Higher-Order Statistics and Neural Network Based Multi-Classifier System for Gene Identification

Teddy Surya Gunawan[1,2], Eliathamby Ambikairajah[1], and Julien Epps[1]

[1]School of Electrical Engineering & Telecommunications
The University of New South Wales
Sydney, NSW 2052, Australia

[2]Department of Electrical and Computer Engineering
International Islamic University Malaysia
Gombak, Selangor 53100, Malaysia

tsgunawan@iiu.edu.my, ambi@ee.unsw.edu.au, j.epps@unsw.edu.au

***Abstract*—This paper presents the use of higher order statistics and a neural network based multi-classifier system for gene and exon identification of a DNA sequence. Newly proposed higher order statistics features, combined with frequency domain analysis, are used to train three different neural networks. Classification results of the three individual neural networks are combined through an aggregation function, of which five variants are compared herein. An evaluation of the multi-classifier system on 117 sequences from the HMR195 database shows that when different opinions of more classifiers on the same input data are integrated within a multi-classifier system, a relative improvement in precision of 5% over the individual performances of the neural networks can be obtained.**

***Keywords*: Higher-order statistics, genomic signal processing, neural networks, multi-classifier systems.**

I. INTRODUCTION

The gene identification problem, which requires the prediction of the protein-coding regions (exons) in DNA sequences through computational means, has attracted significant research attention for some time. Worldwide initiatives on genome sequencing have necessitated the development of new approaches to assess rapidly the potential of a given nucleotide sequence in a functional context. Despite the existence of various data-driven gene finding programs such as FGENES [1], GeneMark.hmm [2], Genie [3], Genscan [4], HMMgene [5], Morgan [6], and MZEF [7], improvements to the accuracy of gene prediction are still highly desirable [8, 9].

A number of methods have been proposed for gene detection based on distinctive features of protein-coding sequences [10-12]. The different methods are based on a variety of contrasting characteristics of exons and introns. These methods employ for example differences in the patterns of codon usage [11], neural networks [12], or the discrete Fourier transform [13]. Furthermore, as higher order statistics (HOS) are able to reveal hidden information not found by normal statistics [14], the use of HOS for gene identification will be investigated.

Neural networks have extensively been used in bioinformatics [10, 12, 15], especially for gene identification. In this paper, we propose a neural network-based multi-classifier system for protein coding identification. The proposed system contains three neural networks that operate on feature vectors from Fourier transform and higher order statistics to deduce coding or non-coding region decisions.

The objective of this paper is to investigate the use of higher order statistics for gene identification and to evaluate the performance of a neural network based multi-classifier system. Section II discusses the feature extraction, i.e. higher order statistics features, and periodicity-3 and periodicity-10.5 features. Section III discusses the neural network based multi-classifier system and the data set used for training and testing the system. The performance evaluation is presented in Section IV, while Section V concludes this paper.

II. FEATURE EXTRACTION

Higher order statistics (HOS) have been applied in many diverse fields, such as radar, plasma physics, biomedicine, array processing, and blind equalization [14]. These statistics not only reveal amplitude information about a process, but also reveal phase information. By using the HOS in a DNA sequence, we hope that we can reveal any hidden information that might be useful for the gene identification. In this section, the features from HOS analysis will be further investigated to determine whether it can be used as discriminative features in identifying the protein coding region of a DNA sequence.

In this paper, we extract features from higher order statistics [14], i.e. mean, variance, skewness, and kurtosis, and signal processing based features, i.e. the periodicity 3 and 10.5 [16] of the Fast Fourier Transform (FFT) spectrum of a DNA sequence. In order to apply digital signal processing techniques and higher order statistics for feature extraction, the character sequences of DNA should be first converted into four binary indicator numeric sequences. The simplest and most popular mapping of a DNA sequence is known as the Voss representation [17]. For example, for a DNA sequence $x[n]$=CGATGACGAA, the binary indicator sequence for each base type, $x_\ell[n], \forall \ell \in \{A, C, G, T\}$, would be

$$x_A[n] = \{0,0,1,0,0,1,0,0,1,1\}, \quad x_C[n] = \{1,0,0,0,0,0,1,0,0,0\}$$
$$x_G[n] = \{0,1,0,0,1,0,0,1,0,0\}, \quad x_T[n] = \{0,0,0,1,0,0,0,0,0,0\}$$
$$x_A[n] + x_C[n] + x_G[n] + x_T[n] = 1$$

where n represents the base index. From a biological perspective, the Voss representation characterizes the frequency of occurrence of each individual base ℓ in the DNA sequence. Other popular DNA representations for genomic signal processing can be found in [18].

A. Higher Order Statistics

To extract the higher-order statistics features for gene identification, we calculate the mean, variance, skewness, and kurtosis of each representation of DNA sequence, $x_\ell[n]\times w[n]$ for a given window length N. The window size N is chosen to be sufficiently large (in the order of few hundred, e.g. 351 as used in [4]). The Bartlett window $w[n]$ is utilized, as it removes the extraneous peaks introduced by the abrupt edges of the rectangular window [19]. The window is then moved by one nucleotide. For each frame, the first moment, the mean, for each base type ℓ is calculated as follows:

$$\mu_\ell = E(x_\ell) = \frac{1}{N}\sum_{n=1}^{N} x_\ell[n]\times w[n] \quad (1)$$

Variance, the second moment, is a measure of the statistical dispersion of a DNA sequence, defined as follows:

$$\sigma_\ell^2 = E\left((x_\ell - \mu_\ell)^2\right) = \frac{1}{N}\sum_{n=1}^{N}\left(x_\ell[n]\times w[n] - \mu_\ell\right)^2 \quad (2)$$

Skewness, the third moment, is a measure of symmetry. A DNA sequence is symmetric if it looks the same to the left and right of the center point of the frequency of occurrence. Skewness is defined as follows:

$$\gamma_\ell = \frac{E\left((x_\ell - \mu_\ell)^3\right)}{\sigma_\ell^3} = \frac{1}{N}\sum_{n=1}^{N}\left(\frac{x_\ell[n]\times w[n] - \mu_\ell}{\sigma_\ell}\right)^3 \quad (3)$$

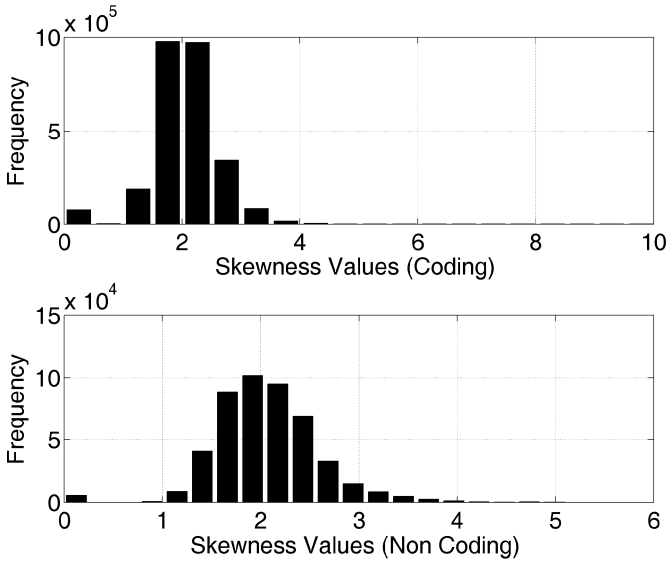

Figure 1. The histogram of skewness features for coding and non coding region of DNA sequences

Kurtosis, the fourth moment, is a measure of whether the data are peaked or flat relative to a normal distribution, i.e. a DNA sequence with high kurtosis tends to have a distinct peak near the mean, decline rather rapidly, and have heavy tails. It is defined as follows:

$$\kappa_\ell = \frac{E\left((x_\ell - \mu_\ell)^4\right)}{\sigma_\ell} - 3 = \frac{1}{N}\sum_{n=1}^{N}\left(\frac{x_\ell[n]\times w[n] - \mu_\ell}{\sigma_\ell}\right)^4 - 3 \quad (4)$$

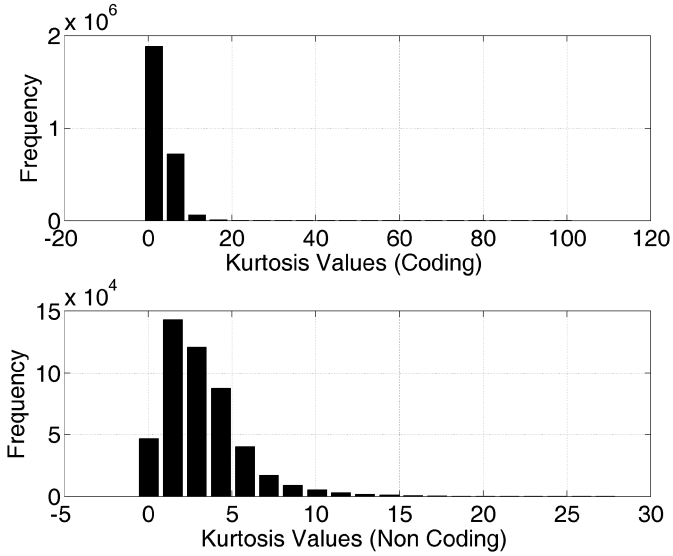

Figure 2. The histogram of kurtosis features for coding and non coding region of DNA sequences

Fig.1 and Fig. 2 show the distribution of skewness and kurtosis features on the HMR195 training sequences (see Section III.B). The figures show the discriminative nature of the skewness and kurtosis values to determine the coding and non coding region in a DNA sequence. Hence, it will be suitable to use the HOS features for gene identification.

B. Periodicity 3 and 10.5 features

The discrete Fourier transform (DFT) of a DNA sequence $x_\ell[n]$ of length N is defined as

$$X_\ell[k] = \sum_{n=0}^{N-1} x_\ell[n]\times w[n]\times e^{-j\frac{2\pi k}{N}} \quad (5)$$

where $k = 0,\ldots,N-1$ and $\ell \in \{A,C,G,T\}$. The DFTs $X_A[k]$, $X_C[k]$, $X_G[k]$ and $X_T[k]$ for the above indicator sequences can thus be calculated using equation (5). The periodicity of 3 and 10.5 in protein coding regions of a DNA sequence suggests that the DFT coefficients corresponding to $k = \frac{N}{3}$ and $k \approx \frac{N}{10.5}$ will be large [16, 20]. Thus if we take the window size N to be sufficiently large (again, for example 351 base pairs), peaks in the magnitude $|X_\ell[k]|$ will be observed at the frequency indices $k = \frac{N}{3}$ and $k \approx \frac{N}{10.5}$ [13, 19], corresponding to coding regions (relatively low values will be found for non-coding regions). However, the values of these peaks vary

significantly even for different DNA sequences derived from the same organisms. To overcome this problem, a 'signal-to-noise ratio' and a threshold were used in [19] to detect the protein coding. Here, the signal-to-noise ratio $SNR_\ell[k]$ is used as a frequency domain feature and is calculated as follows [19]:

$$SNR_\ell[k] = \frac{|X_\ell[k]|^2}{2|\overline{X}_\ell[k]|^2} \quad (6)$$

where $|\overline{X}_\ell[k]|^2$ is the average magnitude spectrum for each base. Note that the window is then moved by one sample until all the DNA sequence has been processed.

III. Neural Network Based Multi Classifier System

The process of classification describes the allocation of previously unknown data into a number of predetermined groups or classes. Artificial neural networks are one of the more popular implementations of the computational intelligence based-classification paradigm [10, 12, 21]. In this paper, the input to the classifier is the extracted features as described in Section II. Moreover, the output of the classifier is "coding" (exon) and "non-coding" (intron) nucleotides. The main goal of learning in classification problems is generalization, i.e. how to accurately classify genes not included in the training set.

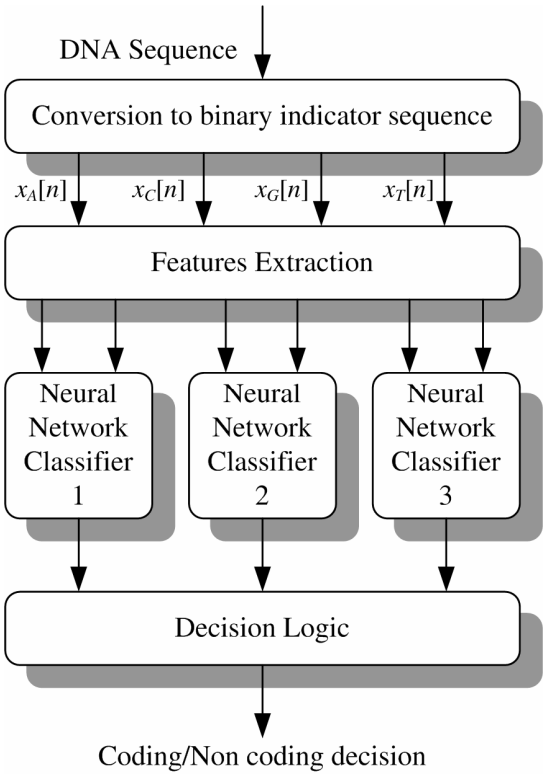

Figure 3. Neural network based multi-classifier system for gene identification

In this paper, we propose a neural network based multi-classifier system as shown in Figure 3. A neural network multi-classifier system was used rather than single neural network. A single neural network approach requires a larger network size as the input feature vectors have dimension 24 rather than 8 for each three neural network multi-classifier system. So, it requires more memory and is slow to train. Informal experiments showed that a single neural network provides reduced accuracy compared with a multi-classifier system.

The proposed system contains three neural networks, to which three groups of features extracted from the same DNA sequence are presented. The outputs of the individual neural networks are then passed through a probability function or decision logic in order to provide an answer as to whether the presented sequence is a coding or non-coding.

A. Neural Network Classifiers

Neural networks have extensively been used in bioinformatics. Multi-layer neural networks trained using the back-propagation algorithm have extensively been used in bioinformatics [10, 12]. The configuration of neural network classifiers (*NNC*) and its input features is shown in Table I. For all networks, a multi-layer perceptron (MLP) with two hidden layers and a network size of 8-8-4-1 was chosen, as this configuration provides good classification and efficient network training.

TABLE I. Configurations of Three Neural Network Based Classifier and Their Input Features.

Network	Input Features ν
NNC_1	$SNR_\ell[k], k \cong \left\{\frac{N}{3}, \frac{N}{10.5}\right\}, \ell \in \{A,C,G,T\}$
NNC_2	μ_ℓ , σ_ℓ^2 , $\ell \in \{A,C,G,T\}$
NNC_2	γ_ℓ , κ_ℓ^2 , $\ell \in \{A,C,G,T\}$

All neurons in both hidden layers have tan-sigmoid transfer functions. The output neuron has a purely linear transfer function. Empirical work found that a further increase in the number of neurons in each layer did not improve the performance of the classifier. Increasing the number of neurons further increases the risk that the networks will overfit the training data.

The output from the neural network classifier is then input to a transfer function that transfers the output of the network into an assertion of the form coding/non-coding region. This function can have various forms depending on the kind of output that the networks produce. In this paper, the networks were trained to produce the value of 1 when faced with a coding and the value of -1 when faced with a non-coding region. In this case the probability function is a simple hard limiter function as follows,

$$hl(y,\tau) = \begin{cases} 1 & \text{if } x > \tau \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

where τ is the threshold value.

All three networks were trained using the training set as defined in Section III.B. Training was said to be complete when the mean square error of the network fell below 0.001 of the training data. The resilient backpropagation algorithm [22]

was utilized to train the networks. This algorithm allows timely training of the networks, because it is especially designed and adapted to work well with multilayered networks with sigmoid transfer functions.

B. Data Sets

The HMR195 dataset [8] contains 195 mammalian sequences with exactly one complete either single-exon or multi-exon gene. The dataset was developed to evaluate different gene-finding programs. All sequences contain exactly one gene which starts with the 'ATG' initial codon and ends with one of the stop codons, i.e. 'TAA', 'TAG', or 'TGA'. There are no in-frame stop codons in coding genes, and introns of multi-exons genes start with dinucleotides 'GT' and end with dinucleotide 'AG'. Sequences longer than 200,000 bp are not included in the set. In this dataset, the ratio of *human:mouse:rat* sequences is 103:82:10, with a mean length of 7096 bp. The set contains 43 single-exon genes, and 152 multi-exon genes. The proportion of coding regions in the sequences is 14% and the mean exon length is 208 bp.

In this paper, the HMR195 dataset was divided into 117 training set sequences (60%) and 78 testing set sequences (40%). The single and multi-exon sequences and human/mouse/rat sequences were evenly and randomly distributed into the training and testing sets. The training set had length 786338 bp, while the testing set had length 603400 bp.

C. Implementation of Multi-classifier System

The basic concept of the multi-classifier system (MCS) is that the shortcomings of one classifier will be compensated by several others, so that the combined classification result will be more accurate than that of a single classifier by itself. The application of a multi-classifier system for DNA analysis has been exhaustively described in [15]. As shown in Figure 3, the decision logic takes all the probability output from three neural networks and combines them into the final result of the overall system.

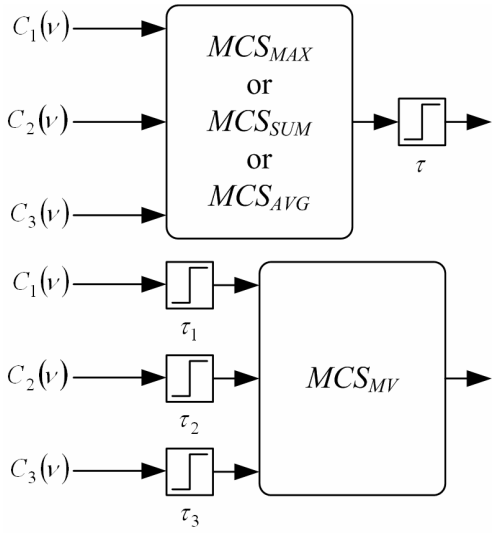

Figure 4. Multi-classifier system using various methods

There are many ways to combine the results of the individual classifiers. In this paper, the most frequently used aggregation functions such as maximum, summation, average, majority voting, and neural network, are explained below. Suppose that we have M classifiers and that $C_m(v), m = 1,\ldots,M$ is the result of a single classifier, while v is the input feature vector.

Figure 4 shows the multi-classifier system using various aggregation methods. The final output is obtained by passing the output from MCS to a transfer function as described in equation (7) with the optimum threshold τ evaluated empirically.

1) Maximum

In this method, the prediction result of the classifier with the highest score is chosen as follows:

$$MCS_{MAX}(v) = \max(w_1 C_1(v),\ldots,w_M C_M(v)) \tag{8}$$

where $MCS_{MAX}(v)$ is the prediction result of the multi-classifier system using the maximum combination method, and w_m is the confidence weight. For simplicity, we set $w_m = 1, \forall m$.

2) Summation

In this method, the sum of all scores achieved by a single classifier is used as follows:

$$MCS_{SUM}(v) = \sum_{m=1}^{M} w_m C_m(v) \tag{9}$$

where $MCS_{SUM}(v)$ is the prediction result of the multi-classifier system using the summation method.

3) Average

In this method, the average of all scores achieved by a single classifier is chosen as follows:

$$MCS_{AVG}(v) = \frac{1}{M}\sum_{m=1}^{M} w_m C_m(v) \tag{10}$$

where $MCS_{AVG}(v)$ is the prediction result of the multi-classifier system using the average combination method.

4) Majority Voting

In this method, the combination of all scores is achieved by following the opinion of the majority of the classifiers. For this method, the output of the neural network classifier is transferred to coding/non-coding ("1" or "0") assertion by a probabilistic function. The result obtained using majority voting is then obtained by

$$MCS_{MV}(v) = \arg\max_{j=1}^{J} \frac{1}{M}\sum_{m=1}^{M} w_m C_m^j(v) \tag{9}$$

where $MCS_{MV}(v)$ is the prediction result of the multi-classifier system using the majority voting combination method, J denotes the number of classes, and $C_m^j(v)$ denotes the certainty of classifier m that input v belongs to class j. Note that, to avoid deadlock, the number of classifiers M should be an odd number.

5) Neural Network

In this method, another neural network is used to obtain the combined result from the output of neural network classifier. In this paper, the multi layer perceptron with a 3-3-1 configuration was utilized as shown in Figure 5.

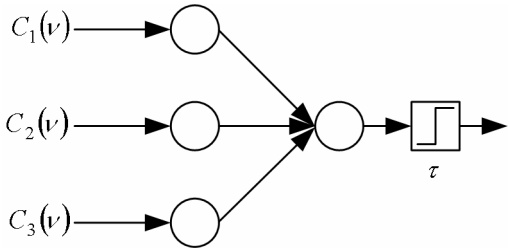

Figure 5. Neural network configuration for the multi-classifier system

The hidden and output layers have pure linear transfer function. The network is trained using the resilient backpropagation algorithm until a mean square error of 0.001 is achieved.

IV. Results and Discussions

In this section, the performance of each individual classifier and the multi-classifier system is evaluated. First, the performance metrics are described. Then, the performance evaluation of individual classifiers is presented. Finally, the performance evaluation of the multi-classifier system is discussed.

A. Evaluation Measures

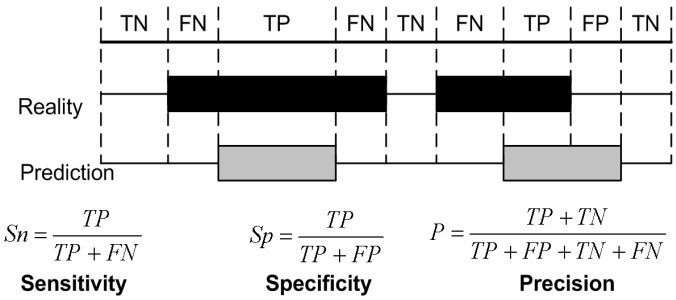

Figure 6. Nucleotide-level measures of prediction accuracy

To evaluate the performance of gene identification, we used prediction accuracy measures similar to [23], as shown in Figure 2. True positive (TP) is the number of coding nucleotides correctly predicted as coding. False negative (FN) is the number of coding nucleotides predicted as non-coding. True negative (TN) is the number of non-coding nucleotides correctly predicted as non-coding. False positives (FP) is the number of non-coding nucleotides predicted as coding. The sensitivity (Sn) provides a measure of the proportion of coding nucleotides that have been correctly predicted as coding. The specificity (Sp) provides the proportion of predicted coding nucleotides that are actually from the coding region. Both Sn and Sp can be viewed as conditional probabilities. Finally, the precision (P) shows the recognition rate of the classifier.

B. Performance Evaluation of Individual Classifiers

The test sequences described in Section III.B were passed through the three neural networks (NNC_n) individually, and the performance results are summarized in Figure 7.

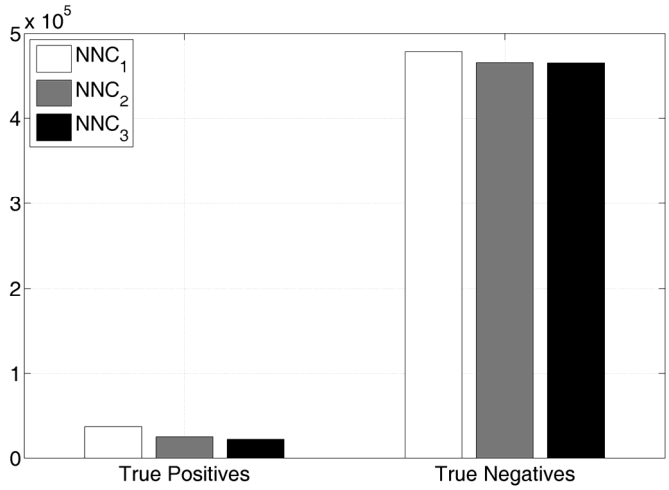

Figure 7. True positives and true negatives recognised by the individual neural networks

Table II shows the performance in terms of specificity (Sn), sensitivity (Sp), and precision (P). Note that the probabilistic function used for each neural network was hard limiter (see equation (7)), for which the optimum threshold values are $\tau_1 = -0.54$, $\tau_2 = -0.49$, and $\tau_3 = -0.47$. From Table II, we can see that the higher order statistics features (NNC_2 and NNC_3) provide a comparable performance with the well-known Fourier analysis features (NNC_1). Hence, the use of HOS features in gene identification is validated.

TABLE II. Performance of Individual Neural Network Classifiers on The HMR195 Test Dataset

Network	Specificity (*Sp*)	Sensitivity (*Sn*)	Precision (*P*)
NNC_1	0.457	0.451	0.854
NNC_2	0.310	0.302	0.813
NNC_3	0.275	0.276	0.808

By combining the HOS features with Fourier analysis features, we expect that a higher recognition rate can be achieved. Two conditions need to be met in order for the application of multiple classifiers to be successful [15]. Firstly, the performance of all classifiers individually needs to exceed 50%. Secondly, the individual classifiers need to be sufficiently different from each other.

C. Performance Evaluation of the Multi-classifier System

The classification performances on the HMR195 test set, including the specificity, sensitivity, correlation, and precision of the combined system, using various aggregation functions described in Section III.C are tabulated in Table III.

TABLE III. PERFORMANCE COMPARISON OF THE MULTI-CLASSIFIER SYSTEM USING VARIOUS AGGREGATING METHODS

Method	*Sp*	*Sn*	*P*	τ
Maximum	0.427	0.436	0.847	-0.34
Summation	0.443	0.444	0.852	-1.6
Average	0.448	0.440	0.854	-0.53
Majority Voting	0.300	0.340	0.830	n/a
Neural Network	0.480	0.472	0.862	-0.51

By comparison with Table II, the combination of these three neural networks provides an improved recognition rate in terms of precision. The MCS using a neural network aggregation function provides the best result, while MCS using majority voting provides the poorest result. The results obtained by the individual neural networks and the MCS using neural network are compared in Figure 8. Furthermore, the performance of the proposed MCS using neural network are compared with the NNPP algorithm [24]. Table IV shows that our algorithm outperforms the NNPP algorithm in terms of specificity, sensitivity, and precision.

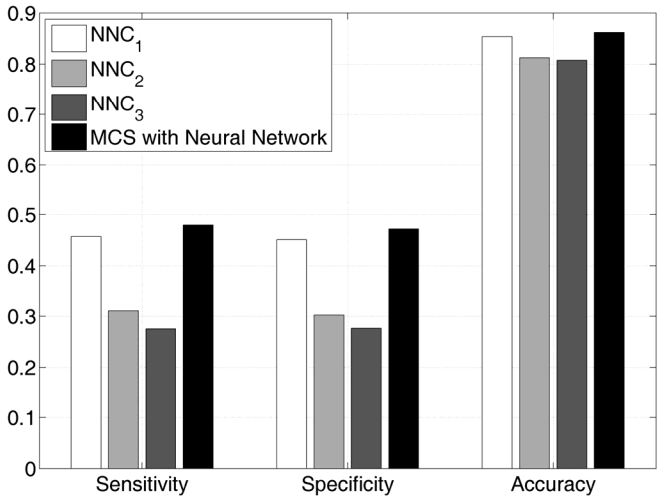

Figure 8. Comparison of the specificity, sensitivity and precision of the three neural networks and the multi-classifier system

TABLE IV. PERFORMANCE COMPARISON OF THE PROPOSED MULTI-CLASSIFIER SYSTEM WITH NNPP ALGORITHM [24].

Method	Specificity (*Sp*)	Sensitivity (*Sn*)	Precision (*P*)
NNPP Algorithm	0.086	0.047	0.806
MCS Neural Net	*0.480*	*0.472*	*0.862*

V. CONCLUSION

In this paper, we presented a novel approach for the recognition of protein coding and non-coding regions in mammalian DNA sequences. Higher order statistics and Fourier analysis features were utilized for the individual neural networks. Evaluation of the proposed system on the HMR195 database revealed that the recognition accuracy of the multi-classifier system can be increased by 5% over that of the individual neural networks. Future work will include the optimization of current neural networks, the application of other classifiers such as support vector machines (SVM), Gaussian mixture models (GMM), hidden Markov models (HMM), and the identification and use of other discriminative features.

ACKNOWLEDGMENT

This research is fully supported by the University of New South Wales, Australia, Faculty Research Grant, 2007 for genomic signal processing research.

REFERENCES

[1] V. V. Solovyev, A. A. Salamov, and C. B. Lawrence, "Identification of human gene structure using linear discriminant functions and dynamic programming," in *Proceedings of the Third International Conference on Intelligent Systems for Molecular Biology*, Menlo Park, CA, pp. 367-375, 1995.

[2] A. V. Lukashin and M. Borodovsky, "GeneMark.hmm: New solutions for gene-finding," *Nucleid Acids Research*, vol. 26, pp. 1107-1115, 1998.

[3] D. Kulp, D. Haussler, M. G. Reese, and F. H. Eeckman, "A generalized hidden Markov model for the recognition of human genes in DNA," in *Proceedings of the Fourth International Conference on Intelligent Systems for Molecular Biology*, Menlo Park, CA., pp. 134-142, 1996.

[4] C. Burge, *Identification of complete structure in human genomic DNA*, PhD Thesis, Stanford, CA, Stanford University, 1997.

[5] A. Krogh, "Two methods for improving performance of an HMM and their application for gene-finding," in *Proceedings of the Fifth International Conference on Intelligent Systems for Molecular Biology*, Menlo Park, CA, pp. 179-186, 1997.

[6] S. Salzberg, A. Delcher, K. Fasman, and J. Henderson, "A decision tree system for finding genes in DNA," *Journal of Computational Biology*, vol. 5, pp. 667-680, 1998.

[7] M. Q. Zhang, "Identification of protein coding regions in the human genome by quadratic discriminant analysis," *Proceedings of the National Academy of Sciences USA*, vol. 94, pp. 565-568, 1997.

[8] S. Rogic, A. K. Mackworth, and F. B. F. Ouellette, "Evaluation of gene-finding programs on mammalian sequences," *Genome Research*, vol. 11, pp. 817-832, 2001.

[9] P. P. Vaidyanathan and B.-J. Yoon, "Digital filters for gene prediction applications," in Proc. Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, California, USA, pp. 306-310, 2002.

[10] E. C. Uberbacher and R. J. Mural, "Locating protein-coding regions in human DNA sequences by a multiple sensor-neural network approach," *Proceedings of the National Academy of Sciences USA*, vol. 88, pp. 11261-11265, 1991.

[11] J. W. Fickett, "The gene identification problem: An overview for developers," *Computers and Chemistry*, vol. 20, pp. 103-118, 1996.

[12] E. E. Snyder and G. D. Stormo, "Identification of protein coding regions in genomic DNA," *Journal of Molecular Biology*, vol. 248, pp. 1-18, 1995.

[13] D. Anastassiou, "Genomic Signal Processing," *IEEE Signal Processing Magazine*, vol. 18, pp. 8-20, 2001.

[14] J. F. Mendel, "Tutorial on higher-order statistics (spectra) in signal processing and system theory: Theoretical results and some applications," *Proceedings of the IEEE*, vol. 79, pp. 278-305, 1991.

[15] R. Ranawana and V. Palade, "A neuro-genetic framework for multi-classifier design: an application for promoter recognition in DNA sequences," *International Journal of Hybrid Intelligent Systems*, vol. 3, pp. 35-61, 2006.

[16] E. N. Trifonov, "3-, 10.5-, 200- and 400-base periodicities in genome sequences," *Physica A: Statistical and Theoretical Physics*, vol. 249, pp. 511-516, 1998.

[17] R. F. Voss, "Evolution of long-range fractal correlations and 1/f noise in DNA base sequences," *Physical Review Letters*, vol. 68, pp. 3805, 1992.

[18] M. Akhtar, J. Epps, and E. Ambikairajah, "On DNA numerical representation for period-3 based exon prediction," in *Proc 5th International Workshop on Genomic Signal Processing and Statistics*, pp., 2007.

[19] S. Datta, A. Asif, and H. Wang, "Prediction of protein coding regions in DNA sequences using Fourier spectral characteristics," in *Proc. IEEE 6th International Symposium on Multimedia Software Engineering*, pp. 160-163, 2004.

[20] L. Kumar, M. Futschik, and H. Herzel, "DNA motifs and sequence periodicities," *In Silico Biologuy*, vol. 6, pp. 71-78, 2006.

[21] E. R. Dougherty, I. Shmulevich, J. Chen, and Z. J. Wang, *Genomic Signal Processing and Statistics*, Hindawi Publishing Corporation, 2005.

[22] M. Riedmiller and H. Braun, "A direct adaptive method for faster backpropagation learning: the RPROP algorithm," in IEEE International Conference on Neural Networks, San Fransisco, pp. 586-591, 1993.

[23] M. Burset and R. Guigo, "Evaluation of gene structure prediction programs," *Genomics*, vol. 34, pp. 353-367, 1996.

[24] M. G. Reese, "Application of a time-delay neural network to promoter annotation in the Drosophila melanogaster genome," *Computers and Chemistry*, vol. 26, pp. 51-56, 2001.