# Optimization method of protein coding region identification based on IHHO-CNN-LSTM

**Siyuan Wu**[1,3]**, Tingting Yang**[2,4,*]

[1] Pomfret School, Pomfret, United States of America

[2] Technology Innovation Center at TAIKANG LIFE INSURANCE Co., Ltd.

[3] name-eman@163.com

[4] Mobphyspde100@oulook.com

[*] corresponding author

**Abstract.** Aiming at the current problem of insufficient i dentification ac curacy of co ding re gions in DNA sequences, this study proposes a protein coding region identification m ethod b ased o n IHHO-CNN-LSTM. Firstly, the data preprocessing of DNA sequences is transformed into feature vectors, and then the protein coding region identification model based on CNN-LSTM i s e stablished. To address the limitations of parameter selection of CNN-LSTM, a hybrid strategy improved Harris Hawk Optimization (HHO) algorithm is introduced to achieve adaptive parameter searching of CNN-LSTM, so as to obtain the optimization model of white matter coding region identification b ased o n I HHO-CNN-LSTM. The improved model was used to accurately distinguish coding and non-coding regions. Two benchmark datasets, HMR195 and BG570, are selected for five-fold c ross-validation, a nd t he r esults s how t hat the AUC values of the model designed in this paper are 0.9854 and 0.9895, the corresponding identification accuracy is 0.9527 and 0.9645, respectively, which are significantly b etter t han o ther m odels, a nd also have a significant advantage i n t erms o f c omputational e fficiency. Th e pr oposed me thod ca n efficiently and accurately identify protein coding regions, which can help promote the related research in the field of genetic engineering.

**Keywords:** protein coding region identification; CNN-LSTM; Harris Hawk Optimization algorithm; hybrid strategy

## 1. Introduction

The vast majority of eukaryotes now have sequences coding for proteins separated by non-coding sequences in the gene sequence. The sequences that can code are also called exons, which conserve genetic information. The sequences that cannot play a non-coding role are also known as Intron. The accurate identification of Exon and Intron is also helpful for the construction of genetic engineering, and also has a certain auxiliary role in the study of gene function in the field of biology, so it is of great significance to carry out the research on the identification method of Exon and Intron [1]. Various methods have been proposed for exon identification in academia so far. Generally, they can be categorized into methods based on homology comparison and methods that do not rely on homology comparison. The method based on sequence homology is to use the existing gene database as the standard to recognize the similarity of DNA sequences to be tested, so as to judge the exon and intron regions of unknown sequences based on the existing experience [2]. BLAST and MUSCLE are

common comparison tools, and in recent years, gene prediction programs such as GeMoMa have also been proposed. Sequence homology-based methods are more accurate, but the high cost of sequencing and the efficiency of comparison have constrained the development of this technology. Based on this, many scholars have shifted their research focus to models that do not rely on comparison techniques. Digital signal processing technology plays a key role in this field. Before digital signal processing, DNA sequences usually need to be mapped numerically. The rapid development of machine learning also provides new solutions for the identification of protein coding regions. Although many identification methods have been proposed, the evaluation parameters of their performance in practical applications have yet to be improved [3-5]. In this paper, based on the data preprocessing of exons and introns, the obtained feature vectors are inputted into the CNN-LSTM classification model for training, and the Harris Hawk Optimization (HHO) algorithm, which is improved by the hybrid strategy, is used for parameter searching of CNN-LSTM, and it is verified through experiments that the designed IHHO-CNN-LSTM model can realize the identification of exons and introns of eukaryotic organisms efficiently and accurately and it has a certain promotion effect on the construction of genetic engineering.

## 2. Data sources and pre-processing

### 2.1. Data sources

In this paper, we analyzed the coding region discrimination of eukaryotic DNA sequences, and two benchmark datasets were used, namely, HMR195 and BG570. HMR195 consists of 195 mammalian DNA sequences, including human, mouse, and rat, with a total of 948 exons, and BG570 refers to a genome test dataset of 570 vertebrate sequences, with a total of 2,649 exons. sequences, totaling 2,649 exons.BG570 refers to a genomic test dataset consisting of 570 vertebrate sequences. The two datasets can be accessed at http://www. imtech.res.in/raghava /genebenchraghava/genebench. The length range of the benchmark dataset, number of exons and introns are shown in Table **??**. In order to ensure the classification of exons and introns, the short (less than 20 bp) exons and introns were classified. bp) exon and intron sequences were also included in the experiment to ensure the comprehensiveness of exon and intron classification.

**Table 1.** The datasets

| Data | Sequence number | Length range | Exons | Introns |
|------|-----------------|--------------|-------|---------|
| HMR195 | 195 | 795 56500 | 948 | 1143 |
| BG570 | 57 | 398 36845 | 2649 | 3211 |
| Total | 765 | 398 56500 | 3597 | 4354 |

### 2.2. Data pre-processing

Before realizing the precise identification of exons and introns, it is necessary to preprocess the DNA sequence. In this paper, we adopt the numerical mapping method to transform the DNA sequence into a numerical representation, and combine with the k-mer technique to transform the exons and introns in the DNA sequence into a feature vector respectively. The process of extracting the feature vector of the data is as follows: Given an exon ACAGCGACC: In step 1, starting from the first nucleotide A, the exon is converted into a specific amino acid sequence by moving one nucleotide at a time, specifically, 'ACA' corresponds to the amino acid T, 'CAG' corresponds to the amino acid Q, and 'CAG' corresponds to the amino acid B. In step 2, the exon is converted into a specific amino acid sequence by moving one nucleotide at a time. CAG' corresponds to amino acid Q, 'AGC' corresponds to amino acid S, 'GCG' corresponds to amino acid A, 'CGA' corresponds to amino acid R, 'GAC' corresponds to amino acid A,

and 'GAC' corresponds to amino acid A. 'GAC' corresponds to amino acid D, 'ACC' corresponds to amino acid T, and thus a specific amino acid sequence is obtained as TQSARDT; in the second step, the classical relative frequency of k characters technique is combined, and the k value is specified to vary from 1 to 5 to transform TQSA RDT into a feature vector.

## 3. Design of protein coding region identification model

### 3.1. CNN-LSTM

Convolutional Neural Network (CNN) is a deep feedforward network that includes convolutional structures due to its powerful feature extraction ability. CNN can extract the relationships between multidimensional time series data in spatial structures, consisting of input layers, convolutional layers, pooling layers, fully connected layers, and output layers. Among them, the feature extraction of convolutional layers is mainly carried out through convolutional kernels. The pooling layer is mainly used for feature dimensionality reduction, reducing the number of parameters, and preventing overfitting. After being processed by convolutional and pooling layers, the features enter the fully connected layer and are further integrated, ultimately transforming into one-dimensional vectors [6]. Long Short Term Memory (LSTM) Neural Network is an improved recurrent neural network. By using Gate to determine the fate of information on a sequence, it is possible to remember information that needs to be memorized for a long time, filter out unimportant information, and solve the long-term dependency problem of recurrent neural networks (RNN). After its proposal, it was also improved by adding additional forget gates. The improved long short-term memory neural network solves the problem of "gradient vanishing" in model training and can learn time series long short-term dependency information. It is currently the most successful recurrent neural network architecture and has been applied in many scenarios[7].

Due to the good feature extraction ability of CNN and the significant advantage of LSTM in dealing with time series problems, as well as the ability to avoid gradient vanishing, this study chooses to combine CNN with LSTM to construct a protein coding region identification model. The specific structure of CNN-LSTM model is shown in Figure 1. The first half of the model is CNN, which is used for feature extraction. The extracted information is processed by the max pooling layer and Dropout layer to effectively prevent overfitting. The latter part of the model is LSTM, which is used for data classification. The CNN -LSTM model has hyperparameters such as convolutional layer kernel size, number of convolutional kernels, number of LSTM layer neurons, batch size, etc [8]. The selection of these hyperparameters will significantly affect the performance of the model. Therefore, this article uses the hybrid strategy improved Harris Eagle Optimization (IHHO) algorithm to optimize the CNN LSTM model, find the optimal hyperparameters, and improve the identification accuracy of the model.

### 3.2. Improvement of HHO

HHO is a new type of swarm intelligence optimization algorithm proposed in recent years, inspired by the three-stage predatory method of search, search and mining switching, and mining in the predatory behavior of birds like Harris eagles. This algorithm has a simple principle, few parameters, and strong global search ability [9]. Once proposed, it has been widely applied in many fields. However, like conventional intelligent optimization algorithms, the standard HHO algorithm still has shortcomings such as difficulty in improving optimization accuracy, slow convergence speed, and susceptibility to falling into local optima when solving high-dimensional complex optimization problems [10]. This paper design an improved Harris Hawk algorithm (IHHO), which first utilizes the random traversal of chaotic systems to improve the initial structure of the population, and proposes a population initialization mechanism based on chaotic Tent mapping to enhance population diversity. Further nonlinear periodic adjustments are made to the energy factor to depict the multi round game of "encirclement escape" between Harris eagles and prey, utilizing multiple rounds of global search and smooth switching between local mining to improve the optimization ability; Finally, the small hole imaging opposition learning mechanism is used to mutate elite individuals in the population, expand the search space, and avoid generating local optimal
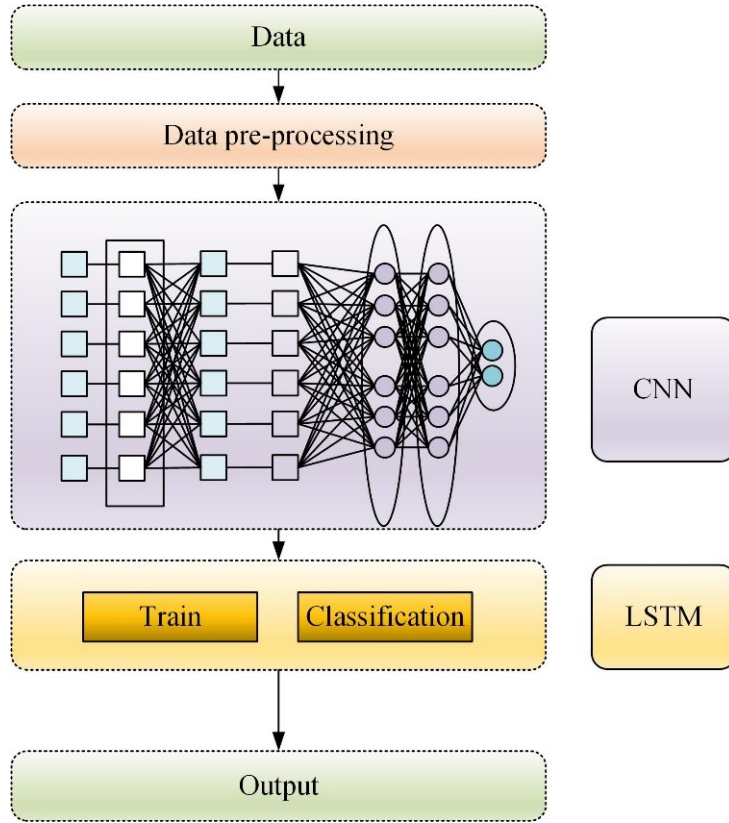
**Figure 1.** The specific structure of CNN-LSTM model

solutions. Using the improved algorithm IHHO to optimize CNN-LSTM, and applying the fusion model to the identification of protein coding regions.

*3.2.1. HHO*   In the HHO algorithm, the position of the Harris eagle individual can be regarded as a candidate solution to the problem, iteratively approaching the optimal solution by utilizing the movement in position. The algorithm itself can still be divided into global search and local mining [11].
(1) Global search
In this stage, individuals will randomly wait at a certain position, search for prey with keen vision in their search space, and search for prey through two equally probable random behaviors. The specific mathematical model is as follows:

$$X(t+1) = \begin{cases} X_{\text{rand}}(t) - r_1 \left| X_{\text{rand}}(t) - 2r_2 X(t) \right|, & \text{if } q \geq 0.5 \\ (X_{\text{rabbit}}(t) - X_m(t)) - r_3 \left( lb + r_4(ub - lb) \right), & \text{if } q < 0.5 \end{cases} \tag{1}$$

which $X_{(t)}$ is the original position of the Harris eagle, $X_{(t+1)}$ is the updated position, $X_{\text{rand}}(t)$ is the randomly selected individual position, $X_{\text{rabbit}}(t)$ is the optimal solution position, $q, r_1, r_2, r_3,$ and $r_4$ are random quantities between (0,1), [lb, ub] is the upper and lower bounds on the search space of the Harris eagle, and is the average position of the population during t-th iteration. If N is the population size, then

$$X_m(t) = \frac{1}{N} \sum_{i=1}^{N} X_i(t). \tag{2}$$

(2)Switching from search to mining

The HHO algorithm utilizes the escape energy factor E to switch from global search to local mining, which is defined as

$$E = 2E_0 \left(1 - \frac{t}{T_{\max}}\right).$$ (3)

In the formula: $E_0$ is the initial energy of the prey, $E_0 = 2$ and -1 , $rand \in (0,1)$ is a random quantity, and $T_{max}$ is the maximum number of iterations. If $|E| \geq 1$, the algorithm enters the global search stage. Otherwise, it will enter local mining. (3)Local mining During this stage, individuals will adopt a surprise attack approach to attack their prey. Specifically, the HHO algorithm will update its position using four strategies to simulate attack behaviour. $\lambda$ is the probability of escape, if $\lambda < 0.5$, then the prey successfully escapes. If $\lambda > 0.5$, then the prey fails to escape. The escape energy factor E represents the individual's soft and hard attack strategies. If $|E| > 0.5$, soft surround is executed. If $|E| < 0.5$, then execute hard surround. There are four predatory strategies as follows: ① Soft surround. If $|E| > 0.5$ and $\lambda > 0.5$ indicates that the target has sufficient energy to escape, and surrounding will consume prey energy. Individuals will gradually choose the optimal position to prey on. The location update method is

$$
\begin{aligned}
X(t+1) &= \Delta X(t) - E[JX_{\text{rabbit}}(t) - X(t)], \\
\Delta X(t) &= X_{\text{rabbit}}(t) - X(t), \\
J &= 2(1 - r_5).
\end{aligned}
$$ (4)

$\Delta X(t)$ is the distance between the position of the prey and the individual Harris eagle, r5 is the random quantity between (0,1), and J is the jumping distance during the prey's escape process.

② Hard surround. If $|E| < 0.5$ and $\lambda > 0.5$ indicates that the prey does not have enough energy to escape, and the Harris eagle will quickly hunt the prey. Its location update method:

$$X(t+1) = X_{\text{rabbit}}(t) - E\left[\Delta X(t)\right].$$ (5)

③ Asymptotically fast diving soft surround. If $|E| > 0.5$ and $\lambda < 0.5$ indicates that the target has sufficient energy to escape, and the individual will intelligently establish a soft surround and encircle the prey through the following two strategies. The specific location update method is

$$X(t+1) = \begin{cases} Y : X_{\text{rabbit}}(t) - E[JX_{\text{rabbit}}(t) - X_m(t)], & \text{if } F(Y) < F(X(t)) \\ Z : Y + S \times LF(D), & \text{if } F(Z) > F(X(t)) \end{cases}$$ (6)

D is the dimension, S is the D-dimensional random row vector, and LF is the Levi's flight function.

④ Asymptotically fast diving hard encirclement. If $|E| < 0.5$ and $\lambda > 0.5$ indicates that the prey has less energy, and the Harris eagle will intelligently establish a hard surround and use the following two strategies to surround the prey. The specific location update method is

$$X(t+1) = \begin{cases} Y : X_{\text{rabbit}}(t) - E[JX_{\text{rabbit}}(t) - X(t)], & \text{if } F(Y) < F(X(t)) \\ Z : Y + S \times LF(D), & \text{if } F(Z) < F(X(t)) \end{cases}$$ (7)

*3.2.2. Algorithm improvements* (1) Population initialization based on chaotic Tent mapping The initial population has a significant impact on the convergence speed and optimization accuracy of swarm intelligence algorithms, and a diverse initial population structure can effectively promote algorithm optimization. However, the HHO algorithm uses random methods for population initialization, which results in insufficient spatial traversal and reduces search efficiency. This article introduces the Tent chaos mechanism in the HHO algorithm, and its mapping method is as follows:

$$y_{k+1} = \begin{cases} y_k \cdot 0.7, & y_k \leq 0.7 \\ \left(\frac{10}{3}\right)(1 - y_k), & y_k > 0.7 \end{cases}$$ (8)

After generating chaotic values, the mapping rule between chaotic values and the population search space is:

$$x_{i,j} = lb_j + v_{i,j} \times (ub_j - lb_j). \tag{9}$$

$x_{i,j}$ is the positions of individual i in dimension j, $ub_j$ $lb_j$ are the position boundaries of individual i in dimension j, $j = 1, 2, ..., d$, d is the position dimension. (2) Nonlinear periodic adjustment mechanism of energy factors The escape energy factor E achieves the switching from global search to local mining, which is an important indicator to measure the optimization ability of the HHO algorithm. E is linearly decreasing in a single cycle during iteration, which does not conform to the natural law that Harris eagle populations require multiple rounds of collaborative hunting of prey. This article introduces a nonlinear periodic adjustment strategy for energy factors to describe the multi round nature of the Harris eagle's game of hunting prey. Specifically, in the update equation of E, the cosine function is used to describe nonlinear periodicity, which is defined as:

(2) Nonlinear periodic adjustment mechanism of energy factors The escape energy factor E achieves the switching from global search to local mining, which is an important indicator to measure the optimization ability of the HHO algorithm. E is linearly decreasing in a single cycle during iteration, which does not conform to the natural law that Harris eagle populations require multiple rounds of collaborative hunting of prey. This article introduces a nonlinear periodic adjustment strategy for energy factors to describe the multi round nature of the Harris eagle's game of hunting prey. Specifically, in the update equation of E, the cosine function is used to describe nonlinear periodicity, which is defined as:

$$E = 2E_0 \left[ 1 - \frac{t}{T_{\max}} \right] \cdot \cos \left[ 2k + \frac{\pi}{2} \frac{t}{T_{\max}} \right]. \tag{10}$$

k is the number of decreasing cycles, and Tmax is the maximum number of iterations. It can be seen that when E approaches 0 periodically, it indicates that the Harris eagle will approach and prey on prey with probability, achieving multiple rounds of global search and local mining processes. (3) Small hole imaging adversarial learning mechanism During the iteration process of the HHO algorithm, individuals in the population tend to approach the global optimal solution region in the later stage of iteration, resulting in a loss of population diversity, iteration stagnation, and ultimately obtaining local optimal solutions. This is also an inherent limitation of intelligent optimization algorithms. Therefore, this article introduces an adversarial learning mechanism based on small hole imaging to mutate and perturb the current optimal solution, enabling the algorithm to have the ability to jump away from local optima. The adversarial learning solution for small hole imaging in d-dimensional space is

$$X'_j = \frac{lb_j + ub_j}{2} + \frac{lb_j + ub_j}{2\delta} \cdot \frac{X_j}{\delta}. \tag{11}$$

$|E| > 0.5$ and $\lambda > 0.5$ The execution steps of the IHHO algorithm are:

Step 1: Initialize the parameters and use the chaotic Tent mapping mechanism to generate the initial population in the search space.
Step 2: Calculate the individual fitness values of the population.
Step 3: Update the prey escape energy factor E and jump distance.
Step 4: If $|E| > 1$, the algorithm enters the global search stage and updates the position of the Harris Eagle using equation (1).
Step 5: If $|E| > 0.5$ and $\lambda > 0.5$, the algorithm enters soft surround and uses equation (4) to update the position of Harris Eagle. Step 6: If $|E| < 0.5$ and $\lambda > 0.5$, the algorithm enters hard surround and updates the position of the Harris eagle using equation (5).
Step 7: If $|E| > 0.5$ and $\lambda < 0.5$, the algorithm enters an asymptotically fast diving soft surround.
Step 8: If $|E| < 0.5$ and $\lambda < 0.5$, the algorithm enters an asymptotically fast diving hard surround.
Step 9: Use the adversarial learning mechanism of small hole imaging to mutate and perturb the current optimal solution, and select the best one to retain.
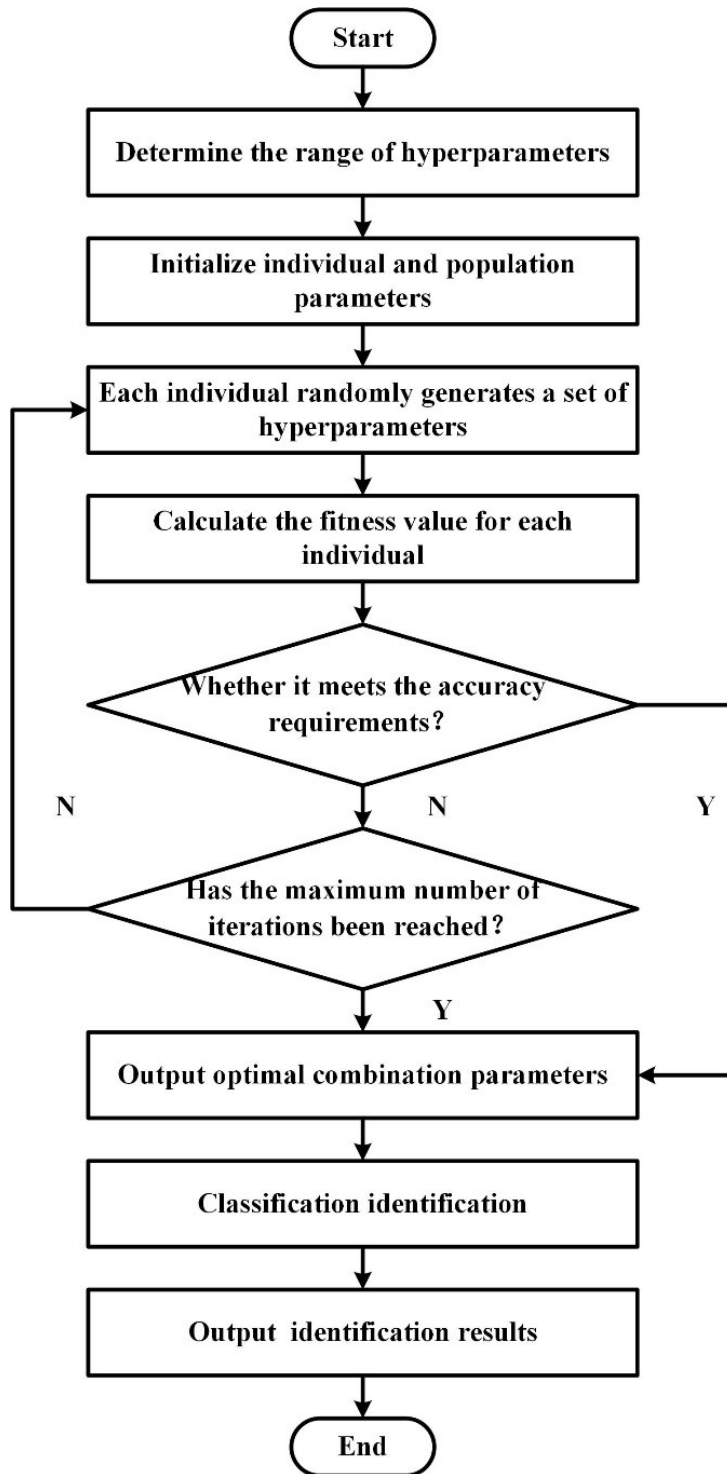Step 10: Update the global optimal solution and its fitness value.

**Figure 2.** The specific steps of IHHO-CNN-LSTM

Step 11: Determine the termination condition of the algorithm. If the maximum number of iterations is reached, the algorithm terminates and outputs the current global optimal solution. Otherwise, proceed to step 3 for execution.

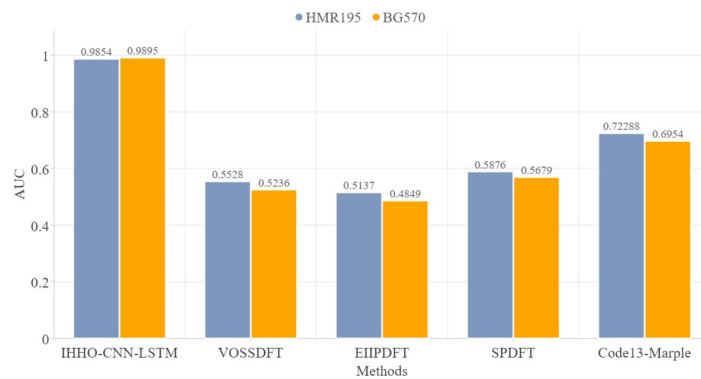The specific steps for optimizing the CNN-LSTM model using the IHHO algorithm are shown in Figure

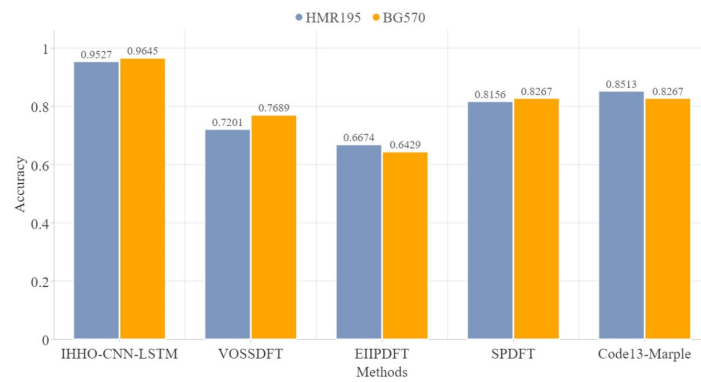**Figure 3.** AUC values of different model identification results



**Figure 4.** Accuracy rate of different model identification results
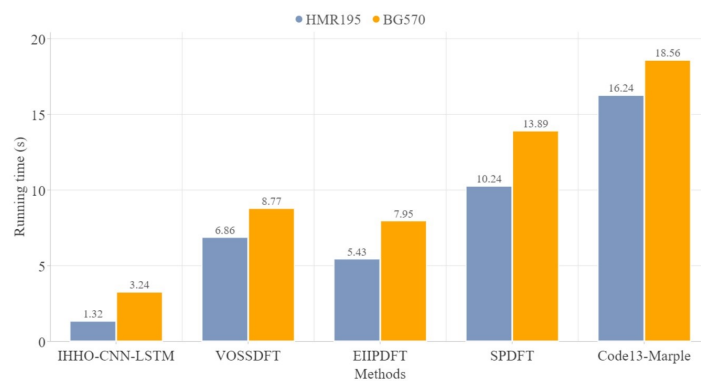


**Figure 5.** Running timeof different model identification results

2. Each Harris eagle individual represents a set of hyperparameters. By calculating fitness values, individuals are sorted and the best performing group is selected as the population for a new round of iteration. The iteration is repeated until the maximum number of iterations is reached or a satisfactory hyperparameter combination is found. The output hyperparameter combination is substituted into CNN-LSTM to ultimately obtain the identification result of the protein coding region.

## 4. Experimental analysis

In order to demonstrate the advantages and disadvantages of the new method in this article compared with other methods, its identification results are compared with classical VOSSDFT, EIIPDFT, SPDFT, and Code13 March. The main evaluation indicators are AUC value, accuracy, and algorithm running time. VOSSDFT [12], EIIPDFT[13], and SPDFT [14] are all based on the Discrete Fourier Transform (DFT) technique to distinguish between exons and introns in eukaryotes. Code13 Marple [15] is an integrated algorithm based on autoregressive spectral analysis and wavelet transform. The calculation of the method in this article is based on a device with an Intel (R) Core (TM) i9-13900H CPU @ 5.40 GHz and 64.0GBRAM processor, programmed using Python 3.8.

From Figure 3, it can be observed that the AUC values of IHHO-CNN-LSTM for the identification of HMR195 and BG570 reached 0.9854 and 0.9895, respectively, far exceeding the maximum values of the other four models and significantly superior to the other three traditional DFT based methods and Code13 Marple. The same situation also occurs in the comparison of identification accuracy in Figure 4. IHHO-CNN-LSTM has high identification accuracy of 0.9527 and 0.9645 for HMR195 and BG570, and also has significant advantages compared to other methods. From the perspective of operational efficiency, it can be seen from Figure 5 that IHHO-CNN-LSTM only takes 1.32 seconds and 3.24 seconds to recognize HMR195 and BG570, which is less time compared to other methods and has an absolute advantage in efficiency.

## 5. Conclusion

This paper proposes a protein coding region identification method based on IHHO-CNN-LSTM, which achieves precise identification of gene exons and introns. Applying the combination model to coding area identification provides a new research perspective for related studies. This article preprocessed two eukaryotic datasets, HMR195 and BG570, using CNN-LSTM as the basic identification model and IHHO algorithm for parameter optimization. As a result, the IHHO-CNN-LSTM identification model was obtained, which achieved good identification performance in data and testing, significantly better than other mainstream methods, and has certain reference value. However, this article has not yet taken the structural information of proteins as an important factor for features, and future research will consider incorporating structural information to further improve the performance of the model. Moreover, further research in this article will expand the sample size and attempt to identify more protein coding regions in more organisms, promoting further progress in biotechnology.

## References

[1] Mudge, J.M.; Jungreis, I.; Hunt, T.; Gonzalez, J.M.; Wright, J.C.; Kay, M.; Davidson, C.; Fitzgerald, S.; Seal, R.; Tweedie, S., et al. Discovery of high-confidence human protein-coding genes and exons by whole-genome PhyloCSF helps elucidate 118 GWAS loci. Genome Res 2019, 29, 2073-2087, doi:10.1101/gr.246462.118.

[2] Zhang, X.; Zhang, K.; Wu, J.; Guo, N.; Liang, J.L.; Wang, X.W.; Cheng, F. QTL-Seq and Sequence Assembly Rapidly Mapped the Gene for the Purple Trait in. Sci Rep-Uk 2020, 10, doi:ARTN 232810.1038/s41598-020-58916-5.

[3] Alfieri, F.; Caravagna, G.; Schaefer, M.H. Cancer genomes tolerate deleterious coding mutations through somatic copy number amplifications of wild-type regions. Nat Commun 2023, 14, doi:ARTN 359410.1038/s41467-023-39313-8.

[4] Badia-Bringué, G.; Canive, M.; Fernandez-Jimenez, N.; Lavín, J.L.; Casais, R.; Blanco-Vázquez, C.; Vázquez, P.; Fernández, A.; Bilbao, J.R.; Garrido, J.M., et al. Summary-data based Mendelian randomization identifies gene expression regulatory polymorphisms associated with bovine paratuberculosis by modulation of the (mediated inflammatory response. Bmc Genomics 2023, 24, doi:ARTN 60510.1186/s12864-023-09710-w.

[5] Yu, R.Z.; Abdullah, S.M.U.; Sun, Y.N. HMMPolish: a coding region polishing tool for TGS-sequenced RNA viruses. Brief Bioinform 2023, 24, doi:10.1093/bib/bbad264.

[6] Bin Syed, M.A.; Ahmed, I. A CNN-LSTM Architecture for Marine Vessel Track AssociationUsing Automatic Identification System (AIS) Data. Sensors-Basel 2023, 23, doi:ARTN 640010.3390/s23146400.

[7] Devi, T.K.; Baluprithviraj, K.N.; Mohan, M.M.; Devi, S.U.; Sakthivel, P.; Rajeshwari, P.; Vinodha, B. Non-invasive method for the prediction of micronutrient deficiency using sequential learning techniques. Comp M Bio Bio E-Iv 2023, 10.1080/21681163.2023.2228914, doi:10.1080/21681163.2023.2228914.

[8] Fan, Y.X.; Xiong, H.; Sun, G.C. DeepASDPred: a CNN-LSTM-based deep learning method for Autism spectrum disorders risk RNA identification. Bmc Bioinformatics 2023, 24, doi:ARTN 26110.1186/s12859-023-05378-x.

[9]  Akpamukcu, M.; Ates, A.; Akdag, O. Combination of electromagnetic field and harris hawks optimization algorithms with optimization to optimization structure and its application for optimum power flow. J Chin Inst Eng 2023, 46, 754-765, doi:10.1080/02533839.2023.2238759.

[10]  Ismail, W.N.; Alsalamah, H.A.; Zhou, X.J.; Nguyen, L.; Zhu, G.H. Efficient Harris Hawk Optimization (HHO)-Based Framework for Accurate Skin Cancer Prediction. Mathematics-Basel 2023, 11, doi:ARTN 360110.3390/math11163601.

[11]  Karnavas, Y.L.; Nivolianiti, E. Harris hawks optimization algorithm for load frequency control of isolated multi-source power generating systems. Int J Emerg Electr P 2023, 10.1515/ijeeps-2023-0035, doi:10.1515/ijeeps-2023-0035.

[12]  Liu, G.C.; Luan, Y.H. Identification of Protein Coding Regions in the Eukaryotic DNA Sequences Based on Marple Algorithm and Wavelet Packets Transform. Abstr Appl Anal 2014, Artn 40256710.1155/2014/402567, doi:Artn 40256710.1155/2014/402567.

[13]  Lehilahy, M.; Ferdi, Y. Identification of exon locations in DNA sequences using a fractional digital anti-notch filter. Biomed Signal Proces 2023, 80, doi:ARTN 10436210.1016/j.bspc.2022.104362.

[14]  Gao, X.H.; Zhang, H.M. Radar coherent integration algorithm for detection of complexly maneuvering target with extended velocity and acceleration scopes. Digit Signal Process 2022, 129, doi:ARTN 10368910.1016/j.dsp.2022.103689.

[15]  Akiyama, Y.; Miyata, H.; Komiyama, M.; Nogami, M.; Ozawa, K.; Oshita, C.; Kume, A.; Ashizawa, T.; Sakura, N.; Mochizuki, T., et al. The identification of affinity peptide ligands specific to the variable region of human antibodies. Biomed Res 2014, 35, 105-116, doi:10.2220/biomedres.35.105.