

Research Article

Recombination Hotspot/Coldspot Identification Combining Three Different Pseudocomponents via an Ensemble Learning Approach

Bingquan Liu,¹ Yumeng Liu,² and Dong Huang^{2,3}

¹School of Computer Science and Technology, Harbin Institute of Technology, Harbin, Heilongjiang, China

²School of Computer Science and Technology, Harbin Institute of Technology Shenzhen Graduate School, Shenzhen, Guangdong 518055, China

³Key Laboratory of Network Oriented Intelligent Computation, Harbin Institute of Technology Shenzhen Graduate School, Shenzhen, Guangdong 518055, China

Correspondence should be addressed to Bingquan Liu; liubq@hit.edu.cn

Received 30 May 2016; Accepted 11 July 2016

Academic Editor: Qin Ma

Copyright © 2016 Bingquan Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Recombination presents a nonuniform distribution across the genome. Genomic regions that present relatively higher frequencies of recombination are called hotspots while those with relatively lower frequencies of recombination are recombination coldspots. Therefore, the identification of hotspots/coldspots could provide useful information for the study of the mechanism of recombination. In this study, a new computational predictor called SVM-EL was proposed to identify hotspots/coldspots across the yeast genome. It combined Support Vector Machines (SVMs) and Ensemble Learning (EL) based on three features including basic kmer (Kmer), dinucleotide-based auto-cross covariance (DACC), and pseudo dinucleotide composition (PseDNC). These features are able to incorporate the nucleic acid composition and their order information into the predictor. The proposed SVM-EL achieves an accuracy of 82.89% on a widely used benchmark dataset, which outperforms some related methods.

1. Introduction

Meiotic recombination describes the process of alleles' exchange between homologous chromosomes during meiosis [1]. It can provide material for natural selection by producing diverse gametes. It might also contribute to the evolution of the genome via gene conversion or mutagenesis [2–4].

Although the exact location where recombination happens in the genome and the mechanism of recombination are still unclear, it has been assured that recombination plays an important role in promoting genome evolution. Therefore, several studies have been performed on chromosomes [5–7] and found that recombination presents a nonuniform distribution across the genome. Genomic regions that present relatively higher frequencies of recombination are called hotspots while those with relatively lower frequencies of recombination are called recombination coldspots [8, 9]. With the number of the sequenced genomes showing explosive

growth, more reliable methods are urgently needed to be developed to identify the recombination spots.

The prediction of recombination hotspots or coldspots is still a challenging task, although much information can be acquired from the experiments. Recently, several computational models have been presented to identify the recombination hotspots/coldspots. For example, Liu et al. [10], based on sequence Kmer frequencies, proposed a model which combines the increment of diversity with quadratic discriminant analysis (IDQD). Later, this method was improved by adding gaps into the kmers [11]. Chen et al. presented a predictor called iRSpot-PseDNC trained with pseudo dinucleotide composition features [12].

The aforementioned methods extracted the features from DNA sequences in different aspects. For example, the model based on oligonucleotide frequencies considers the nucleic acid composition information. The iRSpot-PseDNC incorporates both the local nucleic acid composition information and

TABLE 1: The values of fifteen DNA dinucleotide properties.

	AA/TT	AC/GT	AG/CT	AT	CA/TG	CC/GG	CG	GA/TC	GC	TA
F-roll	0.04	0.06	0.04	0.05	0.04	0.04	0.04	0.05	0.05	0.03
F-tilt	0.08	0.07	0.06	0.10	0.06	0.06	0.06	0.07	0.07	0.07
F-twist	0.07	0.06	0.05	0.07	0.05	0.06	0.05	0.06	0.06	0.05
F-slide	6.69	6.80	3.47	9.61	2.00	2.99	2.71	4.27	4.21	1.85
F-shift	6.24	2.91	2.80	4.66	2.88	2.67	3.02	3.58	2.66	4.11
F-rise	21.34	21.98	17.48	24.79	14.51	14.25	14.66	18.41	17.31	14.24
Roll	1.05	2.01	3.60	0.61	5.60	4.68	6.02	2.44	1.70	3.50
Tilt	-1.26	0.33	-1.66	0.00	0.14	-0.77	0.00	1.44	0.00	0.00
Twist	35.02	31.53	32.29	30.72	35.43	33.54	33.67	35.67	34.07	36.94
Slide	-0.18	-0.59	-0.22	-0.68	0.48	-0.17	0.44	-0.05	-0.19	0.04
Shift	0.01	-0.02	-0.02	0.00	0.01	0.03	0.00	-0.01	0.00	0.00
Rise	3.25	3.24	3.32	3.21	3.37	3.36	3.29	3.30	3.27	3.39
Energy	-1.00	-1.44	-1.28	-0.88	-1.45	-1.84	-2.17	-1.30	-2.24	-0.58
Enthalpy	-7.60	-8.40	-7.80	-7.20	-8.50	-8.00	-10.60	-8.20	-9.80	-7.20
Entropy	-21.30	-22.40	-21.00	-20.40	-22.70	-19.90	-27.20	-22.20	-24.40	-21.30

dinucleotide property index $\mu_1(\mu_2)$; $\bar{P}_{\mu_1}(\bar{P}_{\mu_2})$ represents the average value of $P_{\mu_1}(R_i R_{i+1})(P_{\mu_2}(R_i R_{i+1}))$ for a DNA sequence.

The features of DACC contain global sequence-order information, and it can be generated via Pse-in-One [17] which includes two generation approaches. The generation steps of DACC feature can be described as follows.

For web server approach, firstly, choose the DNA sequences (PseDAC-General) option, then select DACC in the tab of Mode, and set the value of lag. Secondly, upload a user-defined physicochemical index file called user_property and the values of fifteen dinucleotide physicochemical properties are shown in Table 1. Finally, input or upload the DNA sequence file in FASTA format, click the Submit button, and then you will see the results and you can download them as a text file (Figure 2).

For stand-alone approach, DACC features can be easily generated by using the following command line:

```
./acc.py -e user_property -f svm -l +13 DNA DACC'
```

where -e user_property represents the user-defined physicochemical index file, -f svm and -l +1 have the same meaning with the above command line, the parameter lag equals 3, the sequence type is DNA, and the method used is DACC.

2.2.3. Pseudo Dinucleotide Composition (PseDNC). Given a DNA sequence \mathbf{D} represented as (2), the PseDNC feature vector \mathbf{D} can be defined as [17]

$$\mathbf{D} = [d_1 \ d_2 \ d_3 \ \dots \ d_{16} \ d_{16+\lambda} \ \dots \ d_{16+\lambda}]^T, \quad (5)$$

where

$$d_k = \begin{cases} \frac{f_k}{\sum_{i=1}^{16} f_i + w \sum_{j=1}^{\lambda} \theta_j}, & (1 \leq k \leq 16), \\ \frac{w\theta_{k-16}}{\sum_{i=1}^{16} f_i + w \sum_{j=1}^{\lambda} \theta_j}, & (17 \leq k \leq 16 + \lambda), \end{cases} \quad (6)$$

where f_k ($1 \leq k \leq 16$) represents the normalized frequency of dinucleotides along the DNA sequence; w ($0 \leq w \leq 1$)

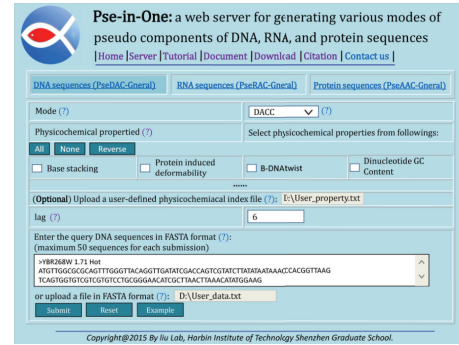


FIGURE 2: An example of the DACC features' generation by using Pse-in-One.

represents the weight factor; λ is the top counted tiers of the correlation in a DNA, θ_j ($1 \leq j \leq \lambda$) measures the correlation between dinucleotides in the DNA, which is defined as

$$\begin{aligned} \theta_1 &= \frac{1}{L-2} \sum_{i=1}^{L-2} \Theta(R_i R_{i+1}, R_{i+1} R_{i+2}), \\ \theta_2 &= \frac{1}{L-3} \sum_{i=1}^{L-3} \Theta(R_i R_{i+1}, R_{i+2} R_{i+3}), \\ \theta_3 &= \frac{1}{L-4} \sum_{i=1}^{L-4} \Theta(R_i R_{i+1}, R_{i+3} R_{i+4}), \\ &\vdots \\ \theta_\lambda &= \frac{1}{L-1-\lambda} \sum_{i=1}^{L-1-\lambda} \Theta(R_i R_{i+1}, R_{i+\lambda} R_{i+\lambda+1}), \end{aligned} \quad (7)$$

$(\lambda < L),$

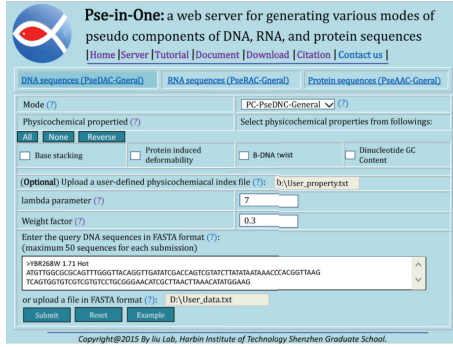


FIGURE 3: An example of the PseDNC features' generation by using Pse-in-One.

where

$$\Theta(R_i R_{i+1}, R_j R_{j+1}) = \frac{1}{\mu} \sum_{\mu=1}^{\mu} [P_{\mu}(R_i R_{i+1}) - P_{\mu}(R_j R_{j+1})]^2, \quad (8)$$

where μ represents the indices of the dinucleotide property; $P_{\mu}(R_i R_{i+1}) (P_{\mu}(R_j R_{j+1}))$ represents the value of dinucleotide $R_i R_{i+1} (R_j R_{j+1})$ at position $i (j)$ for the dinucleotide property index μ .

Pseudo dinucleotide composition (PseDNC) [17] not only incorporates the local nucleic acid composition information and the global or long range information along the DNA sequences, but also incorporates the dinucleotide properties into feature vectors.

For web server approach, the generation steps of the feature vectors are similar to those of the DACC's. For web server approach, an example is shown in Figure 3.

For stand-alone approach, the command line is

```
./pse.py -e user_property -f svm -l +1 7 0.3 DNA PseDNC'
```

where $-e$ user_property, $-f$ svm, and $-l +1$ have the same meaning with the above command line, lambda equals 7, the value of weight equals 0.3, the sequence type is DNA, and the method used is PseDNC.

The meanings of all the parameters for these scripts are described in [17].

2.3. Support Vector Machine (SVM). Support Vector Machine (SVM) is a kind of algorithm based on statistical learning theory proposed by Vapnik [20–22], which has been widely used for many bioinformatics tasks [23–27].

In the current study, the LIBSVM package version 3.21 [18] has been employed. The SVM parameters, the kernel width parameter γ and the regularization parameter C , were optimized via the grid tool provided by LIBSVM [18].

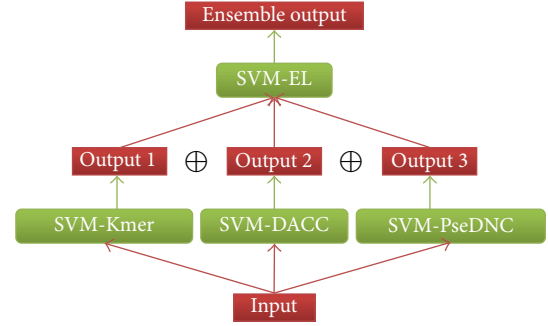


FIGURE 4: The basic framework for an ensemble classifier.

In the current study, three basic predictors are proposed, including SVM-Kmer, SVM-DACC, and SVM-PseDNC. The values of SVM-Kmer's parameters are shown as follows:

$$\begin{aligned} C &= 2^7, \\ \gamma &= 2, \\ k &= 6. \end{aligned} \quad (9)$$

The values of SVM-DACC's parameters are shown as follows:

$$\begin{aligned} C &= 2^3, \\ \gamma &= 2^{-3}, \\ \text{lag} &= 6. \end{aligned} \quad (10)$$

The values of SVM-PseDNC's parameters are shown as follows:

$$\begin{aligned} C &= 2^{13}, \\ \gamma &= 2^3, \\ \lambda &= 7, \\ w &= 0.3. \end{aligned} \quad (11)$$

2.4. Ensemble Learning. In machine learning, ensemble learning is the process by which multiple classifiers are constructed and combined based on the same dataset to obtain a better performance than a single classifier [28, 29] and existing popular multiobjective optimization evolutionary algorithms can be used for ensemble learning [30, 31]. Ensemble classifier also performed well in several bioinformatics problems. In the current study, the basic framework for an ensemble classifier is illustrated in Figure 4. The final results are obtained by fusing three individual classifier outcomes, as illustrated below.

Suppose the ensemble classifier \mathbb{C} is defined as

$$\mathbb{C} = \mathbb{C}_1 \oplus \mathbb{C}_2 \oplus \mathbb{C}_3, \quad (12)$$

where \mathbb{C}_1 represents the classifier SVM-Kmer, \mathbb{C}_2 represents the classifier SVM-DACC, and \mathbb{C}_3 represents the classifier SVM-PseDNC. The symbol \oplus denotes the fusing operator.

Therefore, the process of the ensemble classifier can be formulated as follows:

$$R_j = \frac{1}{3}P_i(\mathbf{S}, L_j), \quad (i = 1, 2, 3; j = 1, 2), \quad (13)$$

where L_1 is the set only containing recombination hotspots and L_2 is the set of recombination coldspots. $P_i(\mathbf{S}, L_j)$ is the probability for DNA sequence \mathbf{S} which belongs to category L_j obtained by the i th basic classifier.

Thus, which category the query DNA \mathbf{S} belongs to is to be determined by using its average probability calculated by (13); that is, suppose that

$$R_\mu = \text{Max} \{R_1, R_2\}, \quad (14)$$

where the operator max represents selecting a larger value in the brackets, and the subscript μ represents the query DNA \mathbf{S} belonging to category L_μ .

2.5. Criteria for Performance Evaluation. The prediction results can be divided into true positive (TP), false negative (FN), false positive (FP), and true negative (TN) [32]. In the current study, jackknife test [33–37] was employed and four kinds of evaluation indexes were adopted, including Sensitivity (Se), Specificity (Sp), Accuracy (Acc), and Matthew's Correlation Coefficient (Mcc). They are described as

$$\begin{aligned} \text{Se} &= \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\%, \\ \text{Sp} &= \frac{\text{TN}}{\text{TN} + \text{FP}} \times 100\%, \\ \text{Acc} &= \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \times 100\%, \\ \text{Mcc} &= \frac{(\text{TP} \times \text{TN}) - (\text{FP} \times \text{FN})}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}}. \end{aligned} \quad (15)$$

3. Results and Discussion

3.1. Performance of the Three Basic Classifiers. As an inherent property, sequence-order is important for the classification of DNA sequences. So, three basic methods based on sequence-order information are adopted to identify recombination hotspots/coldspots. Table 2 shows the performance of the three methods. According to the table, we can see that SVM-DACC and SVM-PseDNC outperform SVM-Kmer on the prediction accuracy index. The main reason is that SVM-Kmer is only based on local sequence-order information, while both of SVM-DACC and SVM-PseDNC also contain global sequence-order information.

3.2. The Performance of the Three Basic Predictors Can Be Further Improved by Using Ensemble Learning. Based on the analysis above, we have proposed three basic predictors for identifying recombination hotspots/coldspots. These methods capture DNA information from different aspects.

TABLE 2: Results on benchmark dataset for different predictors proposed in the current study.

Predictor	Test method	Se (%)	Sp (%)	Acc (%)	MCC
SVM-Kmer ^a	Jackknife	75.92	86.29	81.59	0.628
SVM-DACC ^b	Jackknife	76.12	87.99	82.61	0.649
SVM-PseDNC ^c	Jackknife	72.04	90.69	82.24	0.644
SVM-EL	Jackknife	76.33	88.33	82.89	0.654

^aThe parameters used are $k = 6$ for SVM-Kmer and $C = 2^7$ and $\gamma = 2$ for LIBSVM [18].

^bThe parameters used are lag = 6 for SVM-DACC and $C = 2^3$ and $\gamma = 2^{-3}$ for LIBSVM [18].

^cThe parameters used are $\lambda = 7$ and $w = 0.3$ for SVM-PseDNC and $C = 2^{13}$ and $\gamma = 2^3$ for LIBSVM [18].

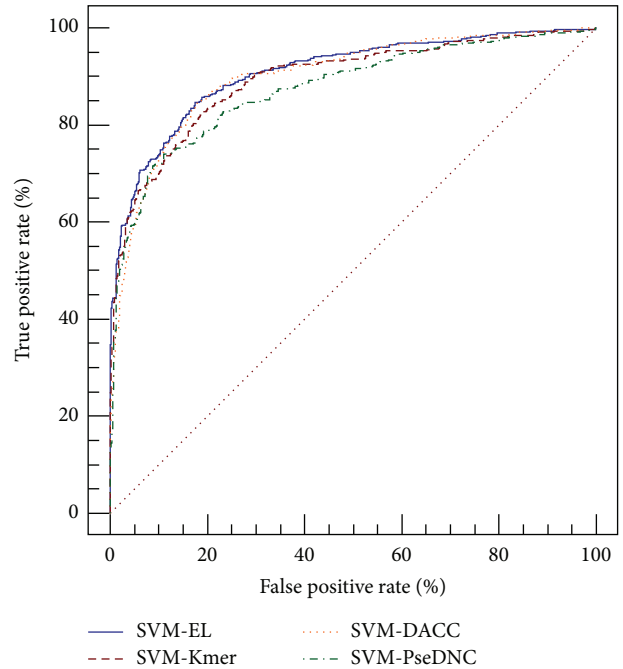


FIGURE 5: The comparison of different predictors for hotspots/coldspots identification. The areas under ROC curves (AUC) of SVM-EL, SVM-DACC, SVM-Kmer, and SVM-PseDNC are 0.91, 0.90, 0.89, and 0.87, respectively.

Therefore, we presented a complementary method SVM-EL which can fuse these basic methods to improve the prediction performance. The performance of SVM-EL is shown in Table 2, from which we can see that SVM-EL outperforms the three basic methods. Besides, the corresponding receiver operating characteristic (ROC) curves of the four classifiers were drawn in Figure 5. AUC, the area under the ROC curve, is often used to indicate the performance of a classifier: the larger the value, the better the classifier.

As shown in Figure 5, the predictor SVM-EL showed the top performance, outperforming three basic methods: SVM-Kmer, SVM-DACC, and SVM-PseDNC.

3.3. Comparison with Other Related Predictors. Two state-of-the-art methods, IDQD [10] and iRSpot-PseDNC, were

TABLE 3: Results on benchmark dataset for different predictors.

Predictor	Test method	Se (%)	Sp (%)	Acc (%)	MCC
IDQD ^a	5-fold	79.40	81.00	80.30	0.603
iRSpot-PseDNC ^b	Jackknife	73.06	89.49	82.04	0.638
SVM-EL	Jackknife	76.33	88.33	82.89	0.654

^aFrom Liu et al. [10].

^bFrom Chen et al. [12].

selected to compare with the proposed SVM-EL. Table 3 shows the results of various methods on the benchmark dataset.

According to Table 3, we can see that SVM-EL outperforms the other methods. The main reason is that IDQD and SVM-Kmer only consider local sequence-order information, and iRSpot-PseDNC, SVM-DACC, and SVM-PseDNC improved them by incorporating global sequence-order information. However, SVM-EL not only incorporates the local nucleic acid information, but also incorporates the global information. Therefore, we conclude that SVM-EL would be a useful tool for hotspots/coldspots identification.

4. Conclusion

In this article, we proposed a predictor called SVM-EL for yeast hotspot/coldspot identification, which combines Support Vector Machine (SVM) with Ensemble Learning (EL). The approach combined with different predictors trained by different features contributes to the improvement of prediction accuracy. SVM-EL is trained by different features, including basic kmer (Kmer), dinucleotide-based auto-cross covariance (DACC), and pseudo dinucleotide composition (PseDNC). All these features can be generated by Pse-in-One [17], which is a powerful web server for generating various DNA, RNA, or protein features. It also provides a stand-alone version to users, which is easy to use. Via jackknife test, it was observed that the predictor outperforms other predictors. In the future, we will consider using other approaches for yeast hotspot/coldspot identification, such as bioinspired computing models [38–45].

Competing Interests

The authors declare no competing financial interests.

Authors' Contributions

Bingquan Liu conceived the study and designed the experiments and participated in designing the study, drafting the manuscript, and performing the statistical analysis. Yumeng Liu participated in coding the experiments and drafting the manuscript. Dong Huang participated in performing the statistical analysis. All authors read and approved the final manuscript. Bingquan Liu and Yumeng Liu contributed equally to this paper.

Acknowledgments

This work was supported by the National High Technology Research and Development Program of China (863 Program) (2015AA015405), the National Natural Science Foundation of China (nos. 61300112, 61573118, 61272383, and 61572151), the Natural Science Foundation of Guangdong Province (2014A030313695), Guangdong Natural Science Funds for Distinguished Yong Scholars (2016A030306008), and Scientific Research Foundation in Shenzhen (Grant no. JCYJ20150626110425228).

References

- [1] A. Lynn, T. Ashley, and T. Hassold, "Variation in human meiotic recombination," *Annual Review of Genomics and Human Genetics*, vol. 5, pp. 317–349, 2004.
- [2] C. C. A. Spencer, P. Deloukas, S. Hunt et al., "The influence of recombination on human genetic diversity," *PLoS Genetics*, vol. 2, no. 9, article e148, 2006.
- [3] N. Galtier, G. Piganeau, D. Mouchiroud, and L. Duret, "GC-content evolution in mammalian genomes: the biased gene conversion hypothesis," *Genetics*, vol. 159, no. 2, pp. 907–911, 2001.
- [4] M. J. Lercher and L. D. Hurst, "Human SNP variability and mutation rate are higher in regions of high recombination," *Trends in Genetics*, vol. 18, no. 7, pp. 337–340, 2002.
- [5] F. Baudat and A. Nicolas, "Clustering of meiotic double-strand breaks on yeast chromosome III," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 94, no. 10, pp. 5213–5218, 1997.
- [6] S. Klein, D. Zenvirth, V. Dror, A. B. Barton, D. B. Kaback, and G. Simchen, "Patterns of meiotic double-strand breakage on native and artificial yeast chromosomes," *Chromosoma*, vol. 105, no. 5, pp. 276–284, 1996.
- [7] D. Zenvirth, T. Arbel, A. Sherman, M. Goldway, S. Klein, and G. Simchen, "Multiple sites for double-strand breaks in whole meiotic chromosomes of *Saccharomyces cerevisiae*," *The EMBO Journal*, vol. 11, no. 9, pp. 3441–3447, 1992.
- [8] E. Mancera, R. Bourgon, A. Brozzi, W. Huber, and L. M. Steinmetz, "High-resolution mapping of meiotic crossovers and non-crossovers in yeast," *Nature*, vol. 454, no. 7203, pp. 479–485, 2008.
- [9] J. L. Gerton, J. DeRisi, R. Shroff, M. Lichten, P. O. Brown, and T. D. Petes, "Global mapping of meiotic recombination hotspots and coldspots in the yeast *Saccharomyces cerevisiae*," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 21, pp. 11383–11390, 2000.
- [10] G. Liu, J. Liu, X. Cui, and L. Cai, "Sequence-dependent prediction of recombination hotspots in *Saccharomyces cerevisiae*," *Journal of Theoretical Biology*, vol. 293, pp. 49–54, 2012.
- [11] R. Wang, Y. Xu, and B. Liu, "Recombination spot identification Based on gapped k-mers," *Scientific Reports*, vol. 6, article 23934, 2016.
- [12] W. Chen, P.-M. Feng, H. Lin, and K.-C. Chou, "iRSpot-PseDNC: identify recombination spots with pseudo dinucleotide composition," *Nucleic Acids Research*, vol. 41, no. 6, article e68, 2013.
- [13] B. Liu, F. Liu, L. Fang, X. Wang, and K.-C. Chou, "repDNA: a Python package to generate various modes of feature vectors for

- DNA sequences by incorporating user-defined physicochemical properties and sequence-order effects,” *Bioinformatics*, vol. 31, no. 8, pp. 1307–1309, 2015.
- [14] Q. Dong, S. Zhou, and J. Guan, “A new taxonomy-based protein fold recognition approach based on autocross-covariance transformation,” *Bioinformatics*, vol. 25, no. 20, pp. 2655–2662, 2009.
- [15] W. Chen, X. Zhang, J. Brooker, H. Lin, L. Zhang, and K.-C. Chou, “PseKNC-General: a cross-platform package for generating various modes of pseudo nucleotide compositions,” *Bioinformatics*, vol. 31, no. 1, pp. 119–120, 2015.
- [16] B. Liu, R. Long, and K.-C. Chou, “iDHS-EL: identifying DNase I hypersensitive sites by fusing three different modes of pseudo nucleotide composition into an ensemble learning framework,” *Bioinformatics*, vol. 32, no. 16, pp. 2411–2418, 2016.
- [17] B. Liu, F. Liu, X. Wang, J. Chen, L. Fang, and K.-C. Chou, “Pse-in-One: a web server for generating various modes of pseudo components of DNA, RNA, and protein sequences,” *Nucleic Acids Research*, vol. 43, no. 1, pp. W65–W71, 2015.
- [18] C.-C. Chang and C.-J. Lin, “LIBSVM: a library for support vector machines,” *ACM Transactions on Intelligent Systems and Technology*, vol. 2, article 27, 2011.
- [19] Y. Guo, L. Yu, Z. Wen, and M. Li, “Using support vector machine combined with auto covariance to predict protein-protein interactions from protein sequences,” *Nucleic Acids Research*, vol. 36, no. 9, pp. 3025–3030, 2008.
- [20] V. N. Vapnik, *Statistical Learning Theory*, John Wiley & Sons, New York, NY, USA, 1998.
- [21] Y. Wu and S. Krishnan, “Combining least-squares support vector machines for classification of biomedical signals: a case study with knee-joint vibroarthrographic signals,” *Journal of Experimental and Theoretical Artificial Intelligence*, vol. 23, no. 1, pp. 63–77, 2011.
- [22] B. Liu, L. Fang, R. Long, X. Lan, and K. Chou, “iEnhancer-2L: a two-layer predictor for identifying enhancers and their strength by pseudo k-tuple nucleotide composition,” *Bioinformatics*, vol. 32, no. 3, pp. 362–369, 2016.
- [23] J. Chen, X. Wang, and B. Liu, “iMiRNA-SSF: improving the identification of MicroRNA precursors by combining negative sets with different distributions,” *Scientific Reports*, vol. 6, article 19062, 2016.
- [24] B. Liu, D. Zhang, R. Xu et al., “Combining evolutionary information extracted from frequency profiles with sequence-based kernels for protein remote homology detection,” *Bioinformatics*, vol. 30, no. 4, pp. 472–479, 2014.
- [25] Q. Zou, Y. Mao, L. Hu, Y. Wu, and Z. Ji, “miRClassify: an advanced web server for miRNA family classification and annotation,” *Computers in Biology and Medicine*, vol. 45, no. 1, pp. 157–160, 2014.
- [26] W. Chen and H. Lin, “Prediction of midbody, centrosome and kinetochore proteins based on gene ontology information,” *Biochemical and Biophysical Research Communications*, vol. 401, no. 3, pp. 382–384, 2010.
- [27] D. Li, Y. Ju, and Q. Zou, “Protein folds prediction with hierarchical structured SVM,” *Current Proteomics*, vol. 13, no. 2, pp. 79–85, 2016.
- [28] B. Liu, S. Wang, and X. Wang, “DNA binding protein identification by combining pseudo amino acid composition and profile-based protein representation,” *Scientific Reports*, vol. 5, Article ID 15479, 2015.
- [29] M. Wu, L. Liao, X. Luo et al., “Analysis and classification of stride patterns associated with children development using gait signal dynamics parameters and ensemble learning algorithms,” *BioMed Research International*, vol. 2016, Article ID 9246280, 8 pages, 2016.
- [30] X. Zhang, Y. Tian, R. Cheng, and Y. Jin, “An efficient approach to nondominated sorting for evolutionary multiobjective optimization,” *IEEE Transactions on Evolutionary Computation*, vol. 19, no. 2, pp. 201–213, 2015.
- [31] X. Zhang, Y. Tian, and Y. Jin, “A knee point-driven evolutionary algorithm for many-objective optimization,” *IEEE Transactions on Evolutionary Computation*, vol. 19, no. 6, pp. 761–776, 2015.
- [32] Y. Wu, P. Chen, X. Luo et al., “Quantification of knee vibroarthrographic signal irregularity associated with patellofemoral joint cartilage pathology based on entropy and envelope amplitude measures,” *Computer Methods and Programs in Biomedicine*, vol. 130, pp. 1–12, 2016.
- [33] B. Liu, J. Xu, X. Lan et al., “iDNA-Prot—dis: identifying DNA-binding proteins by incorporating amino acid distance-pairs and reduced alphabet profile into the general pseudo amino acid composition,” *PLoS ONE*, vol. 9, no. 9, Article ID e106691, 2014.
- [34] W. Chen and H. Lin, “Identification of voltage-gated potassium channel subfamilies from sequence information using support vector machine,” *Computers in Biology and Medicine*, vol. 42, no. 4, pp. 504–507, 2012.
- [35] W. Chen, P. Feng, and H. Lin, “Prediction of ketoacyl synthase family using reduced amino acid alphabets,” *Journal of Industrial Microbiology and Biotechnology*, vol. 39, no. 4, pp. 579–584, 2012.
- [36] B. Liu, J. Chen, and X. Wang, “Application of learning to rank to protein remote homology detection,” *Bioinformatics*, vol. 31, no. 21, pp. 3492–3498, 2014.
- [37] B. Liu, L. Fang, F. Liu, X. Wang, J. Chen, and K.-C. Chou, “Identification of real microRNA precursors with a pseudo structure status composition approach,” *PLoS ONE*, vol. 10, no. 3, Article ID e0121501, 2015.
- [38] T. Song, J. Xu, and L. Pan, “On the universality and non-universality of spiking neural P systems with rules on synapses,” *IEEE Transactions on NanoBioscience*, vol. 14, no. 8, pp. 960–966, 2015.
- [39] T. Song and L. Pan, “Spiking neural P systems with request rules,” *Neurocomputing*, vol. 193, pp. 193–200, 2016.
- [40] X. Wang, T. Song, F. Gong, and P. Zheng, “On the computational power of spiking neural P systems with self-organization,” *Scientific Reports*, vol. 6, Article ID 27624, 2016.
- [41] X. Zhang, L. Pan, and A. Păun, “On the universality of axon P systems,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 11, pp. 2816–2829, 2015.
- [42] X. Zeng, L. Xu, X. Liu, and L. Pan, “On languages generated by spiking neural P systems with weights,” *Information Sciences*, vol. 278, pp. 423–433, 2014.
- [43] X. Zeng, X. Zhang, T. Song, and L. Pan, “Spiking neural P systems with thresholds,” *Neural Computation*, vol. 26, no. 7, pp. 1340–1361, 2014.
- [44] X. Zhang, Y. Liu, B. Luo, and L. Pan, “Computational power of tissue P systems for generating control languages,” *Information Sciences*, vol. 278, pp. 285–297, 2014.
- [45] B. Liu, S. Wang, Q. Dong, S. Li, and X. Liu, “Identification of DNA-binding proteins by combining auto-cross covariance transformation and ensemble learning,” *IEEE Transactions on NanoBioscience*, vol. 15, no. 4, pp. 328–334, 2016.