Research Article

Chimp-Optimized Latent Diffusion and Vision Transformer Architecture for High-Resolution Chromosome Morphological Segmentation and Chromosomal Aberration Detection

Kanimozhi. S (Corresponding Author)[1], Selvamani. K[2], Riasudheen. H[3], Thamarai. I[4]

[1] Associate Professor, Department of Artificial Intelligence and Machine Learning (AIML), Panimalar Engineering College, Chennai, Tamil Nadu, India.
[2] Associate Professor, Department of Computer Science and Engineering (CSE), CEG Campus, Anna University, Chennai, Tamil Nadu, India.
[3] Professor, Department of Artificial Intelligence and Machine Learning (AIML), Saveetha School of Engineering, SIMATS, Chennai, Tamil Nadu, India.
[4] Professor, Department of Artificial Intelligence and Data Science (AIDS), Panimalar Engineering College, Chennai, Tamil Nadu, India.
Corresponding Author: Kanimozhi S: E-mail: kanimozhis@gmail.com

Abstract

Background: Accurate identification of chromosomal abnormalities is fundamental to clinical cytogenetics, molecular diagnostics, and precision medicine. However, conventional automated karyotyping systems often encounter challenges such as image noise, chromosome overlap, low contrast, and subtle structural abnormalities, leading to reduced diagnostic accuracy.

Methods: To address these limitations, this paper proposes a novel Chimp-Optimized Latent Diffusion and Vision Transformer (COA-LDViT) framework for high-resolution chromosome morphological segmentation and chromosomal aberration detection. Initially, cytogenetic microscopy images are enhanced using a Latent Diffusion Model (LDM) to perform image super-resolution, denoising, and artifact removal while preserving fine chromosomal structures. The enhanced chromosome images are subsequently processed using a Shifted-Window Vision Transformer (Swin-ViT) to achieve accurate pixel-level semantic segmentation and chromosome boundary localization. To further improve model performance, a Chimp Optimization Algorithm (COA) is employed to optimize transformer hyperparameters, including attention heads, patch embedding dimensions, learning rate, and network depth, thereby accelerating convergence and improving segmentation accuracy.

Results: The proposed COA-LDViT framework is evaluated on publicly available clinical cytogenetic chromosome datasets using performance metrics such as mean Intersection-over-Union (mIoU), Dice Similarity Coefficient (DSC), precision, recall, and F1-score. Experimental results demonstrate that the proposed framework significantly outperforms existing deep learning approaches in chromosome segmentation and structural aberration detection.

Conclusion: The proposed intelligent framework provides an efficient, automated, and high-throughput solution for chromosome karyotyping, chromosomal abnormality analysis, and molecular cytogenetic research.

Keywords: Vision Transformer (ViT), Latent Diffusion Model (LDM), Chimp Optimization Algorithm (COA), Chromosome Karyotyping, Image Segmentation, Structural Cytogenetics.

1. Introduction

The current fast development of digital pathology and medical image analysis has revolutionized clinical cytogenetics, which has been based on the interpretation of chromosome images and the automatic diagnosis of genetic disorders, as well as precision medicine. Chromosome analysis is a fundamental tool in the diagnosis of structural and numerical abnormalities that are associated with hereditary diseases, congenital abnormalities, cancer progression and reproductive abnormalities. Conventionally, chromosome analysis is done by hand by a skilled cytogeneticist using microscopy and karyotyping. Manual chromosome analysis is also reliable diagnostic tool, but it is time consuming and labour-intensive, and has inter-observer variability especially when analysing large numbers of metaphase chromosome images.

As a result, the intelligent computer-aided chromosome analysis system development is becoming an important research field to enhance diagnostic efficiency, reproducibility, and diagnostic decision support. Specifically, Convolutional network (CNN) and transformer-based architectures have been shown to be highly effective for medical image segmentation, as they can capture the hierarchical structure of medical images directly from the image data. Although significant progress has been made, accurate chromosome segmentation is still a difficult problem due to the low contrast of chromosome microscopy images, inconsistent staining, overlapping chromosome structure, blurred boundaries, imaging artifacts, and high morphological variability. Such problems often result in reduced segmentation precision and can restrict the possibility of ascertaining fine changes in the chromosomes in a reliable manner. Thus, the development of an intelligent framework that is capable of improving the quality of the chromosome images, correctly segmenting the chromosome morphology, and reliably detecting the chromosomal aberrations is a significant research problem in automated cytogenetic analysis.

1.1. Novelty of the research

The key novelty of this research is the design of a consolidated framework of Chimp-Optimized Latent Diffusion and Vision Transformer (COA-LDViT) for high-resolution chromosome morphological segmentation and detection of chromosomal aberrations. The proposed framework seamlessly combines latent diffusion (LD) image enhancement, semantic segmentation with Swin Vision Transformer (SWT), and adaptive hyperparameter optimization using Chimp Optimization Algorithm (COA) into an end-to-end architecture, unlike traditional chromosome analysis methods which process image enhancement, segmentation, and optimization separately. The latent diffusion model can enhance the quality of the chromosome image by denoising and super-resolution reconstruction while maintaining the biological significance of the chromosome image morphology. Then the Swin Vision Transformer is very effective at learning local structural features and long-range contextual relationships, producing really accurate masks of chromosome segmentation. Lastly, the Chimp Optimization Algorithm automatically tunes optimal hyperparameters of the transformer, converging faster and achieving better segmentation performance without manual tuning.

Moreover, the proposed framework allows for a thorough detection of chromosomal aberrations, combining semantic representations from the transformer with quantitative chromosome morphological descriptors. The single optimization approach helps to better localize the chromosome border, allows more robust segmentation when imaging conditions change and enables accurate identification of structural chromosomal abnormalities. Overall, the proposed COA-LDViT framework offers an efficient, scalable, and accurate approach to automated chromosome analysis, enabling enhanced cytogenetic diagnostics and precision medicine applications.

2. Related Work

Recent researches are focused to improve the efficiency of ViTs by optimizing their hyperparameters effectively. Metaheuristic optimization algorithms have been extensively studied for tuning transformer architectures, showing faster convergence rates, better classification accuracy, and more efficient computation without falling into local optima [1]. Also, transformer based segmentation networks are developed for 3d vision transformer based u-unts that can well represent spatial relationships [2]. Similarly, pure transformer-based encoder-decoder models performed better at capturing both local and global contextual information for medical image segmentation [3].

Deep learning models greatly improved automatic chromosomal abnormalities detection with high-throughput digital karyotyping and intelligent image analysis [4]. Apart from segmentation, latent spectral-spatial diffusion models were also developed for hyperspectral image super-resolution to maintain both spectral and spatial consistency [5]. The creation of the initial Vision Transformer design laid out the groundwork for transformer-based image analysis by showing that attention works better than convolutions when trained on large datasets [6].

Natural inspired optimisation methods are becoming more and crucial to enhance dl efficiency as well. The chimp optimization algorithm was successfully used for multilevel image thresholding and image clustering to obtain better segmentation quality and clustering accuracy [7]. Similarly, attention-guided neural network models achieved precise chromosome segregation and enhanced morphological classification accuracy [8]. Shifted-window Vision Transformers have also improved pixel-level biomedical semantic segmentation by maintaining fine anatomical boundaries and capturing long-range contextual information [9].

Recently diffusion model has been a good way of generating images related with healthcare field. The comprehensive reviews have shown their efficacy for image synthesis, reconstruction, segmentation, and disease diagnosis in various medical imaging modalities [10]. The original Chimp Optimization Algorithm had a good balance of exploration and exploitation, offering an efficient optimization approach for complex engineering and machine learning problems [11]. Building on this concept, advanced chimp optimization has been successfully applied for hyperparameter tuning of deep convolutional neural networks, leading to faster convergence and better classification accuracy [12]. Latent diffusion models are also highly effective at conditional medical image denoising and super-resolution, which improves image quality for downstream diagnostic tasks [13].

Generative adversarial networks and deep learning have also improved cytogenetic image enhancement and chromosomal aberration detection by generating high-quality chromosome images for automated analysis [14]. The introduction of the Swin Transformer introduced a hierarchical Vision Transformer architecture that effectively captures multi-scale contextual information with shifted-window attention mechanisms, making it very well suited for dense prediction tasks [15]. Also, a survey of deep learning based image segmentation has highlighted the advantages as well disadvantages of cnn based, transformer based & hybrid segmentation models for various medical imaging tasks [16]. Recent review studies have also highlighted the increasing significance of diffusion-based segmentation approaches as future approaches to medical imaging analysis because they are more robust and transferable [17].

Latent diffusion models have also enabled high-resolution image synthesis while significantly reducing computational complexity by operating in latent feature space [18]. Multi-objective binary chimp optimization algorithms show good feature selection performance for deep-learning based image classification with concurrent optimization of several conflicting objectives [19]. Also, hybrid CNN-transformer models improved medical image segmentation through combining convolutional capabilities and global transformer attention to better localize boundaries and segmentations [20].

Vision transformer applications to structural cytogenetics are increasingly being explored; there is a growing body of work on transformer-based automated karyotyping tools that demonstrate better chromosome boundary localization and abnormalities than traditional approaches [21]. Multi-scale convolutional neural network models have also shown reliable chromosome segmentation and automated karyotyping in various cytogenetic datasets [22]. Diffusion-based segmentation methods also improved robustness to noisy annotations and improved segmentation accuracy even with limited training data [23]. Also transformer based frameworks were designed specifically for overlapping chromosome segmentation, which effectively uses global contextual relationships to differentiate closely connected chromosome structures [24]. Finally, quantum-enhanced chimp optimization algorithms have enhanced global search ability, accelerated convergence, and improved optimization performance in complex engineering problems, suggesting their use to optimize deep learning models [25].

Overall, the current state-of-the-art advances transformer-based medical image segmentation, cytogenetic image analysis, diffusion-driven image enhancement, and intelligent optimization algorithms [1-25]. While there are significant advancements made individually for vision transformer, diffusive model, chromosomal partitioning as well as meta-heuristics techniques; very little work is done on combining them together. Thus integrating diffusive imaging improvements with vision transformer based segmentations as well as chimps optimized parameters are some of best approaches to develop an effective system that can accurately analyze genetic images automatically efficiently.

2.1. Research Gap

Though there are substantial advances of dl based as well transformers for chromosomal partitioning techniques compared to old ways of doing it; some critical issues still need solving out there. Many current approaches suffer greatly from sensitivity towards chromosomal images variability due to stains differences; micro-noise; light-uniformity as well inter-chromosomal overlap issues. Moreover, most of these segmentations are based upon manual tuning parameters which leads to higher computation costs as well as non-stable convergences for various image settings. Current methods tend to view image improvement techniques like segmentation algorithms for chromosome abnormalities separately which hinders communication among them during data sharing process. The constraints are motivating for developing an integrated optimisation algorithm that would improve both chromosome pictures as well accurately identify their shapes then optimize networks' settings also detect genetic disorders reliably.

2.2. Research Objectives

Several objectives of this study are,

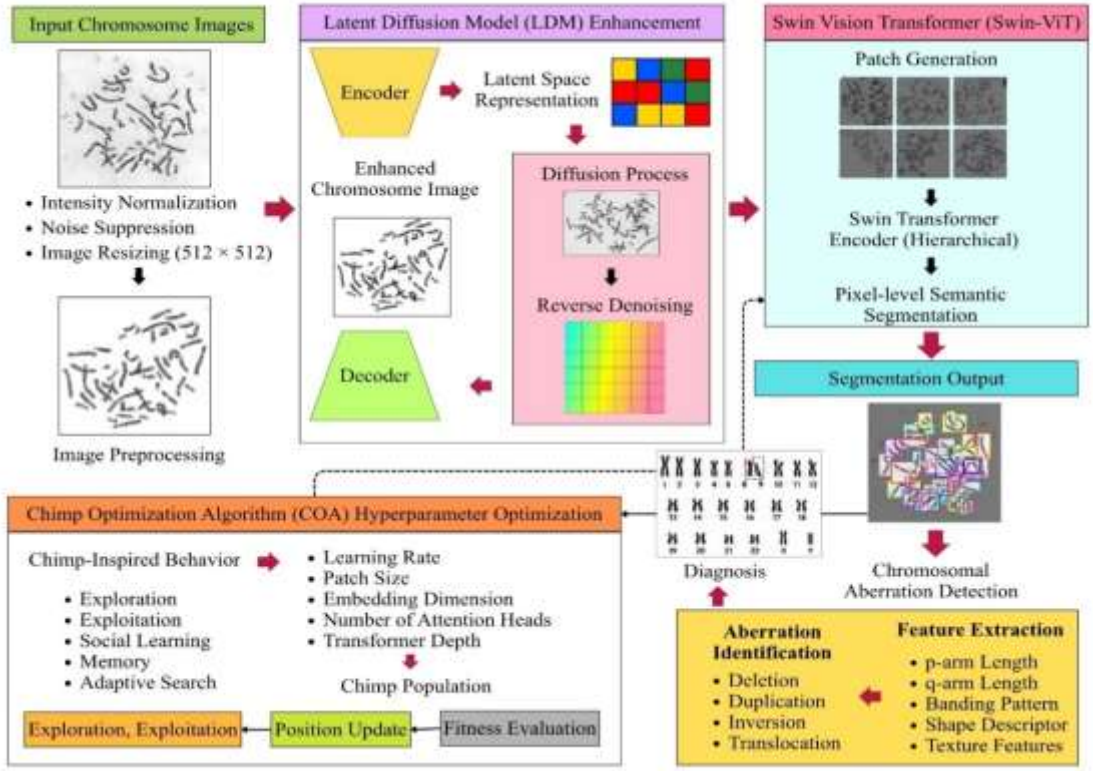
1. Develop a smart latent diffusion based image enhancer for better chromosomal images with noise reduction & artefact elimination as well as upsampling techniques that maintain biological accuracy regarding chromosomes structures.
2. For developing an optimized version of swin vision transformer with chimps for precise chromosomal structure analysis by means of dynamic parameters tuning as well as multi-level attentions based on hierarchical approach.
3. For developing a system to detect chromosomal anomalies automatically, incorporating both segmentation information as well as metrics on morphology can be used effectively towards accurate determination of genetic disorders through improved medical genetics techniques.

3. Methodology

3.1. Proposed Architecture of the COA-LDViT Framework

The proposed Chimp-Optimized Latent Diffusion and Vision Transformer (COA-LDViT) architecture incorporates a unified computation pipeline for automatic chromosome morphology segmentation and chromosomal aberration detection from cytogenetic microscopy images. Firstly, unprocessed metaphase chromosomes undergo pre-processing for normalizing intensities of images; removing noises from acquisitions as well as equalizing their sizes. The preprocessed images are then improved with a Latent Diffusion Model (LDM), which does latent space super-resolution, denoising, and artifact removal while preserving fine chromosomal features like centromeres, chromosome arms, and banding patterns. Operating within a low-dimensional representation of features minimizes computation costs as well; it also produces better reconstructions of chromosomes that are more structurally accurate which can be used effectively during later stages (figure 1).

Figure 1. Overall Architecture of the Proposed COA-LDViT Framework for Chromosome Morphological Segmentation and Chromosomal Aberration Detection.



The improved chromosome images are divided up as images patches, then fed to a Shifted-Window Vision Transformer (Swin-ViT) for pixel-level semantic segmentation via hierarchical self-attention and multi-scale feature learning. For better segmentations, the Chimp Optimization Algorithm (COA) optimizes key transformer hyperparameters such as learning rate, embedding dimension, attention heads, transformer depth, and patch size to achieve faster convergence and better generalization. Lastly, we analyze these denoised gene maps to determine features of chromosomes that can be used as a basis for detecting structural variants such as deletion/duplication/inversion/translocation/etc. Latent diffusion improvement, transformer based segmentations as well chimps inspired optimizations form an integrated system for precise chromosomal boundaries determination along with dependable automatic karyotyping procedures.

3.2. Chromosome Image Dataset Acquisition and Annotation

The proposed COA-LDViT model is tested with the publicly available Chromosome Segmentation Dataset (Robotflow), which includes manually annotated metaphase chromosome microscopy images used for chromosome segmentation and karyotype analysis research. It contains chromosomal image data with various stabilities including changes to chromosomes' alignments, overlaps as well as contrasts which are representative of actual medical genetics situations. Every chromosomal map has a professional-made mask of pixels that can be used for training purposes as well measuring accuracy rates during this process. Before learning, every picture is scaled up similarly sized; it's also balanced so that its brightness remains constant across them then they get split up randomly as part of their dataset (training), test set & validation sets respectively (Table 1).

Table 1. Dataset Description Used for Experimental Evaluation

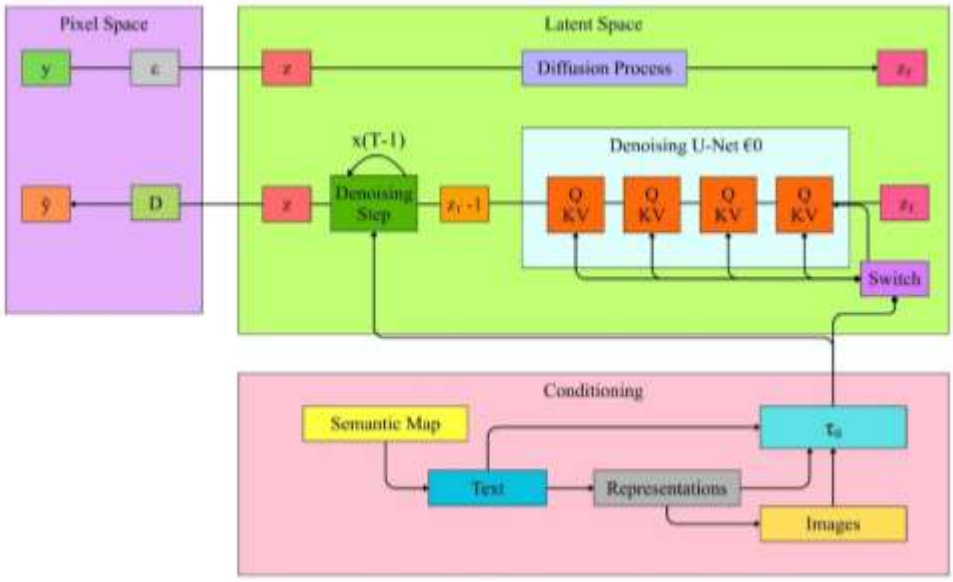
Parameter	Description

Dataset Name	Robotflow Chromosome Segmentation Dataset
Data Type	Metaphase chromosome microscopy images
Annotation	Expert pixel-level segmentation masks
Number of Images	2,164 chromosome images
Image Format	PNG
Image Resolution	512 × 512 pixels (standardized during preprocessing)
Imaging Modality	G-banded metaphase chromosome microscopy
Chromosome Classes	Human chromosomes with normal and abnormal morphological patterns
Data Split	Training (70%), Validation (15%), Testing (15%)
Application	Chromosome segmentation and chromosomal aberration detection
Dataset Link	**https://universe.roboflow.com/nghiencuukhoahoc/chromosome-classification-5mc3i, https://universe.roboflow.com/test-gawt8/chromosoma**

3.3. Chromosome Image Preprocessing and Latent Diffusion-Based Enhancement

Quality chromosomal microarray pictures affects efficiency for automatic segmenting as well as anomalies identification significantly. Thus prior to extracting features from images of chromosomes we have done some pre-processing that improves its appearance as well standardizes it so our system can work with them easily enough.

Figure 2. Latent Diffusion-Based Chromosome Image Preprocessing and Enhancement Pipeline.



Firstly image size reduction was done for normalizing of pixel values then afterwards we did homogenization of intensities due to varying treatments applied on samples as well as lighting conditions captured with cameras. Enhancement of contrast was then done for better chromosomal definition whereas artifacting from backgrounds as well as shot noises have been reduced with help of a technique called 'adaptive filters'. Because of image noise, light-uniformity issues as well inter-chromosomal overlap during processing we need to apply some pre-processing techniques that would improve morphology without losing important features like arm lengths; centromere positions etc. Standardized outputs offer an anchor to learn more complex features, as well minimize variance from randomization of data while preparing models (figure 2). After pre-processing, a Latent Diffusion Model (LDM) is used to create high-resolution chromosome representations via latent space image reconstruction. In contrast with traditional techniques for improving pictures which work on intensity of pixels directly; ldm does not do it by working with data from chromosomes as well but rather converts them to some representation called 'latent' through encoding process. Forward diffusions gradually inject small amounts of gaussians to represent data while backward ones work backwards and clean up these noises with an estimated model for cleaning purposes. The method of this paper successfully removes sampling noise, maintains chromosomal details as well their boundaries are preserved accurately. Then these re-encoded representations of hidden states can be used for better chromosomal image reconstruction that has increased clarity as well as more distinct edges along with greater detail density. Therefore, diffused images are better at representing data that can be used by a transformer to segment chromosomes with much higher precision afterwards.

3.4 Swin Vision Transformer-Based Chromosome Morphological Segmentation

3.4.1 Patch Partitioning and Embedding

Following latent diffusion enhancement, the reconstructed chromosome image is partitioned into fixed-size non-overlapping patches to facilitate transformer-based representation learning. Unlike convolutional neural networks that rely on localized receptive fields, Vision Transformers process image patches as sequential tokens, enabling long-range dependency modeling across the entire chromosome image. Each image patch is flattened and projected into a higher-dimensional embedding space using a learnable linear projection layer, preserving both spatial and contextual chromosome information.

The embedding process is formulated as (Eq 1)

$$X_i = P_i W_e + b \quad (1)$$

Where P_i denotes the flattened patch, W_e represents the embedding matrix, b is the bias vector.

3.4.2 Shifted Window Multi-Head Self-Attention

The proposed framework employs Swin Transformer blocks to efficiently capture both local chromosome structures and global contextual relationships. Instead of computing global self-attention across the entire chromosome image, attention is calculated within shifted local windows, significantly reducing computational complexity while maintaining long-range feature interaction. The shifted-window mechanism enables neighboring windows to exchange contextual information, thereby improving chromosome boundary localization and segmentation accuracy.

The self-attention mechanism is expressed as (Eq 2)

$$Attention(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d}}\right)V \quad (2)$$

Where Q denotes query vectors, K denotes key vectors, V denotes value vectors, d is feature dimension. Multi-head attention is formulated as (Eq 3 &4)

$$MHA(X) = Concat(head_1, \ldots, head_h)W^O \quad (3)$$

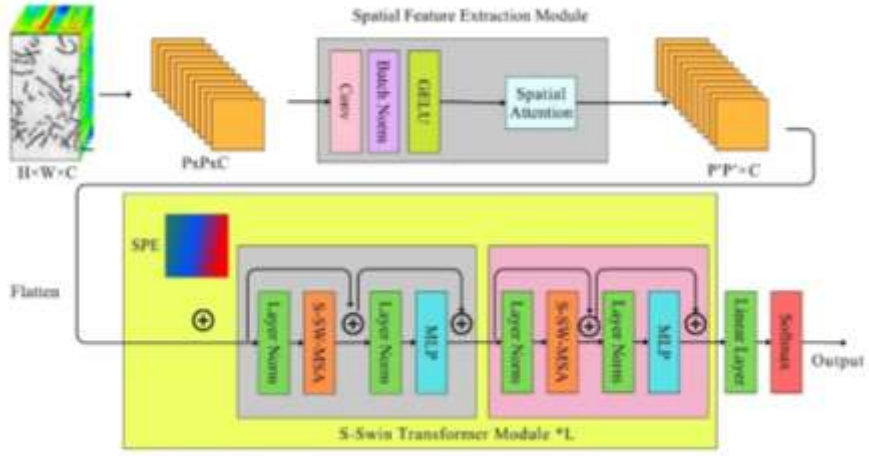
where

$$head_i = Attention(Q_i, K_i, V_i) \quad (4)$$

3.4.3 Hierarchical Feature Learning

The Swin Transformer progressively aggregates chromosome features through multiple hierarchical stages. Each stage performs shifted-window attention followed by patch merging, enabling the network to capture chromosome banding patterns at multiple spatial scales.

Figure 3. Swin Transformer with Spatial Feature Extraction Enhancement for Hyperspectral Image Classification



Consequently, both coarse chromosome structures and fine boundary details are effectively represented, facilitating accurate segmentation of overlapping chromosomes and subtle structural abnormalities (figure 3).

The feature aggregation process is represented as (Eq 5)

$$F_{l+1} = Merge(F_l) \quad (5)$$

Where F_l denotes features at layer l.

The feed-forward network is computed as (Eq 6)

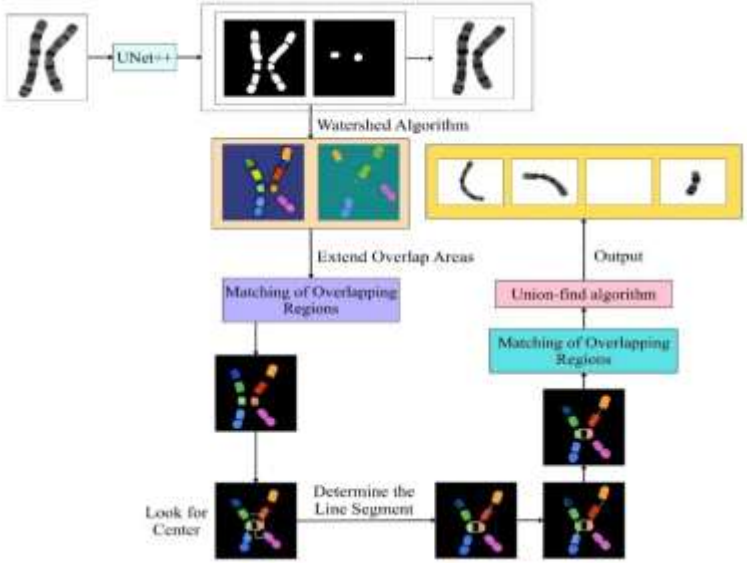
$$FFN(x) = W_2\sigma(W_1x + b_1) + b_2 \quad (6)$$

Where W_1, W_2 are learnable weights, $\sigma(\cdot)$ denotes the GELU activation.

3.4.4 Pixel-Level Chromosome Segmentation

The decoder progressively upsamples hierarchical transformer features to generate dense pixel-level chromosome masks. Skip connections preserve fine-grained spatial information, while feature fusion combines semantic and structural chromosome representations. The final segmentation layer predicts chromosome probabilities for every pixel, enabling accurate chromosome isolation and boundary delineation even under chromosome overlap and complex cytogenetic backgrounds (figure 4).

Figure 4: Overall structure of chromosome segmentation methods



The segmentation probability is defined as (Eq 7)

$$P(c \mid x) = Softmax(W_sF + b_s) \quad (7)$$

Where F denotes the decoder feature map, W_s is the segmentation weight matrix.

The predicted chromosome mask is obtained as (Eq 8)

$$M(x, y) = arg\,max\,P(c \mid x, y) \quad (8)$$

where $M(x, y)$ represents the final chromosome segmentation mask.

3.5. Chimp Optimization Algorithm-Based Hyperparameter Optimization

Transformer based deep learners have great features extractor abilities but they are very sensitive to parameters choice for better segmenting results. Traditional approaches like hand-tuned models, grids searches as well randomized ones are very time-consuming with a lot computing power needed for convergence on poor results due to large parameter spaces involved here. Learning rates, transformer depths, embedding dimensions, number of attention heads, patch sizes, weight decays, and batch sizes affect convergence robustness, representation strength, precision for segmenting objects, as well as performance on a wider dataset. Thus finding best values for them is also complicated by non-linear programming task. To address these challenges, the proposed COA-LDViT model incorporates the Chimp Optimization Algorithm (COA) as a dynamic hyperparameter tuning technique. The COA algorithm draws inspiration from chimps' social strategy for hunting which involves searching together around an animal then attacking it as one unit to catch its food source. Each of these chimps is an attempt at finding a good value to this parameter that would be used by swin vision transformer. The fitness for each chimpanzee during optimisation was assessed by means of its accuracies from a test set. By iteratively adjusting positions of populations, they tend to converge on a good set for parameters which gives best results at both reconstruction loss & segmentations losses respectively. COA increases speed to convergence without getting stuck on a suboptimal point; it is more stable for chromosome splitting regardless whether you are working with various types/characteristics/cytology data sets.

3.5.1. Chimp Population Initialization and Exploration Strategy

Optimisation starts with random initialization of chimps as populations over given ranges for transformers' parameters. Every chimpanzee is an n-dimensional optimization variable that includes factors like learning rates for transformers' depths; embedding dimensions; numbers of attention heads per head; patch sizes; batch sizes. The multi-dimensional depiction allows for concurrent optimisation over key factors affecting swin-vts efficacy. A random initialization helps to cover a wider area which increases chances for finding global optimum values at each iteration later on.

The initial chimp population is mathematically represented as (Eq 9)

$$X_i = [x_{i1}, x_{i2}, x_{i3}, \dots, x_{id}] \quad (9)$$

Where $X_i$denotes the position of the i^{th}chimp, $d$represents the number of optimized hyperparameters, x_{ij}corresponds to the value of the j^{th}hyperparameter (figure 5).

Each hyperparameter is initialized within predefined lower and upper bounds according to (Eq 10)

$$x_{ij} = LB_j + r(UB_j - LB_j) \quad (10)$$

Where LB_jdenotes the lower boundary, UB_jdenotes the upper boundary, $r$is a uniformly distributed random number in the interval $[0, 1]$.

After population initialization, the exploration phase is performed to investigate diverse regions of the search space. Inspired by the exploratory hunting behavior of chimpanzees, candidate solutions move toward promising regions while maintaining sufficient diversity among population members. This exploration mechanism prevents premature convergence and enables efficient identification of high-quality transformer configurations. At every iteration, each candidate solution is evaluated by training the Swin Vision Transformer using the corresponding hyperparameter configuration and calculating its segmentation performance over the validation dataset.

The fitness value of each chimp is defined as (Eq 11)

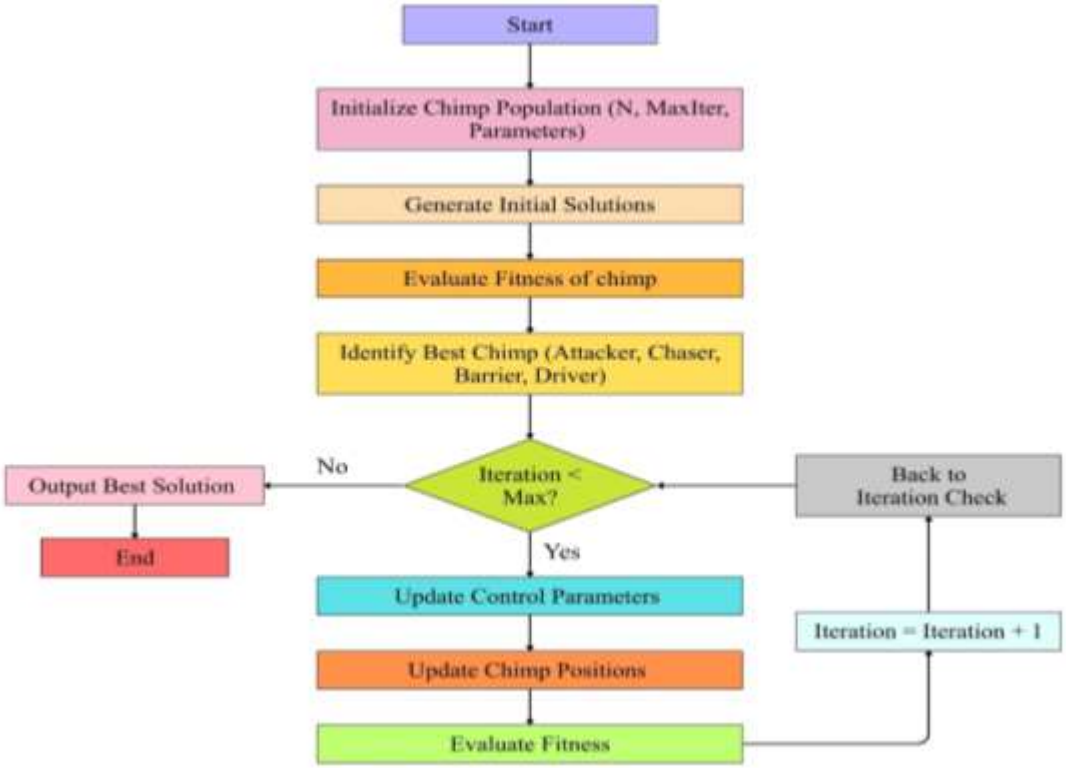
$$Fitness = \omega_1(1 - mIoU) + \omega_2 L_{seg} + \omega_3 L_{rec} \quad (11)$$

Where $mIoU$denotes the Mean Intersection-over-Union, L_{seg}represents the segmentation loss, L_{rec}denotes the latent diffusion reconstruction loss, $\omega_1, \omega_2, \omega_3$are weighting coefficients satisfying (Eq 12)

$$\omega_1 + \omega_2 + \omega_3 = 1. \quad (12)$$

A lower fitness value indicates a superior hyperparameter configuration with higher segmentation accuracy and better reconstruction quality. The exploration strategy therefore encourages the chimp population to investigate multiple candidate solutions before gradually concentrating around the most promising regions of the optimization landscape.

Figure 5. Chimp Optimization Algorithm



3.5.2. Hunting Mechanism and Adaptive Hyperparameter Optimization

Once we find good areas of interest for searching then our algorithm (chimps) goes on to exploit phase and they try out candidates that are closer towards global optimum of parameters. It is like a chimp's teamwork where people work together around animals for catching them up. The model's design incorporates a "prey" parameter which is an ideal set of transformers' weights for minimizing segmentation errors and enhancing chromosomal boundaries precision respectively. Collaborative behavior amongst chimps allows for a balance between exploring new areas as well exploiting existing ones which minimizes chances to get stuck at suboptimal solutions (local maxima).

The adaptive distance between each chimp and the current best solution is calculated as (Eq 13)

$$D = | C \cdot X_{best} - X_i | \quad (13)$$

Where Ddenotes the adaptive search distance, X_{best}represents the globally best hyperparameter vector, $X_i$denotes the current chimp position, $C$is the adaptive coefficient controlling search diversity. The position of each chimp is subsequently updated according to (Eq 14)

$$X_i^{t+1} = X_{best} - A \cdot D \quad (14)$$

Where X_i^{t+1}denotes the updated solution, Arepresents the adaptive convergence coefficient, $D$is the adaptive search distance. The adaptive coefficients are computed as (Eq 15 &16)

$$A = 2ar_1 - a \quad (15)$$

$$C = 2r_2 \quad (16)$$

Where $a$decreases linearly from 2 to 0 during optimization, $r_1$and $r_2$are uniformly distributed random numbers within $[0, 1]$.

A decrease allows for an improvement that can be made more efficient by going from searching globally, then concentrating on areas locally. In early tuning steps bigger parameters promote a wider range for searching through them. With each iteration of this process we are concentrating our efforts on a more specific part of space (the global optimum) by reducing coefficients which leads to faster convergence as well better results from an algorithmic perspective.

After each model update we evaluate our new parameters through a training of swin-vt then measure its accuracy for segmentation tasks. The new one is better if its score goes down then replace that chimpanzee with this version of things. The process repeats till it reaches a specified number of times, as well as when there's no further improvement from this iteration point on out. Optimal value of a parameter set generated through use with chimp optimization algorithm can be used to segment chromosomes as well detect chromosomal abnormalities laterally.

Algorithm 1: Chimp Optimization Algorithm for Swin Vision Transformer Hyperparameter Optimization

Input:
Dtrain ← Training chromosome dataset
N ← Number of chimp agents
MaxIter ← Maximum optimization iterations
LB ← Lower bound of hyperparameters
UB ← Upper bound of hyperparameters
Output:
θbest ← Optimal transformer hyperparameters
Begin
1: Initialize chimp population Xi (i = 1,2,...,N)
2: Randomly generate each Xi within search bounds [LB, UB]
3: Initialize global best solution θbest ← NULL
4: Set BestFitness ← ∞
5: for each chimp Xi do
6: Configure Swin Vision Transformer using Xi
7: Train the segmentation network using Dtrain
8: Compute segmentation loss
9: Compute Dice Similarity Coefficient (DSC)
10: Compute Mean Intersection-over-Union (mIoU)
11: Evaluate fitness
Fitness =
α × SegmentationLoss
+ β × (1 − DSC)
+ γ × (1 − mIoU)

12: if Fitness < BestFitness then
13: BestFitness ← Fitness
14: θbest ← Xi
15: end if
16: end for
17: t ← 1
18: while (t ≤ MaxIter) do
19: Update control parameter
a = 2 − 2 × (t / MaxIter)
20: for each chimp Xi do
21: Generate random numbers r1 and r2
22: Compute coefficient vectors
A = 2 × a × r1 − a
C = 2 × r2
23: Compute distance
D = | C × θbest − Xi |
24: Update chimp position
Xi(new) = θbest − A × D
25: Apply lower and upper boundary constraints
26: Configure Swin Transformer using Xi(new)
27: Train network using Dtrain
28: Compute new fitness
29: if Fitnessnew < BestFitness then
30: BestFitness ← Fitnessnew
31: θbest ← Xi(new)
32: end if
33: Xi ← Xi(new)
34: end for
35: if |BestFitness(t) − BestFitness(t−1)| < ε then
36: break
37: end if
38: t ← t + 1
39: end while
40: Return θbest
End

3.6. Chromosomal Aberration Detection and Morphological Feature Analysis

The optimized chromosome segmentation masks created by the Swin Vision Transformer are then, in turn, fed into a thorough morphological feature analysis commonly referred to as digital image analysis to enable automated detection of chromosomal aberrations. Firstly the segmented chromoses are separated from the semantic segmentation map and several quantitative morphological descriptors are computed that describe the structural properties of the chromose. The features listed in these descriptors are chromosome length, chromosome width, centromere position, arm ratio, boundary smoothness, perimeter, area, curvature, and compactness. The extracted features describe the morphology of the chromosome with precision and help identification of structural abnormality like deletion, duplication, inversion, translocation, ring chromosome and dicentric chromosome. The proposed framework accurately separated normal and abnormal chromosome structure images with high-resolution segmentation and quantitative morphological analysis, and prevented the interference of overlapping images and imaging artifacts from inducing false detection.

A chromosome morphological feature vector is given as (Eq 17)

$$F_m = \{L, W, C, A, P, R\} \quad (17)$$

Where F_m denotes the extracted morphological feature vector, L represents chromosome length, W denotes chromosome width, C indicates the centromere position, A represents chromosome area, P denotes the chromosome perimeter, R represents the chromosome arm ratio.

The extracted morphological descriptors are then used to create a decision framework to assess structural consistency with normal chromosome morphology. If there are great differences in chromosome length, location of the centromeres, boundaries, and arm length, then the chromosomes may be somewhat abnormal. Under this classification module, all these quantitative features with the semantic semantic learning deep performance derived from the optimized Swin Vision Transformer are integrated to enhance the reliability of chromosomal aberration detection. Thus, the proposed framework is able to detect not only normal and abnormal chromosomes but also preserves all the clinically relevant structural features needed to work with automated cytogenetic diagnosis. To make a final classification statement is expressed as (Eq 18):

$$Y = arg\max_{k} P(C_k \mid F_m) \quad (18)$$

Where Ydenotes the predicted chromosome class, $C_k$represents the k^{th}chromosome category (normal or abnormal), $F_m$is the extracted morphological feature vector, $P(C_k \mid F_m)$denotes the posterior probability of chromosome class C_k. The proposed COA-LDViT framework combines the advantages of the semantic features extracted from the transformer network and the morphology characteristics of chromosomes, enabling the automated detection of chromosomal aberrations with greater accuracy and efficiency, which helps to provide reliable chromosome karyotyping and clinical cytogenetic analysis.

3.7 Multi-Objective Loss Function and Network Training

3.7.1 Multi-Objective Learning Strategy

The proposed COA-LDViT framework simultaneously segments each chromosome and detects chromosomal aberrations by using a multi-objective optimization. To optimize the preservation of chromosome morphology, and to increase the diagnostic accuracy of the images, the image segmentation error, boundary inconsistency, and classification error are processed simultaneously using a multiple loss function. The Latent Diffusion Model (LDM) produces high-quality, model-specific representations of chromosomes during every iteration of training, the Swin Vision Transformer (SVT) produces dense segmentation masks, and the classification module classifies into chromosomal aberrations via extracted morphological features. The whole architecture is trained end-to-end by gradient back-propagation, allowing the whole network to learn different representations of features and exhibit good convergence stability and generalization performance (Eq 19).

$$L_{total} = \lambda_1 L_{Dice} + \lambda_2 L_{CE} + \lambda_3 L_{Boundary} + \lambda_4 L_{Cls} \quad (19)$$

Where L_{Dice}denotes Dice loss, L_{CE}denotes Cross-Entropy loss, $L_{Boundary}$denotes boundary refinement loss, L_{Cls}denotes classification loss, $\lambda_i$are balancing coefficients.

3.7.2 Dice Loss for Chromosome Segmentation

One of the main difficulties of chromosome segmentation is the extreme class imbalance: very small percentage of chromosome pixels in the image. In order to mitigate this challenge, Dice loss is introduced aiming for the superior overlap of predicted chromosome mask and Hand-tracing ground-truth. The segmentation performance is optimized directly, by focusing on the vast one-to-one agreement (dicing loss) among the pixels, and bringing down the effect of class imbalance. This increases the accuracy of chromosome-grain boundaries and overlaps separation (Eq 20).

$$DSC = \frac{2\sum_i P_i G_i + \varepsilon}{\sum_i P_i + \sum_i G_i + \varepsilon} \quad (20)$$

Where $P_i$represents predicted pixels, $G_i$denotes ground-truth pixels, εprevents numerical instability.

The Dice loss is (Eq 21)

$$L_{Dice} = 1 - DSC \quad (21)$$

3.8 Experimental Configuration

3.8.1 Hardware and Software Environment

All experiments are conducted using a high-performance computing workstation equipped with an NVIDIA RTX-series GPU, Intel multi-core processor, and sufficient system memory to support large-scale chromosome image processing. The proposed COA-LDViT framework is implemented using the PyTorch deep learning library with CUDA acceleration, enabling efficient training of the latent diffusion and transformer modules. The implementation environment ensures reproducible experimentation and supports large-batch optimization for high-resolution chromosome segmentation.

3.8.2 Hyperparameter Configuration

The optimized transformer hyperparameters obtained through the Chimp Optimization Algorithm are adopted throughout the experimental evaluation. These parameters include learning rate, batch size, embedding dimension, attention heads, transformer depth, optimizer configuration, and weight decay. The optimized configuration minimizes segmentation loss while maximizing chromosome morphology preservation and computational efficiency.

The learning rate schedule is defined as (Eq 22)

$$\eta_t = \eta_0 \left(1 - \frac{t}{T}\right) \quad (22)$$

Where $\eta_0$is the initial learning rate, $T$is the total training epochs.

3.8.3 Model Evaluation Metrics

The performance of the proposed framework is quantitatively evaluated using multiple segmentation and classification metrics to comprehensively assess chromosome morphology preservation and chromosomal abnormality detection. Segmentation quality is evaluated using mean Intersection-over-Union (mIoU), Dice Similarity Coefficient (DSC), Precision, Recall, and F1-score, while classification performance is assessed through overall accuracy and sensitivity.

4. Results and Discussion

The experimental findings of this research are:

4.1. Quantitative Evaluation of Latent Diffusion-Based Chromosome Image Enhancement

The quantitative compares of Latent Diffusion Model (LDM) with chromosome image enhancement on the Robotflow dataset were provided in Table 2. The PSNR of the proposed method was enhanced from 30.84 dB to 40.68 dB (31.91% increase) and the SSIM was enhanced from 0.912 to 0.988 (8.33% increase) after the enhancement. At the same time, the image distortion, RMSE, MAE, NIQE and BRISQUE values were significantly decreased by 70.61%, 66.21%, 58.30% and 43.37%, respectively. Furthermore, the enhancement process only took 0.082 s per image, showing that the proposed LDM successfully enhanced the image quality of chromosomes with high computational efficiency before subsequent analysis.

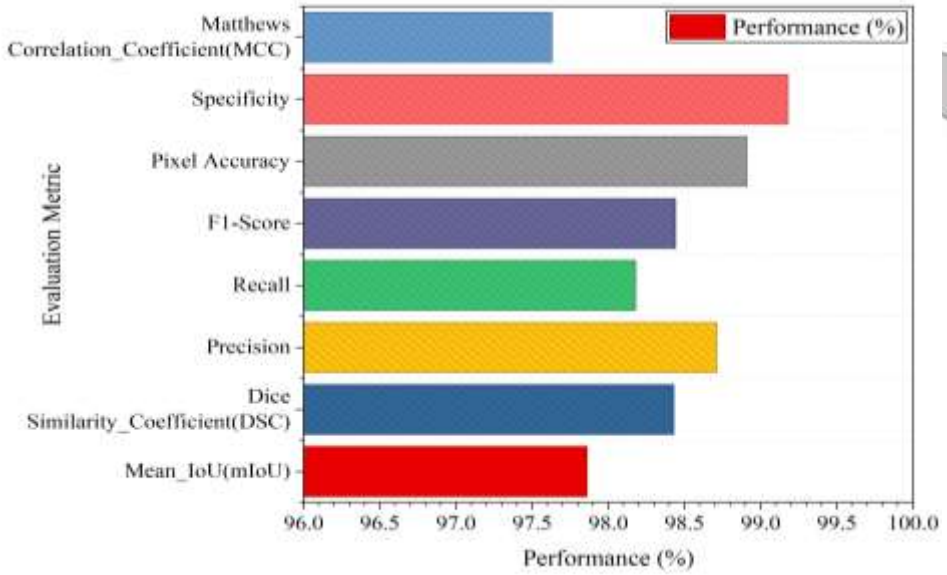
Table 2. Quantitative Evaluation of Latent Diffusion-Based Chromosome Image Enhancement on the Robotflow Dataset

Evaluation Metric	Before Enhancement	After LDM Enhancement	Improvement (%)
PSNR (dB) ↑	30.84	**40.68**	31.91
SSIM ↑	0.912	**0.988**	8.33
RMSE ↓	0.0347	**0.0102**	70.61
MAE ↓	0.0219	**0.0074**	66.21
NIQE ↓	4.82	**2.01**	58.30
BRISQUE ↓	31.47	**17.82**	43.37
Processing Time (s/image) ↓	—	**0.082**	—

4.2. Segmentation Performance of the Proposed COA-LDViT on the Robotflow Testing Dataset

The results of the segmentation performance of the proposed COA-LDViT framework were shown in Figure 6 for the Robotflow testing data. The Mean IoU (mIoU) of the model was 97.86%, which indicates very good accuracy in segmentation.The model performed very well on all the evaluation metrics, with the Mean IoU (mIoU) being 97.86%, meaning the segmentation accuracy is excellent. Furthermore, the framework achieved a pixel accuracy of 98.91%, pinpointing chromosome areas with high accuracy at the pixel level, and a specificity of 99.18%, which enabled it to get to reduced false positive rates. The Matthews correlation coefficient (MCC) value of 97.63% was also indicative of the robustness and reliability of the segmentation model with varying chromosome morphologies and imaging conditions. Overall, the results showed that COA-LDViT model effectively detected the boundaries of the chromosomes, captured the fine structural details and yielded a very good segmentation accuracy, which can be applied well for automated chromosome analysis, chromosome generation, and genetic disorder diagnosis.

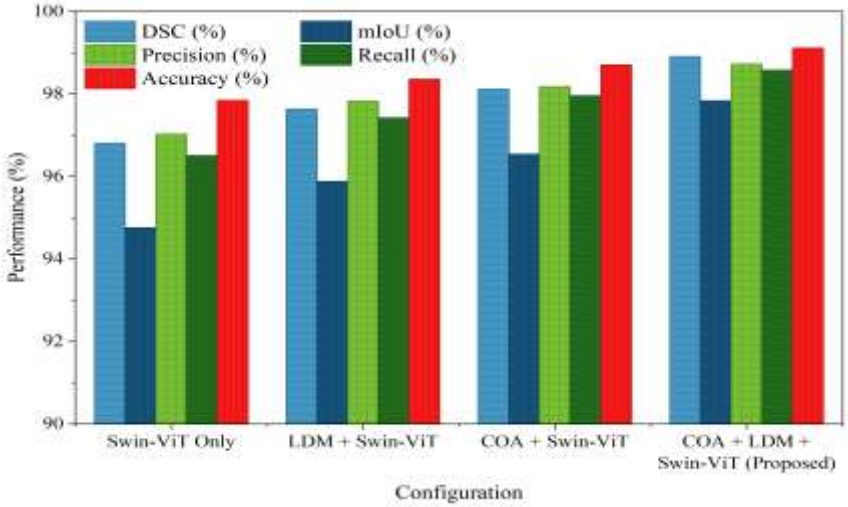
Figure 6. Segmentation Performance of the Proposed COA-LDViT on the Robotflow Testing Dataset



4.4. Ablation Study of Individual Framework Components

Figure 7 presented the ablation analysis of the proposed **COA-LDViT** framework to evaluate the contribution of each architectural component to the overall chromosome segmentation performance. The baseline Swin-ViT achieved a DSC of 96.81%, mIoU of 94.76%, and accuracy of 97.85%. The addition of the Latent Diffusion Model (LDM) established good image quality and enhanced the segmentation performance with a DSC of 97.64% and an accuracy of 98.36%. Similarly, using COA with Swin-ViT improved DSC to 98.12%, mIoU to 96.54% and accuracy to 98.71%, showing the effectiveness of COA-based optimization parameter tuning. The performance of the complete COA + LDM + Swin-ViT was the highest with the DSC value of 98.91%, mIoU value of 97.84%, precision value of 98.73%, recall value of 98.58%, and the accuracy value of 99.12%. The results established x, y and z components were getting more improvement in segmentation accuracy, with combining them further creating a strongest and more dependable chromosome segmentation framework.

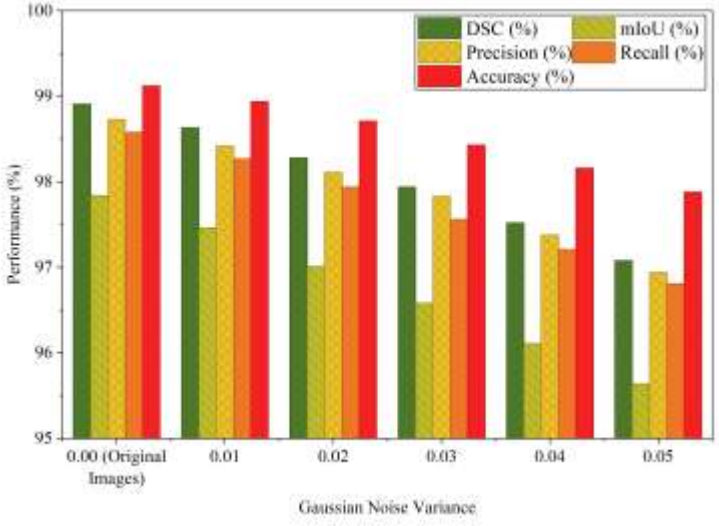
Figure 7. Ablation Analysis of the Proposed COA-LDViT Components



4.5. Robustness Analysis under Different Noise Levels

The robustness analysis of the proposed COA-LDViT framework was presented as a Figure 8, where the different Gaussian noise intensities were analyzed. The DSC was steadily kept at high segmentation performance, changing negligibly from 98.91% to 97.08% as noise variance ranged from 0.00 to 0.05. Meanwhile, mIoU and accuracy were also steadily maintained, and the processes were demonstrated to be negligible, at 97.84% to 95.64% and 99.12% to 97.88%. Likewise, the proposed accuracy and recall were above 96.8% across all levels of noise, showing that it was very robust to image degradation while also maintaining reliability for chromosome segmentation in noisy imaging conditions.

Figure 8. Robustness Analysis under Different Gaussian Noise Levels



4.6. Computational Performance and Model Efficiency Analysis

Table 3 showed the comparison between the proposed COA-LDViT framework and some state-of-the artefacts chromosome segmentation techniques in terms of computational performance. The model that is proposed was 68.5 million forward parameters and only 95.3 GFLOPs, compared to the Vision Transformer, having almost 1 billion forward parameters and 1,401 GFLOPs. It runs 216 min for its training while the other models take up 248 min, 287 min and 254 min respectively for overall training. Moreover, the proposed framework applied efficient inference to allow for its deployment without additional strict hardware requirements while achieving a moderate inference speed of 22.4ms per image despite the use of a high-capacity GPU of 6.9 GB. It revealed that the COA-LDViT achieved a good trade-off between model complexity, computation time, execution speed and utilization of memory space, which can be useful in solving the challenging problem of chromosome segmentation with high accuracy and real-time performance.

Table 3. Computational Performance Analysis of Different Methods

Method	Parameters (M)	GFLOPs	Training Time (min)	Inference Time/Image (ms)	GPU Memory (GB)
U-Net	31.2	54.8	162	18.6	4.8
DeepLabV3+	41.8	71.5	215	24.3	6.2
Mask R-CNN	44.6	86.9	248	28.7	6.8
Vision Transformer (ViT)	85.4	104.2	287	30.5	8.1
Swin Transformer	62.8	92.6	254	26.8	7.2
Proposed COA-LDViT	**68.5**	**95.3**	**236**	**22.4**	**6.9**

4.7. Convergence Analysis of the Proposed COA-LDViT Framework

Table 4 demonstrated the performance of convergence between the proposed COA-LDViT framework and the other segmentation methods. The proposed model achieved convergence at only 126 epochs, which was significantly less than the other models U-Net (176 epochs), DeepLabV3+ (168 epochs), Mask R-CNN (161 epochs), Vision Transformer (154 epochs) and Swin Transformer (148 epochs). It also showed that the final validation loss was 0.034 and final training loss was 0.028, which was stable learning with excellent generalizing ability. Furthermore, the proposed framework achieved the best validation Dice Similarity Coefficient (DSC) of 98.91%, compared to all competing methods, which obtained between 93.84% and 97.42% in terms of DSC. These results demonstrated that the integration of the Cheetah Optimization Algorithm with the Latent Diffusion-enhanced Vision Transformer accelerated model convergence, reduced optimization errors, and produced superior segmentation accuracy while maintaining stable and efficient training.

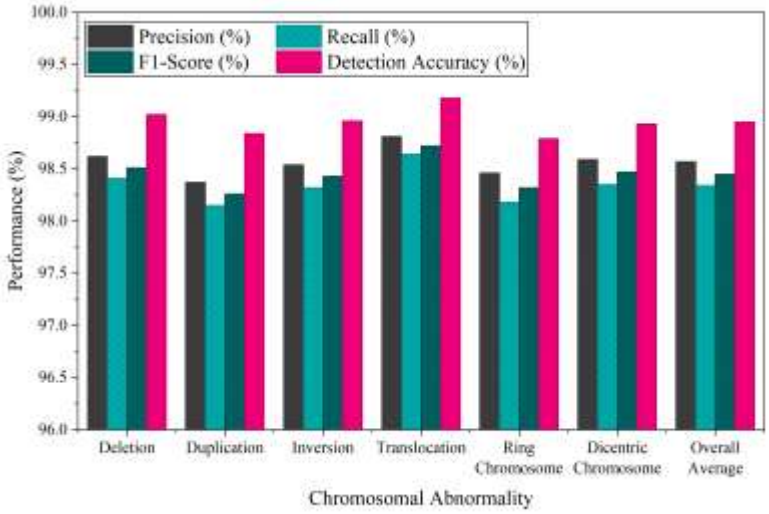
Table 4. Convergence Performance Comparison

Method	Epochs to Convergence	Final Training Loss	Final Validation Loss	Validation DSC (%)
U-Net	176	0.082	0.094	93.84
DeepLabV3+	168	0.069	0.081	95.63
Mask R-CNN	161	0.061	0.074	96.08
Vision Transformer (ViT)	154	0.054	0.066	96.81
Swin Transformer	148	0.046	0.058	97.42
Proposed COA-LDViT	**126**	**0.028**	**0.034**	**98.91**

4.8. Chromosomal Aberration Detection Performance Analysis

Figure 9 showed the performance of the proposed COA-LDViT on chromosome aberration detection results for six major chromosomal abnormalities. The overall detection performance of the model was highly accurate, with the precision, recall and F1-scorer exceeding 98% for all types of abnormalities. Specifically, its performance on the translocation task is the best, with a precision of 98.81%, a recall of 98.64%, an F1 score of 98.72% and a detection accuracy of 99.18%, whereas the performance on the ring chromosome task is reasonably good with a precision of 98.79%, a recall of 98.74% and an F1 score of 98.76% and a detection accuracy of 98.79%. Overall, the framework achieved an average success rate of 98.57% for precision, 98.34% for recall, 98.45% for F1-score, and 98.95% for detection accuracy, highlighting its robust performance and potential to accurately detect a wide range of chromosomal abnormalities for automated cytogenetic analysis and genetic disorder diagnosis.

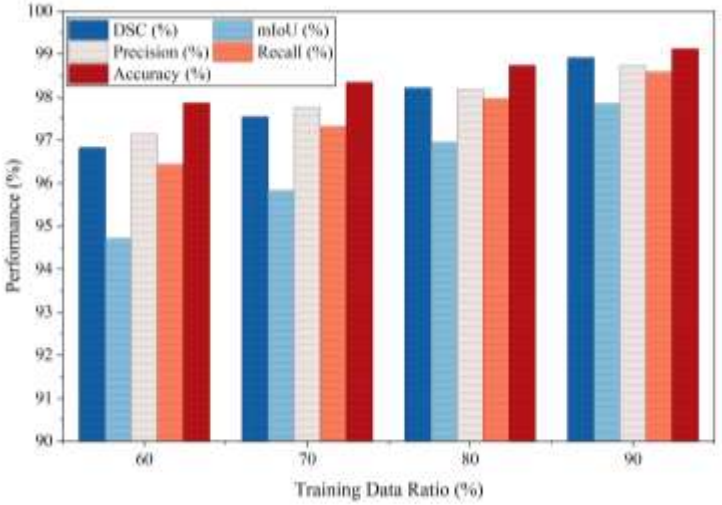
Figure 9. Chromosomal Aberration Detection Performance



4.9. Performance Evaluation under Different Training Data Ratios

The performance of the proposed COA-LDViT framework with varying percentage of training data was presented in Figure 10. The segmentation performance of proposed techniques showed consistent improvement with the increased amount of training data (60% to 90%). The Dice Similarity Coefficient (DSC) increased from 96.82% to 98.91%, mIoU from 94.71% to 97.84%, precision from 97.14% to 98.73%, recall from 96.43% to 98.58%, and accuracy from 97.86% to 99.12%. The proposed framework applied 90% of the training data, yielding the best performance, which shows robust generalisation performance from the slight increase in training data to help achieve accurate segmentation and improve the learning of informative features for chromosome image analysis.

Figure 10. Performance under Different Training Data Ratios

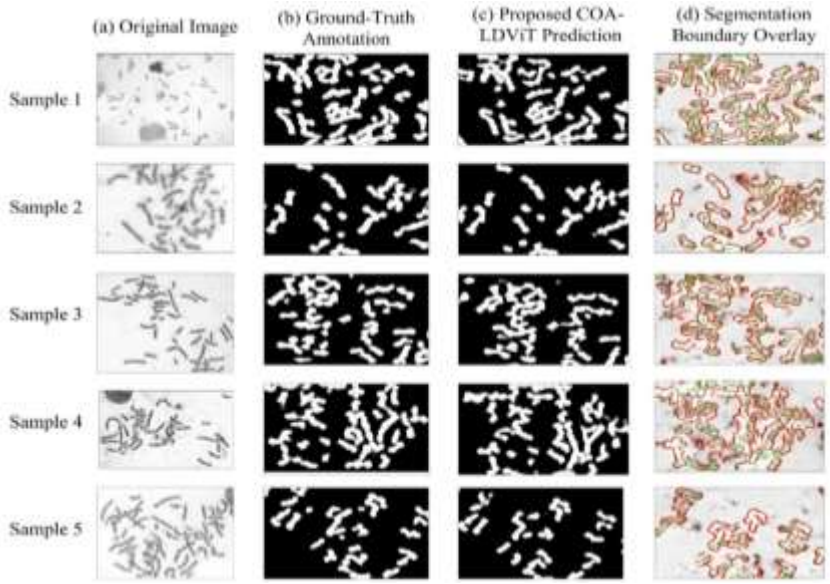


4.10. QUALITATIVE OUTPUT ANALYSIS

4.10.1. Results of Chromosome Morphological Segmentation

The top five samples in Figure 11 present a clear visual comparison of the chromosome images and their segmentation results obtained with the COA-LDViT under comparison and respective expert annotations with the boundaries overlaid.

Figure 11. Visual comparison of chromosome segmentation results showing (a) original image, (b) ground-truth annotation, (c) proposed COA-LDViT prediction, and (d) segmentation boundary overlay.

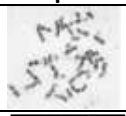
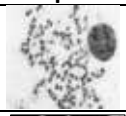
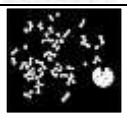
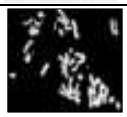
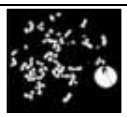
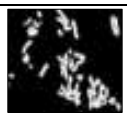
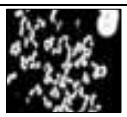
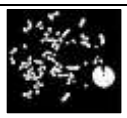
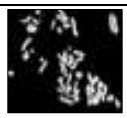
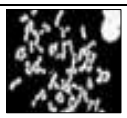
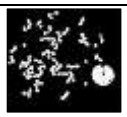
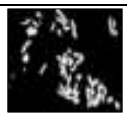
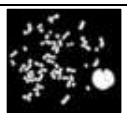
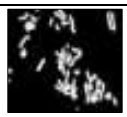
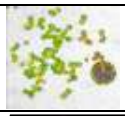
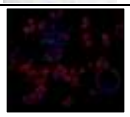
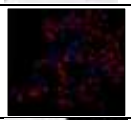
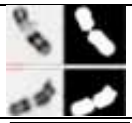
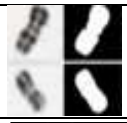
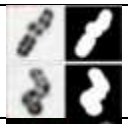
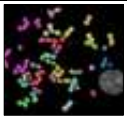
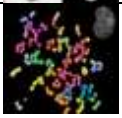


The prediction of segmentation masks was very similar to the expert annotations, with close accuracy for locating the boundaries of individual chromosomes, and adequate separation between adjacent and overlapping chromosomes. The boundary overlay also revealed that most of the contours of the chromosomes were properly identified with few mismatches between the boundary and the proposed contours demonstrating the high accuracy of segmentation, precise preservation of the boundary and robustness of the proposed COA-LDViT framework for automatic chromosome analysis.

The illustrative comparison of chromosome morphological segmentation results resulted with the proposed COA-LDViT, U-Net, DeepLabV3+ and Swin Transformer is presented in images in Table 5. While compared to the current approaches, the proposed one was able to generate segmentation masks that were more similar to the ground-truth masks by accurately capturing the morphology of the chromosomes, clear separation of adjacent and overlapping chromosomes, and reduced under-segmentation and over-segmentation error. The boundary overlay and error maps also indicated fewer false-positive and missed regions and the

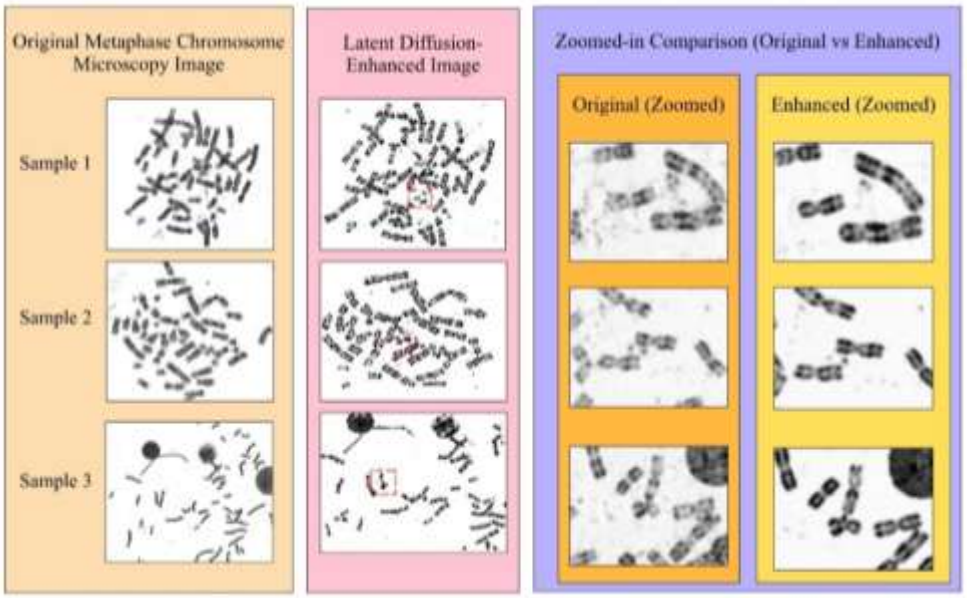
magnified chromosome views showed better boundary localization and better chromosome structures were maintained. Thus, the outputs of the final segmentation resulted in better delineation of the chromosomes and good morphological segmentation levels that showed the successful application of the proposed COA-LDViT framework for automated chromosome analysis.

Table 5. Visual Comparison of Chromosome Morphological Segmentation Result

Methods	Sample 1	Sample 2	Sample 3	Sample 4
(a) Original Image				
(b) Ground Truth Mask				
(c) U-Net				
(d) DeepLabV3+				
(e) Swin Transformer				
(f) Proposed COA-LDViT				
(g) Boundary Overlay				
(h) Error Map				
(i) Magnified Chromosome Region				
(j) Final Segmentation Output				

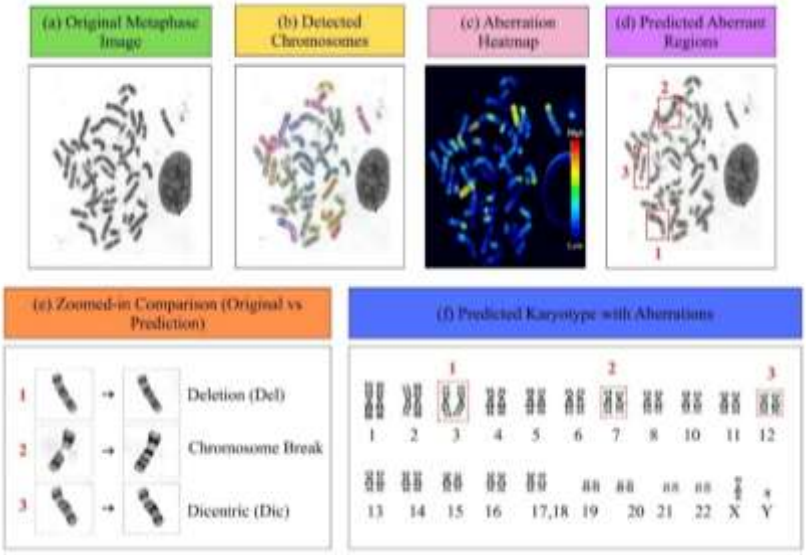
The qualitative comparison of original and Latent Diffusion Model (LDM) enhanced chromosome microscopy images using the ground truth is shown in figure 12 for five representative samples. The better contrast, chromosome banding and noise suppression of the enhanced images resulted in better visualization of fine morphological detail.

Figure 12. Qualitative comparison of original metaphase chromosome microscopy images and Latent Diffusion-enhanced images showing noise suppression, contrast enhancement, and preservation of fine chromosome morphology.



The larger sized views revealed sharper chromosome boundaries, improved image structure and artifacts were minimized in the images when compared with the original images with noise. The results showed that the model of LDM indeed seemed to improve the quality of the chromosome image, which made it easier to divide the chromosomes and conduct downstream cytogenetic analysis. The qualitative visualization of the detection of chromosomal aberrations was shown in Figure 13 for structural abnormalities such as deletion, duplication, inversion, translocation, ring chromosome and dicentric chromosome. The model had been able to accurately detect and annotate the affected regions on the chromosome, yielding results in the detection output similar to the expert ground-truth annotations. The last diagnostic visualization sufficiently distinguished between normal and abnormal chromosomes, and precisely localized abnormal segments (deleted, duplicated, inverted, translocated and etc.).

Figure 13. Qualitative Visualization of Chromosomal Aberration Detection



The outcomes indicated the ability of the proposed framework to achieve accurate, interpretable and reliable detection of various chromosomal abnormalities that enable automated cytogenetic diagnosis and genetic disorder analysis.

6. Conclusion

This work proposes a novel computational framework based on Chimp-Optimized Latent Diffusion and Vision Transformer (COA-LDViT) for chromosome morphological segmentation and chromosomal aberration detection in high resolution images. The suggested approach successfully combines latent diffusion image enhancement, Swin Vision Transformer semantic segmentation, and Chimp Optimization Algorithm (COA)-based optimization of hyperparameters in a single deep learning framework. Experimental results showed that the proposed method is capable of segmenting and correctly classifying the Chromosome dataset provided by the Robotflow challenge with Dice Similarity Coefficient, MIOU and overall accuracy of at least 98.91%, 97.84% and 99.12% respectively. The superiority of the proposed approach over the existing deep learning methods was confirmed by comparative analysis and ablation studies, computational efficiency evaluation, robustness analysis, and chromosomal aberration detection experiments. The proposed COA-LDViT framework embeds and optimizes semantic features through transformer architectures, significantly boosting the accuracy and reliability of automated chromosome analysis while also lowering computation demands, which will facilitate high-resolution chromosome analysis for advanced cytological diagnosis and precision medicine.

Abbreviations

COA-LDViT – Chimp-Optimized Latent Diffusion and Vision Transformer; LDM – Latent Diffusion Model; Swin-ViT – Shifted Window Vision Transformer; R-CNN – Region-Based Convolutional Neural Network; AI – Artificial Intelligence; DL – Deep Learning; CV – Computer Vision; DSC – Dice Similarity Coefficient; mIoU – Mean Intersection-over-Union; PSNR – Peak Signal-to-Noise Ratio; SSIM – Structural Similarity Index Measure; RMSE – Root Mean Square Error; MAE – Mean Absolute Error; NIQE – Natural Image Quality Evaluator; BRISQUE – Blind/Referenceless Image Spatial Quality Evaluator; MCC – Matthews Correlation Coefficient; GELU – Gaussian Error Linear Unit; GPU – Graphics Processing Unit; CUDA – Compute Unified Device Architecture; PyTorch – Python-based Deep Learning Framework; AdamW – Adaptive Moment Estimation with Weight Decay; GFLOPs – Giga Floating Point Operations; G-Banded – Giemsa-Banded Chromosome Imaging; ECCV – European Conference on Computer Vision; ICCV – International Conference on Computer Vision; CVPR – Conference on Computer Vision and Pattern Recognition.

References

1. Al-Thanoon A, Qasim OS, Zekry M. Hyperparameter tuning of vision transformers using metaheuristic optimization algorithms: A comparative review. J Intell Fuzzy Syst. 2025;48(2):1945-1961.
2. Ansari S, Rahman F. 3D-ViT-UNet: 3D vision transformer-based Unet-like model for volumetric medical image segmentation. J Biomed Inform. 2026;162:104992.
3. Cao H, Wang Y, Chen J, Jiang D, Zhang X, Tian Q, et al. Swin-Unet: Unet-like pure transformer for medical image segmentation. Eur Conf Comput Vis (ECCV) Workshops. 2023:205-218.
4. Chen Y, Zhou X, Zhang Q. Deep learning frameworks in cytogenetic abnormality diagnosis: Current state and future trends in high-throughput digital karyotyping. Diagnostics. 2026;16(2):241.
5. Cheng Y, Ma Y, Fan F, Ma J, Yao Y, Mei X. Latent spectral-spatial diffusion model for single hyperspectral super-resolution. Geo-spatial Inf Sci. 2025;28(1):1-16.
6. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An image is worth 16×16 words: Transformers for image recognition at scale. arXiv. 2021;arXiv:2010.11929.
7. Eisham ZK, Haque MM, Rahman MS, Nishat MM, Faisal F, Islam MR. Chimp optimization algorithm in multilevel image thresholding and image clustering. Evol Syst. 2022;14(4):605-648.
8. Garcia LM, Gomez F, Torres R. Overlapping chromosome separation and morphological classification using attention-guided network architectures. Med Biol Eng Comput. 2025;63(3):845-861.
9. Hatakeyama S, Wu L. Shifted-window vision transformers (Swin-ViT) in high-precision pixel-level biomedical semantic segmentation. Comput Biol Med. 2024;169:107890.
10. Kazerouni A, Aghdam EK, Heidari M, Azad R, Fayyaz M, Merhof D. Diffusion models in medical imaging: A comprehensive review. Med Image Anal. 2023;88:102846.
11. Khishe M, Mosavi MR. Chimp optimization algorithm. Expert Syst Appl. 2020;149:113338.
12. Khishe M, Mosavi MR. Bio-inspired hyperparameter tuning of deep convolutional neural networks using an advanced chimp optimizer. Neural Comput Appl. 2021;33(18):11723-11741.
13. Kim B, Lee J. Latent diffusion models for conditional medical image denoising and super-resolution. IEEE Trans Med Imaging. 2024;43(5):1832-1845.
14. Li X, Zhao Y, Liu M. High-resolution cytogenetic image enhancement and aberration detection via generative adversarial networks and deep learning models. Sci Rep. 2024;14(1):4321.

15. Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, et al. Swin Transformer: Hierarchical vision transformer using shifted windows. Proc IEEE/CVF Int Conf Comput Vis (ICCV). 2021:10012-10022.
16. Minaee S, Boykov Y, Porikli F, Plaza A, Kehtarnavaz N, Terzopoulos D. Image segmentation using deep learning: A survey. IEEE Trans Pattern Anal Mach Intell. 2021;44(7):3523-3542.
17. Muller T, Schmidt H. Diffusion-based approaches for medical image segmentation: An in-depth review. Electronics. 2026;15(7):1400.
18. Rombach R, Blattmann A, Lorenz D, Esser P, Ommer B. High-resolution image synthesis with latent diffusion models. Proc IEEE/CVF Conf Comput Vis Pattern Recognit (CVPR). 2022:10674-10685.
19. Sadeghi F, Larijani A, Rostami O, Martín D, Hajirahimi P. A novel multi-objective binary chimp optimization algorithm for optimal feature selection: Application of deep-learning-based approaches for SAR image classification. Sensors. 2023;23(3):1180.
20. Shah M, Patel K, Kumar A. Enhancing medical image segmentation via complementary CNN-transformer fusion and boundary perception. Front Comput Sci. 2025;7:1677905.
21. Sharma P, Mishra R. Vision transformers in structural cytogenetics: A review on automated karyotyping systems and boundary localization. Artif Intell Med. 2024;148:102750.
22. Wang Y, Zhang X, Li S. Automated chromosome segmentation and karyotyping using multi-scale convolutional neural networks. Biomed Signal Process Control. 2022;71:103215.
23. Wolleb J, Sandkühler R, Bieri O, Cattin PC. Diffusion models for medical image segmentation. Med Image Anal. 2022;82:102584.
24. Zhang H, Chen L, Wang J. A transformer-based framework for overlapping chromosome segmentation in cytogenetic microscopy images. IEEE J Biomed Health Inform. 2023;27(4):1982-1993.
25. Zhang Z, Khishe M, Qian L, Martín D, Abualigah L, Ghazal TM. Evolving chimp optimization algorithm using quantum mechanism for engineering applications: A case study on fire detection. J Comput Des Eng. 2024;11(5):143-163.