.34481703 & 0.02662722 & 0.0078125 \\
0.04995867 & 0.3921969 & 0.04828444 & 0.33125 \\
0.087193 & 0.07002601 & 0.09171598 & 0.0678125 \\
0.02339181 & 0.02240696 & 0.0255858 & 0.0360255 \\
0.0645274 & 0.07282013 & 0.06521392 & 0.0646875 \\
0.087193 & 0.10406226 & 0.15976331 & 0.140625 \\
0.0645274 & 0.10644258 & 0.11242654 & 0.0925 \\
0.03508577 & 0.06442577 & 0.06508876 & 0.06875 \\
0.0938775 & 0.0481703 & 0.06504734 & 0.10375 \\
0.20467896 & 0.10644258 & 0.12139178 & 0.1484275 \\
0.087193 & 0.10644258 & 0.1055893 & 0.1766875 \\
0.03508577 & 0.00645335 & 0.03550296 & 0.0360255 \\
0.00584795 & 0.0140056 & 0.00591716 & 0.0234275 \\
0.05263165 & 0.06942017 & 0.0147929 & 0.0546875 \\
0.0999125 & 0.07002601 & 0.0443787 & 0.03125 \\
0.05847955 & 0.02512008 & 0.01384312 & 0.0234275 \\
\end{bmatrix}
$$

**Figure 3. Feature of DNA Leukaemia**

### 3.2. Training Data

Support Vector Classification (SVC) training was the main point of the research. In this process, training data were processed to make the vector. Training data contains 40 sample data, which were sequences of 10 Acute Myeloid Leukaemia (AML), 10 Acute Lymphocytic Leukaemia (ALL), 10 Chronic Myelocytic Leukaemia (CML), and 10 Chronic Lymphocytic Leukaemia (CLL).

First, SVC trained the training data into acute and chronic classes, and then, the acute class was divided into ALL and AML, and the chronic class was divided into CLL and CML.

![Plotting Training Data into Acute and Chronic](image)

**Figure 4. Plotting Training Data into Acute and Chronic**

Figure 4. shows the division of training data, acute and chronic. Blue crosses represent acute data and red squares represent chronic data. Principle Component Analysis (PCA) method used to compres the multidimensionl data into 2-dimensional in order to make it easy get view of training data, but still has the feature of the data.
Figure 5. Plotting Acute Training Data into AML and ALL

Figure 5 shows that data in acute class are divided into 2 subclasses, AML and ALL. The blue dots represent data in ALL and the red squares represent data in AML.

Figure 6. Plotting Chronic Training Data into CML and CLL

Figure 6. shows that the data in chronic class are divided into 2 subclasses, CML and CLL. The blue dots represent data in CLL and the red squares represent CML.

3.3. Support Vector Classification

In this paper applied the multilevel classification. The first level, the classifier divided data into 2 classes, acute and chronic. And the second level, the classifier divided the acute data into 2 subclasses, AML and ALL, and chronic data into 2 subclasses too, CLL and CML.

The linear kernel is the default kernel of SVC. $C$ is the margin parameter that determines the trade-off between the maximization of the margin and minimization of classification error. Low-value of $C$ would made the decision surface smoother, while high-value $C$ would classifying every training sample correctly. From the training with several different $C$, the higher value $C$ make the classifier more sensitive and thorough for creating support vector.
Figure 7. Result from SVC with Linear Kernel for Dividing Acute and Chronic

In Figure 7, training data is classified with linear kernel into acute and chronic classes. The area with red indicates the class from chronic and the blue one shows the acute class.

Figure 8. Result from SVC with Linear Kernel for Dividing AML and ALL

When the test data will be entered into acute class, the data will be reclassified to AML and ALL class as shown in Figure 8. The area on figure shows blue area is class of ALL while the red one shows the class from AML.

Figure 9. Result from SVC with Linear Kernel for Dividing CML and CLL

When the initial classification of the data belongs to the chronic class, the data will be classified again into CLL and CML, as in Figure 9. The class of CML data is represented by a red area, while the data with CLL class is represented with a blue area.

Gaussian RBF kernel creates the future space from training data of support vector into Gaussian form. As defined before, the $\gamma$ value will determine how far the effect of training data reached, low value as ‘far’ and high value as ‘near’. The $\gamma$ parameter can be seen as the inverse of the sample effect radius that chosen by the model as the support vector.
Similar to the previous kernel use "linear kernel", the classification with the used of the RBF kernel also carried out a multilevel classification. The difference only in the used of the kernel, so that the formed hyperplane did not the same but they have the same level of the class used. The data first classified into chronic and acute. If the data has been classified in the acute class then the data was reclassified to ALL and AML. If the data has been classified as a chronic class, then the data was classified again into CML and CLL.
The results using Gaussian RBF kernel were variates, depends on the value of $C$ and $\gamma$. Bigger $\gamma$ value made the hyperplane in classification very detail towards training data. As in Figure 10, Figure 11 and Figure 12, experiments using small $\gamma$ had some error, there were data that should belong to the other class of classification. While the bigger $\gamma$ value, made error become smaller in SVC making process.

Parameters used were $\gamma$, $C$, and $d$, with $d$ as a degree. In this research, $d$ set as 3, while $C$ and $\gamma$ were independent. Similar to the RBF kernel, in the polynomial kernel, the $\gamma$ value greatly affected the shape of hyperplane from support vector.

![Figure 13. Result from SVC with Polynomial Kernel for Dividing Acute and Chronic](image)

The classification applied to SVC with the polynomial kernel also the same as before in the RBF kernel and linear. In Figure 13, the training data classified into two classes first, namely acute and chronic classes, blue indicates acute and red indicates chronic class.

![Figure 14. Result from SVC with Polynomial Kernel for Dividing AML and ALL](image)

After data has been classified at acute and chronic levels. then at the next level the training data classified into sub-classes from acute and chronic. Figure 14 shows if the data at the beginning has been classified as acute then the data was classified again into AML with the red area class and while the ALL class belongs to the blue area.
In addition, if the data has been classified in the chronic class, then the data classified again against the sub class of chronic namely CML and CLL. As in Figure 15, the area in blue indicates the class of CLL and the area in red shows the class of CML.

### 3.4. Performance Results

The two important parameters namely C (regularization constant) and kernel type were optimized using the grid search method. The observed trend was that the performance of the training set increases with the use of linear kernel type.
Figure 16. accuracy of SVM with several kernel (a) $\gamma = 0.1$ (b) $\gamma = 0.5$ (c) $\gamma = 1$ (d) $\gamma = 5$ (f) $\gamma = 10$, where the value of $\gamma$ useful for Gaussian Radial Basis Function kernel and Polynomial kernel.

Figure 16 shows that radial basis function kernel has better than other kernel. In several value of parameter $C$ and $\gamma$, the RBF has highest accuracy from other kernel. Indeed, when $C \geq 60$ and $\gamma \geq 5$ the RBF kernel perform 100% accuracy. Polynomial kernel has worse performance then linear kernel when $\gamma \leq 1$, but when $\gamma > 1$ its performance increase over linear kernel.

The value of $C$ controls the trade-off between margin and classification error. A high value of $C$ gave a bigger penalty toward classification error using training data as a comprehensive support vector. While the $\gamma$ value from RBF and polynomial kernel would affect the hyperplane. High-value $\gamma$, make hyperplane more precision toward the complexity of training data which became a support vector, while low-value $\gamma$, make the classification process hardly read the complexity of the data.

Higher $\gamma$ value doesn’t always mean better for classification. The precision toward training data will narrow the feature of each data. Therefore, the classified data should have a very high similarity from training data. While the existing DNA leukaemia data does not always similar to the existing training data, because each human DNA has differences.

4. Conclusion

Feature extraction using a second-order Markov chain generate 64 features from DNAs sequences. That feature useful for training data to classify with SVM. Classification divide the type of leukaemia become 4 class (Acute Lymphocytic Leukemia, Acute Myeloid Leukemia, Chronic Lymphocytic Leukaemia and Chronic Myeloid Leukaemia). SVM use 3 of kernels and analysing the parameters $C$ and $\gamma$ affected the result of training and classification. The higher value of $C$, the training process need more training data that become support vector, the higher value of $\gamma$, the more precision hyperplane created for each support vector.

The best kernel is radial basis function, that on several parameter the accuracy better then other kernel. Indeed, when $C \geq 60$ and $\gamma \geq 5$ the RBF kernel perform 100% accuracy. Polynomial kernel has worse performance then linear kernel when $\gamma \leq 1$, but when $\gamma > 1$ its performance increase over linear kernel.

5. Acknowledgment
We would like to thank Ministry of Technology, Research and Higher Education thank the Republic of Indonesia for financial support for our research.

REFERENCES

[1] Kemenkes RI. (2016). *Pusat Data Dan Informasi Kementrian Kesehatan RI 2015*. Jakarta: Kementrian Kesehatan RI. http://www.depkes.go.id/resources/download/pusdatin/profil-kesehatan-indonesia/Profil-Kesehatan-Indonesia-2016.pdf

[2] Dyrrskjot, L., Zieger, K., Real, F. X., Malats, N., Carrato, A., Hurst, C., Orntoft, T. F. (2007). Gene Expression Signatures Predict Outcome in Non-Muscle-Invasive Bladder Carcinoma: A Multicenter Validation Study. *Clinical Cancer Research, 13*(12), 3545–3551. https://doi.org/10.1158/1078-0432.CCR-06-2940

[3] Catto, J. W. F., Abbod, M. F., Wild, P. J., Linkens, D. A., Pilarsky, C., Rehman, I., Hamdy, F. C. (2010). The Application of Artificial Intelligence to Microarray Data: Identification of a Novel Gene Signature to Identify Bladder Cancer Progression. *European Urology*. https://doi.org/10.1016/j.euro.2009.10.029

[4] Ransohoff, D. F. (2004). Rules of evidence for cancer molecular-marker discovery and validation. *Nature Reviews Cancer*. https://doi.org/10.1038/nrc1322

[5] Nacong, N., Lusiyanti, D., Irawan, M.I., (2018). Sequence analysis of Leukemia DNA, AIP Conference Proceedings

[6] Haryanto T, Rijzaani H, and M.L. Fajar (2013). Identifikasi DNA dengan Rantai Markov Orde Satu dan Probabilistic Neural Network, *Seminar Nasional Teknologi Informasi* https://www.researchgate.net/profile/Toto_Haryanto2/publication

[7] Robin, S., Rodolphe, F., & Schbath, S. (2005). *DNA, Words and Models*. New York, NY, USA: Cambridge University Press. https://www.amazon.com/DNA-Words-Models-Statistics-Exceptional/dp/052184729X

[8] Shigeo Abe.(2005). *Support Vector Machines for Pattern Classification*. Kobe, Japan:Springer-verlag https://www.springer.com/gp/book/9781846282195

[9] Griffiths, A., Wessler, S., Carroll, S., & Swynghedauw, B. (2007). Introduction to genetic analysis. *Biofutur*. https://doi.org/10.1016/S0294-3506(00)90098-X

[10] Hsu, C., & Lin, C. (2002). A comparison of methods for multiclass support vector machines. *Neural Networks, IEEE Transactions On*, 13(2), 415–425. https://doi.org/10.1109/TNN.2002.1021904