Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 45 No.4 (2024)

A Hybrid Ant Colony Optimization Algorithm for Human Monkeypox DNA Codon Selection

Akshaya Kumar Mandal [1,*], Nurulla Mansur Barbhuiya[1],[**] and Pankaj Kumar Deva Sarma[1]**

[1]Department of Computer Science, Assam University, A Central University of India, Silchar- 788011, Assam, India

Abstract: **-** Effective codon selection is a pivotal aspect of heterologous gene expression, significantly impact on protein synthesis. Common strategies often rely on the prevalent usage of host genome codons, but concerns persist about their reliability. The substantial asymmetry between gene dimensionality and sample size can result in inaccuracies in disease diagnosis within clinical settings. This paper introduces a modified hybrid ant colony optimization and the support vector machine (ACO-SVM) algorithm, utilize as a classifier on the extracted codon in the monkeypox virus DNA sequence. Experimental outcomes on monkeypox virus DNA datasets reveal that the proposed approach outperforms in recognizing monkeypox virus codon selection. This underscores the efficacy of the modified ACO as a valuable tool for codon selection in the monkeypox virus and the extraction of meaningful information from high-dimensional data. In the context of vaccines design, optimized codons in a viral vector escalate the production of viral antigens, fostering a more potent and effective immune response and ultimately enhancing vaccine efficacy. This research ensures that viral agents are meticulously tuned for optimal efficiency and adaptability across diverse applications, ranging from gene therapy to vaccine development.

Keywords: *ACO-SVM; Monkeypox; DNA sequence; Ant Colony System (ACS); DNA Codon Optimization*

1. Introduction

The development of DNA microarray technology has significantly enhanced the capability to analyze gene expression levels across multiple DNAs simultaneously, enabling molecular-level diagnosis of illnesses such as monkeypox [1, 2]. However, classifying microarray data presents unique challenges due to the vast number of DNA sequences (genes), greatly outnumbering the samples—a phenomenon known as the "curse of dimensionality" leading to overfitting [3]. Successful disease diagnosis relies on selecting a subset of discriminative DNA sequences [4], not only improving classification accuracy but also reducing clinical costs [5] and enhancing interpretability for biologists [6]. Monkeypox virus codon selection is pivotal for developing an effective diagnostic system based on microarray data. Several methods for selecting DNA codons have evolved, broadly categorized as filter (or DNA ranking) and wrapper (or DNA subset selection) approaches [7]. In filter approaches, each DNA is individually assessed based on parameters such as t-statistics, $\chi 2$-statistics, informative gain, signal-noise ratio, Pearson correlation coefficient, or a combination of filtering algorithms [8]. Bioinspired methods, on the other hand, conduct a search in the DNA space, evaluating subsets using a specific classifier's accuracy percentage, in this case, the support vector machine (SVM) [9]. While bioinspired approaches may yield superior classification performance, they come with higher computational costs. In this context, a hybrid ant colony optimization (ACO) algorithm is proposed to efficiently search for the optimal monkeypox virus DNA codon, addressing the limitations of other bioinspired-based approaches like the computational cost of genetic algorithms (GA) and local optima issues with particle swarm optimization (PSO) [15-17]. In 1990, Dorigo et al. [10] successfully adapted ACO, inspired by real ant foraging behavior applied in combinatorial optimization problems. The proposed ACO algorithm is specifically designed for monkeypox virus DNA codon subset selection, with support vector machine (SVM) serving as the classifier. SVM's efficacy

in handling high-dimensional and small-sample data makes it a suitable choice [11, 13-16]. The experimental application of this approach to NCBI monkeypox DNA virus datasets demonstrates excellent codon selection.

2. Related Works

Chiang et al. [11] explores the application of DNA microarrays in cancer classification, addressing the challenge of handling high-dimensional gene expression data. The research introduces an ACO algorithm for gene selection, enhancing the accuracy of classifiers such as multi-layer perceptrons (MLP) and SVM. The experiments, conducted on prostate tumor and human lung carcinoma datasets, demonstrate the effectiveness of ACO-based gene selection, especially in SVM classification with 100 genes. The study compares ACO with other selection methods, highlighting its superior performance. However, the optimal number of selected genes varies with datasets. Overall, the ACO algorithm proves robust in improving classification accuracy, emphasizing its potential in leveraging microarray data for cancer research. Alwan et at. [12], proposed a novel approach, ACOR-SVM, combining ACO with SVM for parameter tuning without discretizing continuous values. Evaluating on seven UCI datasets, the proposed algorithm demonstrates enhanced classification accuracy compared to grid search techniques. Optimal values for regularization parameter (C) and gamma (γ) are provided, showcasing improved accuracy across datasets. The algorithm's efficiency is highlighted by its computational speed. The study suggests future extensions, including simultaneous optimization of SVM parameters and feature subsets using mixed-variable ACO and exploring alternatives like incremental continuous ACO. Overall, ACOR-SVM presents a promising hybrid technique for optimizing SVM parameters with potential applications in various scenarios and further research directions. In this work, Yu et al. [13] proposed a modified ACO method to address the challenge of dimensionality asymmetry in microarray data when selecting tumor-related marker genes. They applied the technique to a subset of the 100 most informative genes and subsequently classified them using SVM. The results, when compared to existing approaches, including GA, demonstrated superior performance in terms of classification accuracy. The modified ACO showed faster convergence, and a further enhancement, modified ACO, effectively balanced intensification and diversification. The study showcased the stability and efficacy of the proposed algorithms across multiple tumor microarray datasets, suggesting their potential in improving disease diagnosis through marker gene selection. Prasad et al. [14] describe the utilization of SVM classifiers augmented by GA, ACO, and PSO to analyze siRNA datasets, including the Huesken, wine, and wdbc breast cancer gene benchmarks datasets. The models demonstrated superior performance over traditional SVM techniques in terms of accuracy. Notably, PSO-SVM outperformed GA-SVM and ACO-SVM in accuracy, particularly on the Huesken dataset. Effective prediction relied on critical feature selection, with both sequential and thermodynamic characteristics playing significant roles. The proposed models, GA-SVM and PSO-SVM, underscored the importance of sequence and thermodynamic properties, while ACO-SVM emphasized sequence features. The study highlighted the importance of optimal feature subsets and reported improved stability in ACO-SVM and PSO-SVM over GA-SVM. Overall, the hybridization of evolutionary computing methods with SVM proved effective in enhancing siRNA efficacy predictions and feature selection. Puigbo et al. [15] proposed an OPTIMIZER which is a cutting-edge online tool for enhancing gene expression by optimizing codon usage. It provides three optimization strategies, including a novel approach to maximizing optimization with minimal sequence modifications. The server utilizes pre-computed codon usage tables from over 150 prokaryotic species, emphasizing highly expressed genes under translational selection. Two key indices, CAI and ENc, assess the optimization process. Outputs include sequence alignments, codon frequency charts, and cleavage site information for selected restriction enzymes. Compared to other tools, OPTIMIZER stands out with its extensive pre-computed reference sets and incorporation of tRNA gene-copy numbers. It is a versatile tool for optimizing gene expression and designing genes with diverse metabolic capabilities in specific species. Srilatha et al. [16] developed a hybrid strategy for diagnosing and classifying brain tumors in MRI images. The strategy comprises pre-processing, feature extraction using local binary pattern and histogram, and classification using SVM and ACO. The experiments, conducted on 158 MR images, demonstrate a classification accuracy of 98.99%, providing efficient and effective brain tumor identification. The system's outcomes indicate its potential

application in other medical imaging contexts. The dataset includes various benign and malignant tumor classes, showcasing the system's capability to distinguish between them. The study emphasizes the importance of confidence-based decision synthesis, offering a novel perspective on automated medical examination.

3. Methods

3.1 Canonical ACO

Dorigo et al.'s 1990 proposed ACO algorithm [10], proven effective in discrete combinatorial optimization problems, is now applied to the path selection of travelling salesman problem (TSP) problem [17]. When a single path search is concluded, an ant's path becomes a potential solution to the Traveling Salesman Problem (TSP). Using the Ant Colony System (ACS) as an example [18, 19], the method initiates by randomly distributing m ants among n cities. Consequently, at time t, the kth ant utilizes a roulette technique with the probability P_{ij}^k, to select the next city on its journey as:

$$P_{ij}^k = \begin{cases} \frac{\tau_{ij}^{\alpha} \eta_{ij}^{\beta}}{\sum_{j \in N_i^k} \tau_{ij}^{\alpha} \eta_{ij}^{\beta}} \\ 0 \quad otherwise \end{cases} \tag{1}$$

In this context, τ_{ij} represents amount of pheromone between city i and city j at time t. The parameter α signifies the relative significance of the pheromones, $\eta_{ij} = \frac{1}{d_{ij}}$ denotes the heuristic factor from city i to city j, d_{ij} represents the distance between city i and city j, β indicates the importance of the heuristic factor, and k signifies the set of cities not yet traversed by the kth ant, i.e., the set of permissible cities.

The ant completes the search for the path, influenced by both pheromones and heuristic values. Throughout the search, ants produce pheromones, contributing to positive feedback within the system and allowing for faster algorithm improvement. The pheromone update technique consists of three models: an ant quality model, an ant density model, and an ant cycle model. The ant cycle model is commonly accepted because it is more oriented towards global information [10, 18, and 19]. The following formula is used to update pheromones using the ant cycle model:

$$\tau_{ij}(t+1) = \rho \tau_{ij}(t) + \sum_{k=1}^{m} \Delta \tau_{ij}^k (t) \tag{2}$$

Where, $\Delta \tau_{ij}^k(t) = \begin{cases} \frac{Q}{L_k} & if\ ant\ k\ select\ edge\ (i.j) \\ 0 & otherwise \end{cases}$

Here, ρ denotes the evaporation coefficient. Increasing ρ heightens the algorithm's randomness, while decreasing it accelerates convergence, albeit with an increased risk of falling into a local optimum. Q represents the total pheromone amount, and (i, j) signifies the path from city i to city j. L_k denotes the length of the path visited by the kth ant in the ongoing iteration.

3.2 Traditional SVM

Vapnik's SVM is a powerful tool in tackling pattern recognition and classification challenges [20]. Unlike traditional methods, SVM have significant advantages such as strong classification capabilities, the absence of local minima, and suitability for small-sample datasets. C = $\{(x_i, y_i) \mid x_i \in R^n, y_i \in \{-1, 1\}, i = 1,\ldots,N\}$ are given a dataset , where x_i is a n-dimensional sample, is the matching class label, N is the number of samples, and the discriminant function of SVM is written as: $S(x) = |(\sum_{i=1}^{p} \propto_i y_i K(x_i, y_i + b)|$. Where, P stands for the number of support vectors, $\propto_i$ represents the Lagrange multiplier, b is the bias of the optimal classification hyperplane, and K denotes the kernel function in this case. The Radial Basis Function (RBF) was used as the kernel in the tests conducted in this research [21]:$K(x_i, x_i) = \exp\{\frac{|x_i - x_j|}{2\sigma^2}\}$.This specific kernel function introduces non-linearity and is commonly employed in SVM for its effectiveness [20].

3.3 Codon Optimization Process

DNA codon optimization involves strategically selecting a gene's codon sequence to enhance its expression or function in a chosen host organism. The process begins by selecting the target gene and identifying the host organism. Utilizing tools like Codon Usage Analyzer and Rare Codon Calculator, researchers analyze codon usage bias and identify rarely used codons. Optimization tools such as OPTIMIZER and COOL are employed to replace rare codons with preferred synonymous ones. Predicting mRNA secondary structure with tools like RNAfold helps avoid stability issues, while optimizing the Kozak sequence using tools like Kozak Sequence Optimization Tool improves translation initiation [22, 23]. Codon pair bias is considered with tools like Codon Pair Bias Calculator to enhance translation efficiency. Customization and refinement of the optimized gene sequence are facilitated by sequence analysis software like Benchling or Geneious. Experimental design tools like SnapGene or ApE aid in planning experiments, and after synthesis, the optimized gene undergoes experimental validation. For an iterative optimization process, propose a hybrid ACO algorithm. Here, we delve into the identification of high-expression genes to delineate the individual codon usage (ICU) selection for the monkeypox virus, enhancing confidence in the efficient expression of the optimized recombinant gene based on experimental insights, ultimately streamlining the entire process for efficient customization to the specific characteristics of the target gene and host organism.

4. Monkeypox Virus DNA codon selection algorithm based on hybrid ACO-SVM

In this research, proposed a hybrid approach for marker monkeypox virus DNA codon selection by integrating hybrid Ant Colony Optimization (ACO) and Support Vector Machine (SVM) as follows:

Step 1.Set the initial pheromone levels across all routes.
Step 2.Each ant uses equation (8) to generate various feature subsets by randomly searching a path from the nest to the food sources.
Step 3. Evaluate the fitness of each feature subset acquired in step 2 through SVM analysis, and compare the best-performing subset with previous search outcomes, updating the overall best outcome if it outperforms the previous result.
Step 4.If the termination requirement is met, the best outcome is returned; otherwise, modify the pheromone levels for each route, return to step 2, and continue the iterative process.

4.1 Description of Proposed ACO-SVM Algorithm

In this article, proposed a hybrid ACO algorithm which gave the process of selecting monkeypox virus DNA codons is metaphorically framed as an ant foraging for food (as shown in Figure 1). The ant moves from the nest to the food, passing through each gene in a candidate gene subset. At each DNA bases, the ant faces two pathways: pathway 1 signifies selecting the next codon, and pathway 0 denotes filtering the next codon. When the ant reaches the food, selected DNA bases (codon) form the monkeypox virus DNA subset, while others are filtered. For instance, a binary set S = {1, 0, 0, 1, 0, 1} indicates the 1st, 4th, and 6th codons are selected for the usage [17]. Evaluate the selected feature subset using a fitness function; a higher fitness value indicates a more desirable feature subset. Ants collaborate through the intensity of pheromones in pathways during the search for the optimal feature subset. In this hybrid ACO algorithm, multiple ants simultaneously explore pathways, selecting them based on the quantity of pheromones. The experimental steps of codon selection hybrid algorithm ACO-SVM is illustrated in Figure 2. Figure 2 visually represents the ACO and SVM-based Monkeypox Virus DNA codon selection algorithm, providing a clear explanation of the experimental procedures. The application of ACO algorithms to fine-tune SVM with RBF parameters, involves a systematic process in the experiments as the kernel. In each phase, a smart synthesis of ideas is applied to thoroughly explore and optimize the search space, directing a dependent heuristic through a sequence of creative and iterative operations.

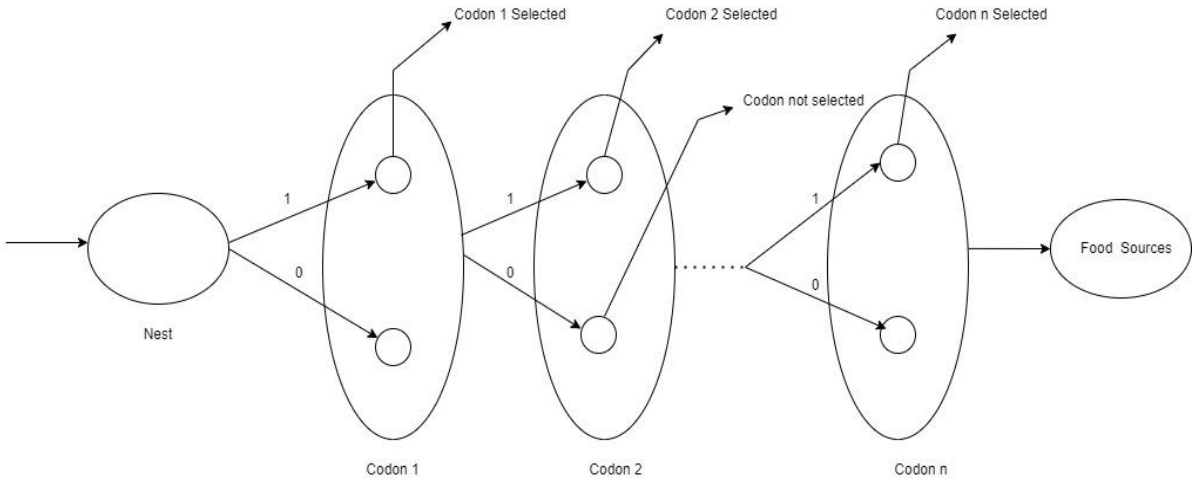

Fig. 1 The feature selection procedure in the hybrid ACO-SVM algorithm: '1' denoting the selection of the corresponding codon and '0' representing the non-selection of the codon.

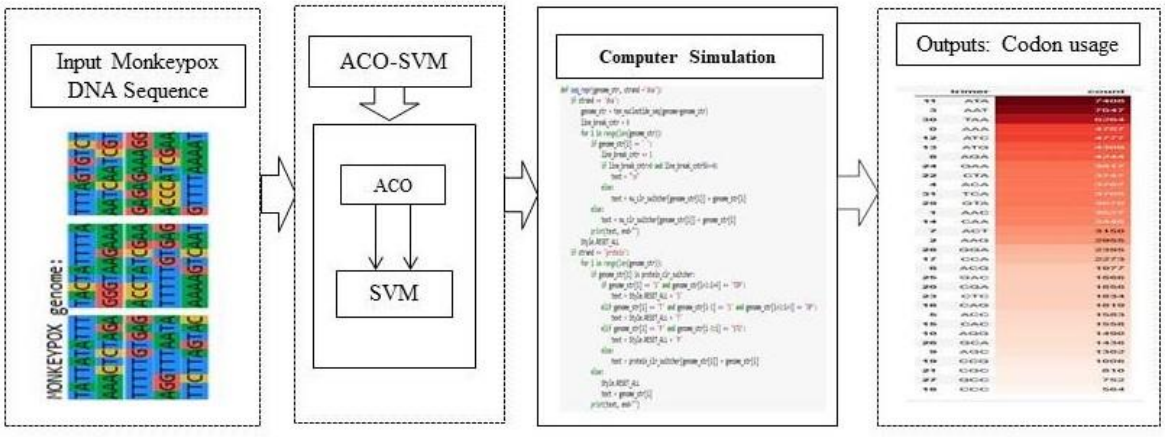

Fig.2 The Experimental steps of codon selection algorithm based on hybrid ACO-SVM.

Learning models are employed to construct knowledge, facilitating the efficient attainment of near-optimal solutions. The hybrid ACO approach iteratively generates SVM parameter values and integrates them into the SVM for pattern classification. The algorithm concludes its operation when the classification accuracy meets the user-specified threshold or the maximum iteration limit is reached. If these conditions are not met, the hybrid ACO algorithm persists in the search for optimal SVM parameter values to further enhance performance. The proposed ACO-SVM algorithm for codon selection in the context of analyzing the critical DNA associated with the monkeypox virus, specifically leveraging data from NCBI accession MT903343.1 datasets. In this method, real ants' efficient path finding, particularly their ability to navigate obstacles from a food source to the nest, serves as inspiration. Ants communicate through a chemical trail known as a pheromone, encompassing crucial traits of ant colonies such as distributed computing, constructive greedy heuristic and positive feedback [17, 24]. ACO was initially applied to address the traveling salesman problem (TSP), with the goal of finding the shortest closed tour that visits each town exactly once. Subsequently, the concept of the ant system was adapted to handle the codon selection problem. Here, each DNA base is represents to a city or node in the TSP. The ant colony generates a tour, and the nodes on this tour represent the selected DNA bases for codon selection. Considering a set of n DNA bases, the intensity of the pheromone trail between DNA bases pairs (i, j) at time t is denoted as τ_{ij}. In each iteration, an ant at time t selects the next DNA base, determining its position at time t + 1. To choose a specified number of DNA bases (d < n) from the original DNA set, the algorithm progresses

through cycles. In each cycle, consisting of d iterations, every ant completes a tour. Following each cycle, the trail intensity is updated based on a prescribed equation (3).

$$\tau_{ij}(t+1) = \rho\tau_{ij}(t) + \sum_{k=1}^{m} \Delta\tau_{ij}^{k}(t) \quad (3)$$

The coefficient ρ, where (1-ρ) signifies the trail evaporation between time t and t + 1, plays a crucial role in the algorithm. The term $\Delta\tau_{ij}^{k}$, representing the pheromone deposited by the k[th] ant on the edge (i, j) during the interval from time t to t + 1, is defined as follows:

$$\Delta\tau_{ij}^{k}(t) = \begin{cases} \frac{Q.L_k}{n_k} & if\ ant\ k\ select\ edge\ (i.j) \\ 0 & otherwise \end{cases} \quad (4)$$

Here Q remains a constant and k[th] ant's tour length represents by L_k, the tour length is calculated by Euclidean distance between two cities (nodes). Instead, the article proposes a new definition for the tour length of the k[th] ant. Then, σ_i denotes the standard deviation of the i[th] (i = 1,…, n), DNA bases across monkeypox virus DNA sequence. In this article, the tour length of ant (k[th]) is determined by:

$$L_k = \sum_{i=1}^{d} \sigma_i \quad (5)$$

To prevent the trail from accumulating indefinitely [17], it is necessary to assign a value to the coefficient ρ that is less than 1. Let $L^k(t)$ represent the set of DNA bases chosen by the k[th] ant at the time t, and let L(t) denote the set of DNA bases chosen by the m ants at the same time t.

$$L(t) = \bigcup_{k=1}^{m} L^k(t) \quad (6)$$

In experiments, specify the initial trail intensity at time 0 as:

$\tau_{ij}(0) = (C + \frac{\sigma_i + \sigma_j}{2})$, Here C is a constant (7)

The transition probability from DNA base i to j for the kth ant can be expressed as:

$$P_{ij}^{k} = \begin{cases} \frac{[\tau_{ij}(t)]^{\propto}[\eta_{ij}]^{\beta}}{\sum_{k=1}^{d}\{[\tau_{ij}(t)]^{\propto}[\eta_{ij}]^{\beta}\}} \\ 0,\ otherwise \end{cases} \quad (8)$$

Here, the term η_{ij}, is defined as $\eta_{ij} = \sigma_i + \sigma_j$, where k is a set comprising DNA bases not chosen by the k[th] ant. Additionally, ∝ and β are parameters regulating the pheromone trail in the context of the transition probability (Equation 8) from DNA bases i to j.

5. Experimental Results and Discussion

The monkeypox virus, a member of the Orthopoxvirus family, contains various genes that contribute to its ability to infect host cells. In this paper, the proposed method ACO-SVM experiments the NCBI datasets [27] and find the DNA based codon sequence of the monkeypox virus gene encoding the major envelope protein, let's look at a short segment: ATG TAC GGA CTA TGG AAA AGC CGC TAC GTT GCA TGT TGA, depicts in Table 1 and Figure 3 depicts the line graph represents the timer, codon count, and normalized frequency. The study conducted codon selection experiments using a hybrid ACO algorithm toolbox in Python. The ACO algorithm parameters included m = 50, α = 1, Q = 500, β = 1, ρ = 0.5, and d = selected DNA bases of monkeypox virus as NCBI monkeypox virus datasets MT903343.1. The research presented an ACO-based codon selection method to determine the number of each codon in the NCBI monkeypox virus datasets. To expedite codon selection and reduce computational burden, the ACO-SVM approach was employed to select informative DNA bases of monkeypox virus from the MT903343.1, NCBI datasets. Figure 4 depicts the experimental results from propose hybrid ACO-SVM algorithm for DNA of Monkeypox virus: (a) the query sequences of monkeypox virus DNA for evaluating the codon selection process, (b) Graphical representation of codon weight chart, (c) codon usage tables for monkeypox virus DNA sequence, (d) Code for query and

optimized sequence alignment for codon selection. Subsequently, the hybrid ACO-SVM algorithm was utilized to identify optimal monkeypox virus codons within these selected DNA bases.

Table 1 Detailed description of codon extracted by ACO-SVM

Sl. No.	Timer	Codon	Count	Norm-frequency
1	11	ATA	7408	0.050341252063812555
2	3	AAT	7047	0.037090786720089176
3	30	TAA	6264	0.031075496103732215
4	0	AAA	4787	0.07410795974382434
5	12	ATC	4777	0.03898371034062109
6	13	ATG	4309	0.016647211618344534
7	8	AGA	4244	0.02079061109884216
8	24	GAA	3817	0.03312616335930845
9	22	CTA	3747	0.04463093247520796
10	4	ACA	3707	0.013692147521847494
11	31	TCA	3705	0.01566920108106971
12	29	GTA	3678	0.07790432322722445
13	1	AAC	3527	0.05023608964044968
14	14	CAA	3448	0.045314488227066706
15	7	ACT	3150	0.03626000357552239
16	2	AAG	2955	0.016384305559937324
17	28	GGA	2395	0.019129044809708597
18	17	CCA	2273	0.02390341883038353
19	6	ACG	1977	0.005931160677666656
20	25	GAC	1866	0.010579339790306128
21	20	CGA	1856	0.019518145776151265
22	23	CTC	1834	0.008518156292393601
23	16	CAG	1819	0.039404360034072626
24	5	ACC	1583	0.01928678844475292
25	15	CAC	1558	0.04014049699761281
26	10	AGG	1490	0.01962330819951415
27	26	GCA	1436	0.015101323994910138
28	9	AGC	1302	0.007908214236888875
30	19	CCG	1006	0.025186400395410712
31	21	CGC	810	0.03867873931286873
32	27	GCC	752	0.06587374199451052
33	18	CCC	564	0.038962677855948515

The ACO-SVM algorithm successfully identified 33 codons for MT903343.1, and testing on the NCBI datasets revealed the number of codons within each codon group for example: "ATG" remains the start codon, "TAC" still encodes Tyrosine, "GGG" now encodes Glycine, which is a more frequently used codon in the host, "CTA" remains as Leucine, "TGG" still encodes Tryptophan, "AAA" remains as Lysine, "AGC" is changed to "CGC" to encode Arginine, a more common codon in the host, "CGT" now encodes Arginine, aligning with host preferences, "TAC" remains Tyrosine, "GTT" is used instead of "GTC" to encode Valine, optimizing for host cell translation, "GCA" remains as Alanine, "TGT" encodes Cysteine, "TGA" remains the stop codon. This is a simplified example, but it illustrates how codon optimization might be approached to enhance the expression and translation efficiency of a critical gene in the monkeypox virus, ultimately contributing to the development of a more effective vaccine. By aligning the viral DNA codon usage with the codon preferences of the host primate, the efficiency of protein expression can be enhanced. This increased efficiency can lead to higher levels of the major surface protein being produced in the host cells, ultimately contributing to a more effective immune response when the host encounters the virus.

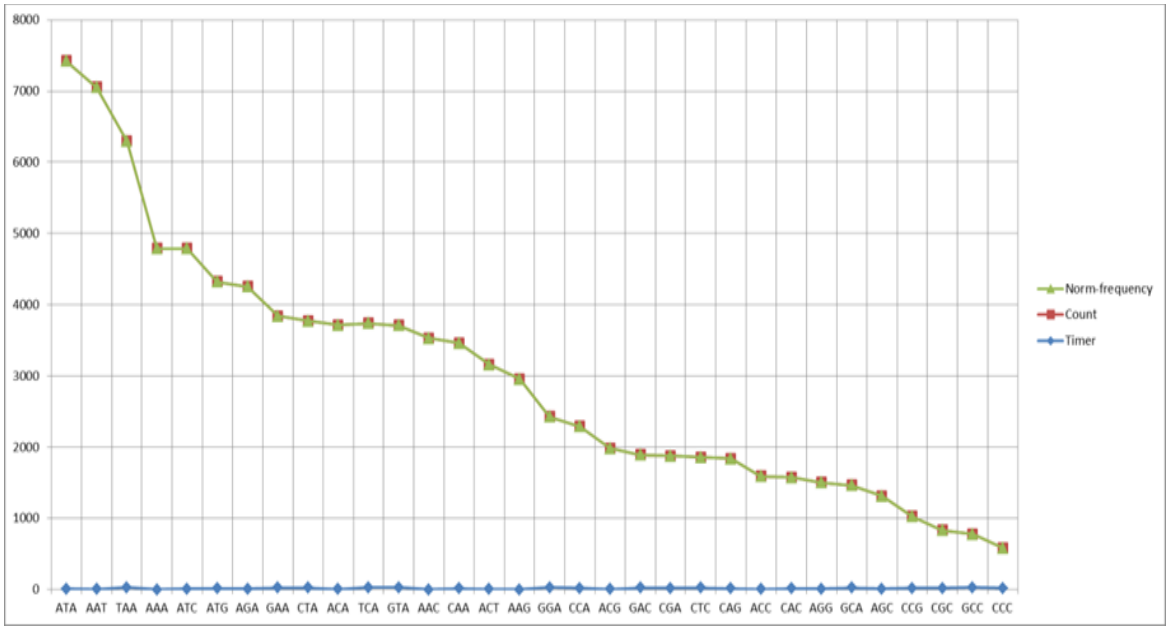

Fig.3 The line graph represents the timer, codon count, and normalized frequency.

5.1 Discussion

In this article proposes a hybrid ACO-SVM algorithm to optimize codon selection in the monkeypox virus gene, aligning it with host preferences for enhanced protein expression. This approach, demonstrated on NCBI datasets, showcases the potential for improving vaccine development by increasing the efficiency of viral gene translation in host cells. The development of a more effective vaccine against the monkeypox virus involves several key steps, and optimizing the virus's genetic code (codon optimization) is one strategy to enhance the vaccine's efficacy [25, 26]. Here's how the process contributes to the development of a more effective monkeypox vaccine:

Identification of Target Genes: The hybrid ACO-SVM methods identify key genes within the monkeypox virus genome that are crucial for its ability to infect host cells. This often includes genes encoding surface proteins, enzymes essential for replication, or proteins involved in viral entry.

Codon Optimization for Host Cells: The DNA sequences of these target genes are analyzed, and the codon usage is optimized to align with the preferences of the host cells, which are typically primate cells in the case of monkeypox. This involves selecting codons that are more frequently used in the host's genome, improving the efficiency of gene expression.

Increased Protein Expression: Codon optimization leads to enhanced expression of viral proteins within host cells. This increased expression is crucial for the vaccine's efficacy, as higher levels of the targeted antigens stimulate a stronger and more specific immune response.

Improved Antigen Presentation: The optimized viral proteins are processed and presented more efficiently on the surface of host cells. This improves the interaction with immune cells, such as T cells and B cells, leading to a more robust adaptive immune response.

Stimulation of Immune Response: The enhanced expression and presentation of viral antigens stimulate both cellular and humoral immune responses. This includes the activation of cytotoxic T cells, which can eliminate infected cells, and the production of antibodies that can neutralize the virus.

Memory Immune Response: The optimized vaccine aims to establish a long-lasting immune memory. This memory ensures that the immune system "remembers" the viral antigens, providing rapid and effective protection upon exposure to the actual monkeypox virus in the future.

Potential Cross-Species Adaptability: Codon optimization may also consider variations in codon usage between different monkeypox strains and various primate species. This adaptability can be crucial for developing a vaccine that is effective against diverse monkeypox strains and is potentially applicable to different primate hosts.

Finally, by optimizing the viral genes for efficient expression and immune recognition, the vaccine can induce a potent and targeted immune response, ultimately contributing to the development of a more effective monkeypox vaccine. It's worth noting that codon optimization is just one aspect of vaccine development, and a comprehensive approach involves multiple strategies to ensure safety, efficacy, and broad applicability.

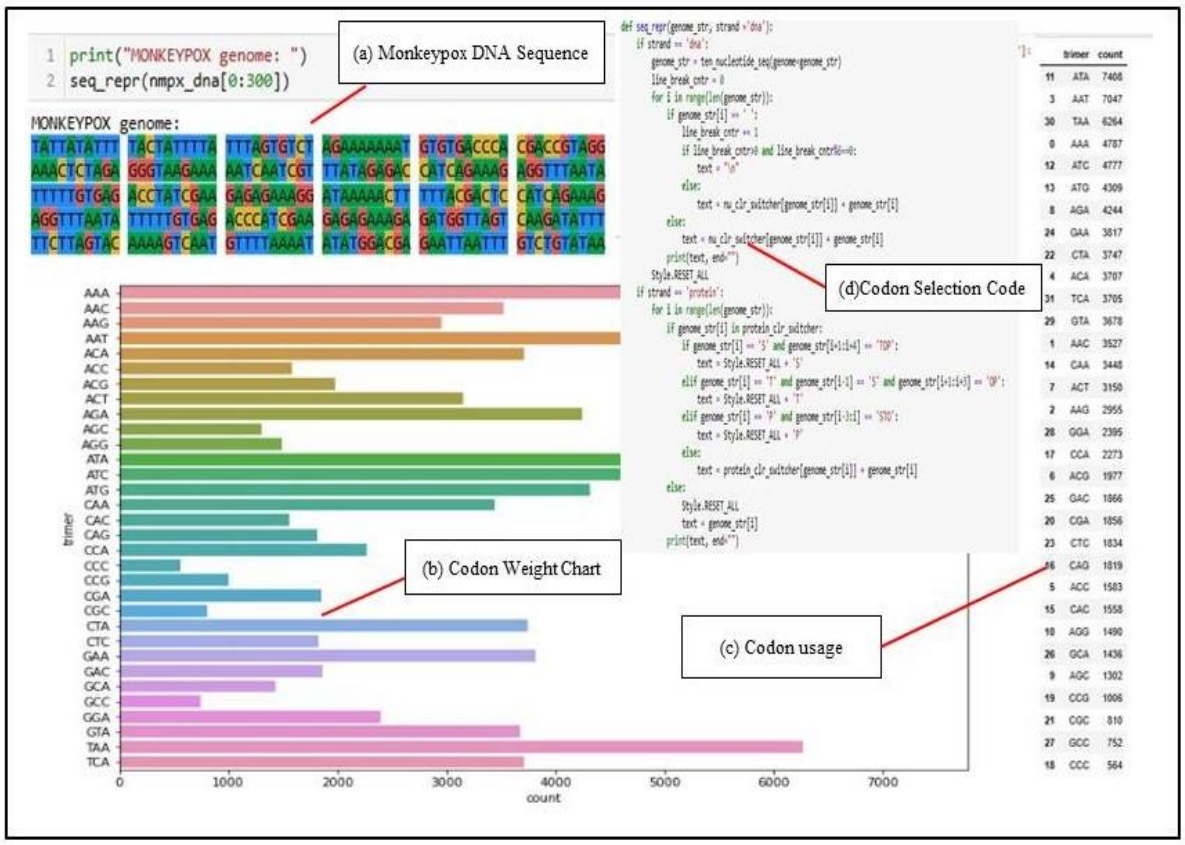

Fig. 4 Experimental results from hybrid ACO-SVM algorithm for DNA of Monkeypox virus: (a) the query sequences of monkeypox virus DNA for evaluating the codon selection process, (b) Graphical representation of codon weight chart, (c) codon usage tables for monkeypox virus DNA sequence, (d) Code for query and optimized sequence alignment for codon selection.

6. Conclusion

In conclusion, the hybrid ACO-SVM approach presented in this study demonstrates its effectiveness in optimizing codon selection for the monkeypox virus, particularly in the context of vaccine development. By aligning the viral genetic code with host cell preferences, the proposed algorithm enhances the efficiency of gene expression, leading to increased production of viral antigens. This optimization contributes to a more potent and targeted immune response, laying the groundwork for the development of a more effective monkeypox vaccine. In future, further exploration and refinement of the ACO-SVM algorithm can be pursued to address specific challenges and nuances in codon optimization. Additionally, extending the applicability of this approach to different strains of monkeypox and diverse primate hosts would enhance the vaccine's versatility. Integration of experimental validation and in vivo studies can provide concrete evidence of the algorithm's impact on vaccine efficacy. Moreover, considering the dynamic nature of viral genomes, continuous monitoring and adaptation of codon optimization strategies based on emerging data will be crucial. Collaboration with virologists, immunologists, and bioinformaticians can foster a multidisciplinary approach for comprehensive vaccine development. In essence, the proposed ACO-SVM algorithm serves as a stepping stone, and future endeavors can build upon its success to create even more robust and adaptable tools for optimizing gene expression in the pursuit of effective vaccines against emerging viral threats like monkeypox.

CRediT Authorship Contribution Statement: Dr.Akshaya Kumar Mandal: (https://orcid.org/0000-0003-0955-4096) the general idea of literature survey, Data collection, Resources, Implementation, Writing – review & editing – original draft. **Mr.Nurulla Mansur Barbhuiya**: The general idea of literature survey, Data collection, review & editing – original draft.**Prof.Pankaj Kumar Deva Sarma**: (https://orcid.org/0000-0002-9748-3787) the general idea of literature survey, Review, and editing.

Conflict of Interest: The writers confirm that they have no financial interests or interpersonal conflicts that may have seemed to influence the research presented in this article. This article complies with ethical standards since none of the writers conducted any experiments using human subjects.

References

[1] Kugelman JR, Johnston SC, Mulembakani PM, Kisalu N, Lee MS, Koroleva G, Rimoin AW. Genomic variability of monkeypox virus among humans, Democratic Republic of the Congo. Emerg Infect Dis. 2014;20(2):232. https://doi:10.3201/eid2002.130118 .

[2] Mandal AK, Sarma PK, Dehuri S. Machine learning approaches and particle swarm optimization-based clustering for the human monkeypox viruses: a study. In: Innovations in Intelligent Computing and Communication: First International Conference, ICIICC 2022, Bhubaneswar, Odisha, India, December 16-17, 2022, Proceedings. Cham: Springer International Publishing; 2023. p. 313–332. https://doi:10.1007/978-3-031-28435-9_27 .

[3] Dougherty ER. Small sample issues for microarray-based classification. Comp Funct Genomics. 2001;2:28-34. https://doi:10.1002/cfg.98 .

[4] Xing EP, Jordan MI, Karp RM, et al. Feature selection for high-dimensional genomic microarray data. In: Proceedings of the 18th International Conference on Machine Learning. San Francisco, USA: Morgan Kaufmann Publishers Inc.; 2001. p. 601–608.

[5] Roth FP. Bringing out the best features of expression data. Genome Res. 2001;11(11):1801-1802. https://doi:10.1101/gr.240701 .

[6] Guyon I, Weston J, Barnhill S, Vapnik VN. Gene selection for cancer classification using support vector machines. Mach Learn. 2002;46:389–422. https://doi:10.1023/A:1012487302797 .

[7] Nouri-Moghaddam B, Ghazanfari M, Fathian M. A novel bio-inspired hybrid multi-filter wrapper gene selection method with ensemble classifier for microarray data. Neural Comput Appl. 2023;35(16):11531–61. https://doi:10.1007/s00521-022-07763-0.

[8] Carreras J, Kikuti YY, Miyaoka M, Hiraiwa S, Tomita S, Ikoma H, Nakamura N. A single gene expression set derived from artificial intelligence predicted the prognosis of several lymphoma subtypes; and high immunohistochemical expression of TNFAIP8 associated with poor prognosis in diffuse large B-cell lymphoma. AI. 2020;1(3):342–360. https://doi:10.3390/ai1030023 .

[9] Dashtban M, Balafar M, Suravajhala P. Gene selection for tumor classification using a novel bio-inspired multi-objective approach. Genomics. 2018;110(1):10–17. https://doi:10.1016/j.ygeno.2017.08.008 .

[10] Dorigo M, Maniezzo V, Colorni A. The ant system: optimization by a colony of cooperating agents. IEEE Trans Syst Man Cybern B Cybern. 1996;26(1):29–42. https://doi:10.1109/3477.484436 .

[11] Chiang YM, Chiang HM, Lin SY. The application of ant colony optimization for gene selection in microarray-based cancer classification. In: 2008 International Conference on Machine Learning and Cybernetics. IEEE; 2008. p. 4001-4006. https://doi:10.1109/ICMLC.2008.4620985 .

[12] Alwan HB, Ku-Mahamud KR. Optimizing support vector machine parameters using continuous ant colony optimization. In: 2012 7th International Conference on Computing and Convergence Technology (ICCCT). IEEE; 2012. p. 164-169. https://doi:10.1109/ICCCT.2012.58 .

[13] Yu H, Gu G, Liu H, Shen J, Zhao J. A modified ant colony optimization algorithm for tumor marker gene selection. Genomics Proteomics Bioinformatics. 2009;7(4):200-208. https://doi:10.1016/S1672-0229(08)60045-8 .

[14] Prasad Y, Biswas KK, Jain CK. SVM classifier based feature selection using GA, ACO, and PSO for siRNA design. In: Advances in Swarm Intelligence: First International Conference, ICSI 2010, Beijing, China, June 12-15, 2010, Proceedings, Part II. Springer Berlin Heidelberg; 2010. p. 307-314. https://doi:10.1007/978-3-642-13495-1_38 .
[15] Puigbo P, Guzmán E, Romeu A, Garcia-Vallve S. OPTIMIZER: a web server for optimizing the codon usage of DNA sequences. Nucleic Acids Res. 2007;35(suppl_2). https://doi:10.1093/nar/gkm219 .
[16] Mandal AK, Sarma PKD. Usage of particle swarm optimization in digital images selection for monkeypox virus prediction and diagnosis. Malaysian J Comput Sci. 2024;37(2):124–138. https://doi.org/10.22452/mjcs.vol37no2.2.
[17] Mandal AK, Sarma PKD. Novel applications of ant colony optimization with the traveling salesman problem in DNA sequence optimization. In: 2022 IEEE 2nd International Symposium on Sustainable Energy, Signal Processing and Cyber Security (iSSSC). IEEE; 2022. p. 1-6. https://doi:10.1109/iSSSC55617.2022.9936749 .
[18] Mandal AK, Sarma PK, Dehuri S. A study of bio-inspired computing in bioinformatics: a state-of-the-art literature survey. Open Bioinform J. 2023;16(1). https://doi:10.2174/18750362-v16-e230517-2022-17 .
[19] Mandal AK, Dehuri S. A survey on ant colony optimization for solving some of the selected NP-hard problems. In: Biologically Inspired Techniques in Many-Criteria Decision Making: International Conference on Biologically Inspired Techniques in Many-Criteria Decision Making (BITMDM-2019). Springer International Publishing; 2020. p. 85–100. https://doi:10.1007/978-3-030-43519-1_6 .
[20] Vapnik VN. Statistical learning theory. John Wiley & Sons, Inc.; 1998.
[21] Ding X, Liu J, Yang F, Cao J. Random radial basis function kernel-based support vector machine. J Franklin Inst. 2021;358(18):10121-10140. https://doi:10.1016/j.jfranklin.2021.07.032 .
[22] Grzegorski SJ, Chiari EF, Robbins A, Kish PE, Kahana A. Natural variability of Kozak sequences correlates with function in a zebrafish model. PLoS One. 2014;9(9) https://doi:10.1371/journal.pone.0108475 .
[23] Zouridis H, Hatzimanikatis V. Effects of codon distributions and tRNA competition on protein translation. Biophys J. 2008;95(2):1018–1033. https://doi:10.1529/biophysj.107.125468 .
[24] Mandal AK, Sarma PKD, Dehuri S. Image-based skin disease detection and classification through bioinspired machine learning approaches. Int J Recent Innov Trends Comput Commun. 2023;12(1):85-94. https://doi:10.17762/ijritcc.v12i1.7914 .
[25] Bonidia RP, Machida JS, Negri TC, Alves WA, Kashiwabara AY, Domingues DS, Sanches DS. A novel decomposing model with evolutionary algorithms for feature selection in long non-coding RNAs. IEEE Access. 2020;8:181683-181697. https://doi:10.1109/ACCESS.2020.3028211 .
[26] Chung BKS, Lee DY. Computational codon optimization of synthetic gene for protein expression. BMC Syst Biol. 2012;6(1):1-14. https://doi:10.1186/1752-0509-6-134 .
[27] https://www.ncbi.nlm.nih.gov/nuccore/MT903343.1.