Scientific Reports

https://doi.org/10.1038/s41598-026-37678-6

Article in Press

A method for structural variant detection using Hi-C contact matrix and neural networks

Received: 30 October 2025

Accepted: 23 January 2026

Published online: 05 February 2026

Cite this article as: Shen J., Wang H., Zhai H. *et al.* A method for structural variant detection using Hi-C contact matrix and neural networks. *Sci Rep* (2026). https://doi.org/10.1038/s41598-026-37678-6

Jiquan Shen, Haojie Wang, Haixia Zhai, Junfeng Wang & Junwei Luo

A method for Structural Variant Detection using Hi-C Contact Matrix and Neural Networks

Jiquan Shen[1, 2], Haojie Wang[1], Haixia Zhai[1], Junfeng Wang[1,*], Junwei Luo[1, *]

1, School of Software, Henan Polytechnic University, Jiaozuo, 454003, China.

2, College of Chemical and Environmental Engineering, Anyang Institute of Technology, Anyang 455000, China

*Correspondence to luojunwei@hpu.edu.cn.

Abstract

Background: Structural variations (SVs) play a key role in many human diseases and are major causative factors of malignant tumors. High-throughput chromatin conformation capture (Hi-C) technology captures spatial interactions between genomic fragments, thereby enhancing SV identification and localization and compensating for the limitations of sequencing-based approaches in detecting complex variants. However, existing methods based on Hi-C data still suffer from low accuracy, limited applicability, and difficulties in handling multiple types of SVs simultaneously.

Results: In this study, we propose VarHiCNet, a novel method for detecting structural variations from Hi-C data. Contact matrices are preprocessed and converted into image-like representations. These representations are then input into an improved RT-DETR network to identify candidate SV regions. Subsequently, a filtering and classification network is applied for precise breakpoint detection. Evaluated on six cancer cell lines, VarHiCNet demonstrates high accuracy and stability in SV identification, with overall performance surpassing that of existing methods. The source code is available at https://github.com/000425/VarHiCNet.

Conclusions: Experimental results indicate that VarHiCNet achieves superior performance in detecting structural variations compared to other methods, offering a robust and accurate tool for genomic studies.

Keywords: structural variation; Hi-C; target detection; deep learning; RT-DETR

Background

Structural variations (SVs) refer to changes in DNA sequence longer than 50 base pairs (bp) within the genome. SVs can be classified into deletions, duplications/tandem repeats, inversions, translocations, and more complex combinations[1]. With the continuous advancement of high-throughput sequencing technology, researchers have discovered that structural variations exhibit significant enrichment patterns in disease states. During the occurrence and development of tumors, SVs can induce carcinogenic effects through various molecular mechanisms, as follows: (I) Deletions can directly lead to heterozygous loss (LOH) or homozygous loss of tumor suppressor genes[2]; (II) Duplications can increase the copy number of proto-oncogenes (e.g., EGFR amplification) [3]; (III) Inversions can cause

abnormalities in the three-dimensional chromatin structure of gene regulatory regions [4]; (IV) Balanced translocations can generate oncogenic fusion genes (e.g., EML4-ALK) [5]. As such, accurately detecting structural variations in the genome, particularly in highly complex samples such as cancer cell lines, is of great significance for understanding disease mechanisms and advancing precision medicine.

Although whole genome sequencing (WGS) has become the primary method for detecting structural variations [6], WGS methods based on short reads still have limitations in many respects. On one hand, the mapping quality of short-read sequencing is suboptimal, particularly in highly repetitive regions, making it challenging to identify many breakpoints [7] accurately. On the other hand, in the detection of complex SVs (such as multi-breakpoint rearrangements) and copy number neutral events (such as balanced translocations and inversions), WGS often lacks sufficient resolution and contextual information to make comprehensive judgments[8]. While long-read technologies (such as PacBio and Nanopore) have partially addressed the mapping issues associated with short reads, their high costs and high error rates limit their application in large-scale sample studies[9]. Furthermore, existing WGS methods primarily focus on the precise localization of variant breakpoints and fail to fully leverage the three-dimensional structural features of the genome for global analysis. This detection method, based on local information, may overlook important structural variation features.

High-throughput Chromosome Conformation Capture (Hi-C) technology can capture three-dimensional spatial contact information of DNA sequence across the entire genome, forming a high-dimensional contact matrix that reflects spatial interactions between different genomic locations[10]. Hi-C provides a genomic structural perspective for structural variation detection by capturing physical interaction information of chromatin in three-dimensional space. Several attempts have been made to detect structural variations based on Hi-C data. Early methods, such as HiC_breakfinder [11], are based on statistical modeling of interaction frequencies, aiming to identify breakpoints from regions with abnormally high interaction frequencies. HiNT-TL[12] analyzes the distribution of outliers and potential features in the Hi-C matrix to detect copy number variations (CNVs) and interchromosomal translocations. HiSVision[13] introduced deep learning strategies to improve identification accuracy and adaptability to some extent. The HiSV method[14] was the first to propose the use of a significance segmentation mechanism to divide potential SV regions from the perspective of pixel similarity in the overall image, achieving good generalization results. However, these methods still have room for improvement in terms of the accuracy of candidate variation localization, the ability to distinguish between variation types, and the accuracy of identifying complex variations.

In this study, we developed a deep learning framework that transforms the task of detecting structural variations based on Hi-C data into a classic target detection problem in the field of image processing. In the field of SV detection using deep learning methods, there have been a number of innovative studies that have yielded important results[15][16].

Our core idea is that structural variations in the Hi-C contact matrix form unique spatial feature patterns, which are highly similar to object detection tasks. Hence, we transform the SV detection

problem into an object detection problem in images. First, the raw Hi-C contact matrices were preprocessed to correct for distance-dependent biases. Second, the preprocessed matrices were converted into image representations. Using a sliding-window strategy, the entire chromosome was systematically scanned, and each raw submatrix (160×160 for intra-chromosomal regions, 200×200 for inter-chromosomal regions) was resized into an 800×800 image for model input. These images were then processed by an improved RT-DETR network[17] to detect candidate SV regions. The precise breakpoint positions were subsequently determined using principal component analysis (PCA), and a Transformer model[18] was employed to capture the dynamic changes in contact frequency around the breakpoints, thereby further enhancing the ability to accurately identify and distinguish different types of SVs.

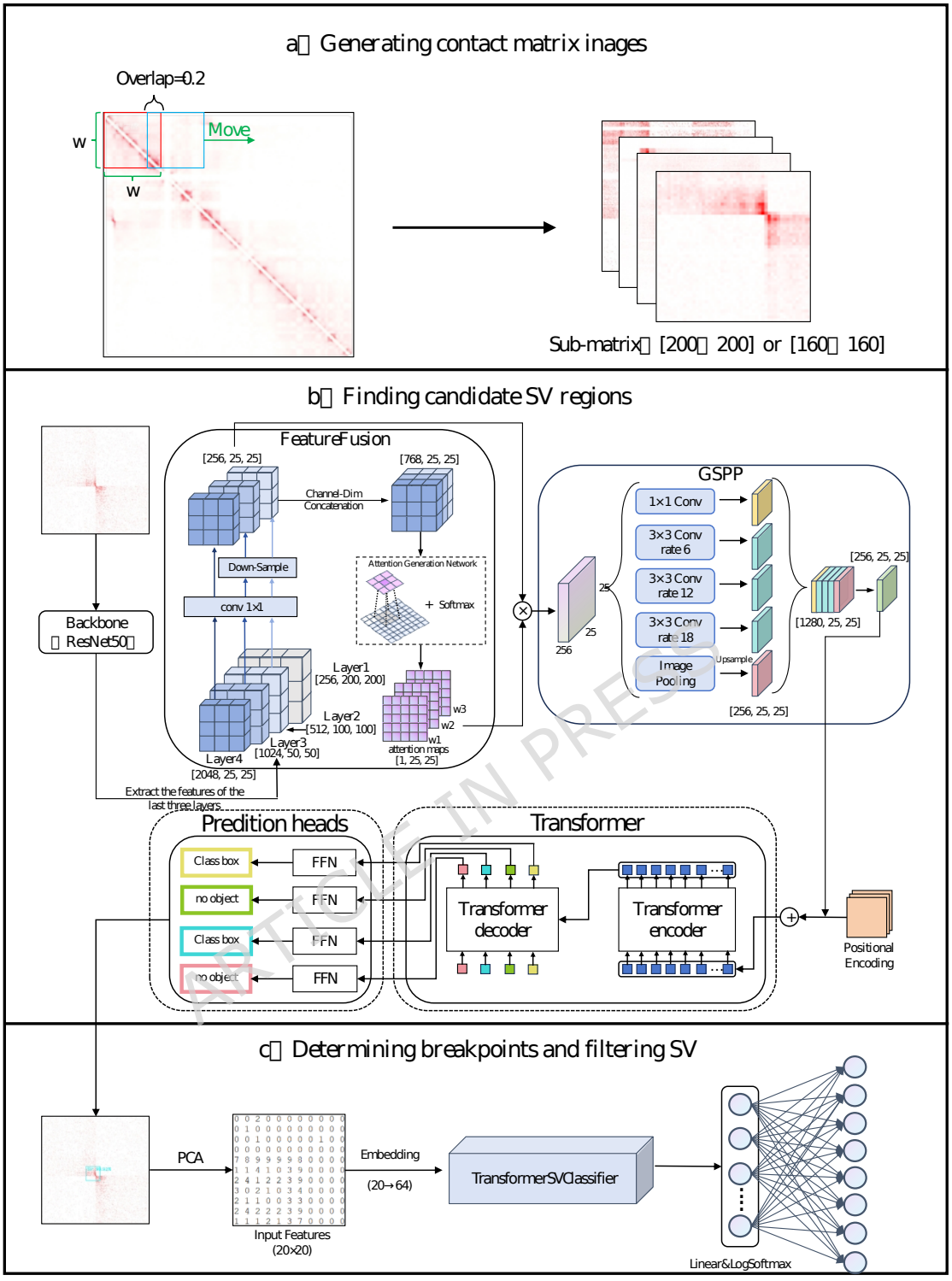

Fig 1. This figure illustrates the complete workflow of the proposed VarHiCNet model for structural variation (SV) detection from Hi-C data: (a) Generating contact matrix images: The preprocessed Hi-C contact matrix is cropped using sliding windows (with an overlap rate of 0.2) to generate 200×200 (for inter-chromosomal regions) or 160×160 (for intra-chromosomal regions) sub-matrix images, which serve as inputs for subsequent detection. (b) Finding candidate SV regions: Multi-scale features are extracted via a ResNet50 backbone; these features are then integrated (shallow/deep information) by a Feature

Fusion Module, followed by spatially adaptive feature sampling via the GSPP module. Finally, a Transformer encoder-decoder and prediction heads output the coordinates and confidence scores of candidate SV regions. (c) Determining breakpoints and filtering SV: Features of candidate regions are dimensionally reduced using PCA, then fed into the TransformerSVClassifier to achieve precise breakpoint localization and final classification of SV types (e.g., deletions, inversions, translocations).

Methods

This study transforms SV detection from “genomic coordinate alignment” to “image object detection”. The representation of structural variation in Hi-C maps is shown in Figure 2. The input is a Hi-C contact matrix, and the output is breakpoint coordinates and category labels. The overall approach can be divided into four steps:(I) Contact matrix preprocessing. Perform preprocessing operations such as distance-dependent bias correction on the original Hi-C contact matrix; (II) Generate sub-matrix. Convert the preprocessed contact matrix into an image format; (III) Searching for candidate SV regions. Use an improved RT-DETR network to perform object detection on the image, identifying candidate regions where SVs may exist; (IV) Filtering candidate SV regions. Establish a mapping relationship between the candidate regions and genomic coordinates using the final filter, precisely locate the candidate regions, and classify structural variations based on local features in the breakpoint regions.

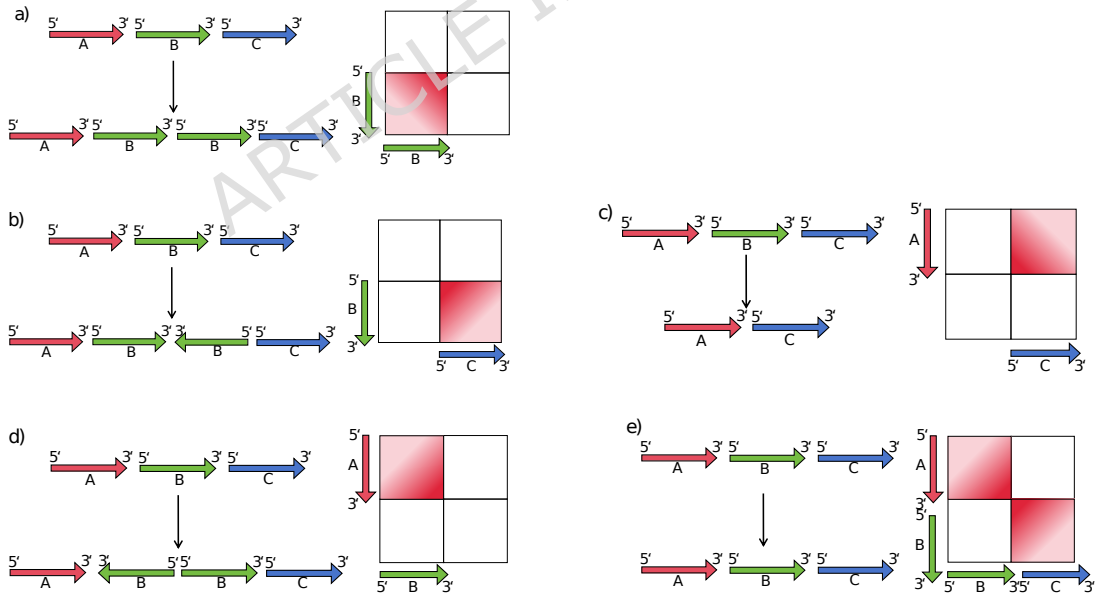

Fig 2. Distinct patterns of intra-chromosomal and inter-chromosomal SV on Hi-C map. Subfigure a (Deletion): Its core feature is a continuous blank band of contact signals at the diagonal of the Hi-C matrix, reflecting the interruption of interaction signals caused by genomic fragment deletion; Subfigure b (Forward Tandem Duplication): A symmetric signal-enhanced region is formed adjacent to the diagonal, and the extension direction of the

enhanced region is consistent with the original sequence direction of the genome; Subfigures c and d (Backward Tandem Duplication): Although local signal enhancement also exists, the spatial direction of the enhanced region is opposite to that of the original sequence, showing the signal feature of reverse distribution on both sides of the diagonal in the matrix; Subfigure e (Inversion): It is characterized by the reversal of the interaction direction of contact signals in the diagonal region, where the signal pattern originally extending orderly along the diagonal shows an obvious turning block—this is a direct manifestation of chromatin spatial conformation changes caused by inversion.

2.1 Contact matrix preprocessing

To enhance the detectability of structural variations (SVs) within Hi-C contact matrices, we applied a two-step preprocessing procedure to the raw matrices prior to candidate region detection. This process consisted of (i) Observed/Expected (OE) normalization for banded regions near the main diagonal and (ii) Z-score standardization for regions distal to the diagonal [19]. All Hi-C data used in this study were preprocessed into contact matrices with a 50kb resolution. While 25kb resolution retains more local details, it also amplifies noise in Hi-C contact signals (e.g., random sequencing errors), leading to a decrease in recall; 100kb resolution smooths noise but compresses the feature signals of small-scale SVs (e.g., 500kb inversions), also reducing recall. In contrast, 50kb resolution avoids noise interference from overly fine resolution while preserving the core interaction patterns of most SVs, making it the optimal choice for "noise-signal balance" in current experiments.

(1) OE transformation for near-diagonal regions

Because Hi-C contact frequency naturally decays with increasing genomic distance (bin separation) [20], we performed OE normalization within a fixed-width band surrounding the main diagonal. M is the Hi-C contact matrix. Mij is the element in the i-th row and j-th column of M. E|i-j| represents the expected contact value for all positions with a distance of |i - j|. For each entry (i,j) in this region, the OE value was calculated as follows:

$$E_d = \frac{1}{|\{(i,j)||i-j| = d, M_{i,j} > 0\}|} \sum_{|i-j|=d, M_{i,j}>0} M_{i,j} \quad (1)$$

$$OE_{i,j} = \begin{cases} \log\left(1 + \frac{M_{i,j}}{E_{|i-j|}}\right), & \text{if} E_{|i-j|} > 0 \\ 0, & \text{if} E_{|i-j|} = 0 \text{ or undefined} \end{cases} \quad (2)$$

(2) Z-score standardization for distal regions

In contrast, regions far from the diagonal (|i - j|>b) are typically sparse and noisy. To harmonize signal distributions across different genomic distances and to improve comparability of long-range weak signals, we applied distance-specific Z-score normalization to the O/E contact matrix M. For each genomic distance d, with Nd denoting the number of elements along the corresponding offset diagonal, we computed the mean and standard deviation as:

$$\mu_d = \frac{1}{N_d} \sum_{|i-j|=d} M_{i,j}, \sigma_d = \sqrt{\frac{1}{N_d} \sum_{|i-j|=d} (M_{i,j} - \mu_d)^2} \quad (3)$$

$$Z_{i,j} = \begin{cases} \frac{M_{i,j} - \mu_{|i-j|}}{\sigma_{|i-j|}}, & \text{if} \sigma_{|i-j|} > 0 \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

In Equations (3) and (4), the parameter d represents the absolute difference |i-j| between row index i and column index j in the Hi-C contact matrix (unit: bin), whose values cover all possible |i-j| in the matrix and correspond to different genomic distances (under the 50kb resolution of this study, d=1 represents a genomic distance of 50kb). Here, we define the band-width threshold b=50 (optimized via performance validation in Tables S16-S17): elements satisfying $d \leq b$ belong to the band region (processed by OE normalization in Equations (1)-(2)), while elements with $d > b$ are assigned to the non-band region (processed by the Z-score standardization in Equations (3)-(4)). During calculation, the mean and standard deviation of elements on the corresponding offset diagonal are calculated separately for each d to achieve distance-specific Z-score standardization, eliminating the bias of Hi-C signals decaying with genomic distance. This differential preprocessing strategy preserved the original information for inter-chromosomal interactions while simultaneously enhancing both local and long-range intra-chromosomal features. As a result, the processed matrices provide more robust inputs for downstream SV detection.

2.2 Generate sub-matrix

This study designed an image transformation strategy based on sliding windows to address the characteristics of Hi-C contact matrices. During data processing, we first divide the contact matrix into two types based on chromatin spatial interaction features: intra-chromosomal and inter-chromosomal. Then we use dynamic window scanning technology to systematically traverse the entire contact matrix with predefined detection window sizes. Inter-chromosomal regions default to a window size of 200×200, while intra-chromosomal regions are set to 160×160. We adopted a 20% window overlap to ensure that key structural variation signals are not lost due to boundary effects. In the specific implementation, starting from the top-left corner of the matrix, submatrices are sequentially extracted in a sliding manner to divide the matrix into multiple subregions. In the formulas (5)–(8), *w* denotes the size of the sliding window (unit: bin), with a value of 200 for inter-chromosomal contact matrices (corresponding to a 200×200 window) and 160 for intra-chromosomal contact matrices (corresponding to a 160×160 window), which is set based on the spatial interaction characteristics of different matrix types, m represents the submatrices obtained after partitioning, and M refers to the entire matrix being partitioned.The following formula can express the submatrix m in the i-th row and j-th column of M:

$$c_{top_l} = (w \times 0.8) \times (i - 1) \quad (5)$$

$$c_{top_r} = c_{top_l} + w \quad (6)$$

$$c_{bottom_l} = (w \times 0.8) \times (j - 1) \quad (7)$$

$$c_{bottom_r} = c_{bottom_l} + w(8)$$

$$m = M[c_{top_l}:c_{top_r}, c_{bottom_l}:c_{bottom_r}](9)$$

The corresponding region is defined by four boundary coordinates: Ctop_l, Ctop_r, Cbottom_l, and Cbottom_r. Ultimately, each submatrix is converted into an 800×800-pixel RGB image. The 800×800 dimension represents the unified image resolution for model input, derived from resizing 160×160 (intra-chromosomal) or 200×200 (inter-chromosomal) raw sub-matrices via bilinear interpolation. Instead of using a single-channel grayscale representation, we chose to convert the submatrices into RGB format, which leverages the intrinsic characteristics of Hi-C signals and better aligns with the downstream model. Hi-C data encode contact-frequency distributions between genomic bins, and structural variations (SVs) introduce characteristic local abnormalities such as signal interruptions or abnormal intensification. Mapping these signals into a single grayscale channel compresses their dynamic range, making subtle SV-related deviations more susceptible to being masked by background noise. In contrast, RGB images provide three complementary intensity channels, enabling a more expressive representation of high, medium, and low contact-frequency patterns, thereby enhancing the separability between SV-affected and normal regions. Furthermore, RGB representation ensures compatibility with the ImageNet-pretrained ResNet-50 backbone used in this study, whose pretrained filters are optimized for multi-channel inputs and are effective in capturing diverse spatial features such as edges and textures. Consequently, converting the Hi-C submatrices into RGB format not only improves feature expressiveness but also enables efficient utilization of pretrained weights.

This design leverages the intrinsic characteristics of Hi-C signals and better fits the downstream model.This processing method not only preserves the spatial interaction features in the original data but also converts complex genomic interaction information into an image format suitable for deep learning models, laying an important foundation for subsequent variant detection.

2.3 Searching for candidate SV regions

We introduced feature fusion and spatial adaptive fusion mechanisms based on the original RT-DETR model and applied the improved architecture to detect structural variation (SV) regions. The overall framework is illustrated in Figure 1(b). The model mainly comprises five modules: the Backbone, Feature Fusion Module, GSPP Module, Transformer Encoder [21], and Prediction Heads.

Originally proposed by Zhou et al., RT-DETR (Real-Time DETR) is a Transformer-based end-to-end object detector that eliminates the anchor-dependent design of traditional detectors. It offers two notable advantages: (1) real-time inference (≥30 FPS on GPUs) through a “query-attention mechanism” that directly matches object queries to feature maps, avoiding complex anchor generation; and (2) the ability of the Transformer encoder–decoder to capture long-range dependencies, enabling robust modeling of global spatial patterns. However, when applied directly to Hi-C–based SV detection, RT-DETR exhibits fundamental limitations. It relies solely on top-layer backbone features (e.g., ResNet50 Layer 4) for single-scale extraction, making it difficult to simultaneously capture “local breakpoint details”

(e.g., subtle signals from small deletions) and “global SV patterns” (e.g., symmetric interaction signatures of large inversions). High-resolution lower-layer features retain local details but lack global context, whereas low-resolution deeper features capture global structure but lose fine-grained information. Moreover, its fixed convolutional receptive fields cannot adapt to core characteristics of Hi-C data—namely, heterogeneous SV spans (50–200+ bins) and distance-dependent contact decay—leading to missed detection of large translocations with small receptive fields and blurred small SV signals with excessively large receptive fields.

To address the “single-scale feature deficiency” in the original RT-DETR, we designed a Feature Fusion Module tailored to the properties of Hi-C matrices, with the goal of integrating multi-level features to preserve both local and global SV information. The module first extracts feature maps from three key layers of the ResNet50 backbone (Layer 2, Layer 3, Layer 4): Layer 2 (resolution 1/8, 512 channels) provides high-resolution features for precise breakpoint localization; Layer 3 (resolution 1/16, 1024 channels) offers a balance between local and regional patterns; and Layer 4 (resolution 1/32, 2048 channels) captures global SV interaction structures. Feature alignment is then performed by applying 1×1 convolutions to unify all feature channels to 256 (reducing computational cost and preventing channel imbalance), followed by bilinear interpolation to upsample Layer 3 and Layer 4 to the spatial resolution of Layer 2, ensuring that all feature maps are spatially consistent. A lightweight attention network (two 1×1 convolutions and a Sigmoid activation) is then applied to learn pixel-wise fusion weights. For regions containing small-scale SVs (<100 bins), the network assigns higher weights to Layer 2 features to highlight fine breakpoint details; for regions containing large-scale inter-chromosomal translocations, it increases the weight of Layer 4 features to enhance global pattern representation. The final fused feature map is obtained by weighted summation, effectively combining low-level detailed cues with high-level semantic information.

To overcome the “fixed receptive field mismatch” in the original RT-DETR and mitigate spatial distortion caused by distance-dependent contact decay in Hi-C data, we propose the GSPP (Genomic Spatial Pyramid Pooling) Module. Its design is specifically tailored for SV detection in genomic contact matrices. The module achieves accurate multi-scale SV representation through multi-branch feature extraction and fusion. It first includes a 1×1 convolution branch that compresses channel dimensions while retaining essential signals, avoiding unnecessary computation. Three parallel atrous convolution branches with dilation rates of 6, 12, and 18 are then incorporated. These heterogeneous dilation rates expand the receptive field flexibly, enabling targeted extraction of small-scale (e.g., deletions), medium-scale (e.g., intra-chromosomal inversions), and large-scale (e.g., translocations) SV features. Padding and dilation are carefully designed to maintain consistent output sizes across scales. Additionally, a global average pooling branch compresses the input feature map into a 1×1 representation to capture global SV interaction patterns, adjusts the channels via a 1×1 convolution, and upsamples the output back to the original spatial size using bilinear interpolation, ensuring complementarity between local and global information. After concatenating all branch outputs along the channel dimension, the

combined features pass through a fusion block consisting of a 1×1 convolution, batch normalization, ReLU activation, and a Dropout layer. This not only integrates multi-branch information but also mitigates overfitting through regularization, producing a unified, highly discriminative feature map. To further mitigate overfitting in the deep architecture, multiple regularization strategies were incorporated into VarHiCNet. Dropout layers (rate = 0.3) were inserted into the fusion blocks of the GSPP module to reduce neuron co-adaptation, and L2 weight decay (5×10^{-5}) was applied to all fully connected layers in the Transformer encoder to constrain parameter magnitudes. In addition, batch normalization was added after all convolutional layers to stabilize intermediate feature distributions and enhance generalization. Together, these measures effectively prevent the model from overfitting to redundant local patterns in the training data.Compared with traditional feature extraction modules, this design covers the typical SV span range in Hi-C data without requiring additional complex structures and enhances the separability between SV signals and background noise through multi-branch cooperation, making it highly suitable for the spatial heterogeneity of genomic interaction data.

In the final stage of the model architecture, multi-scale features that have been deeply integrated are input into a Transformer-based decision module. This module utilizes self-attention mechanisms to model long-range dependencies between features, enabling it to capture cross-regional interaction patterns in Hi-C matrices. Through feature transformations performed by a multi-layer encoder, the system ultimately outputs detection results that include predicted box spatial coordinates and reliability assessments. During application, we set a confidence threshold of 0.8 to filter out high-confidence candidate regions. These regions are marked as potential genomic structural variation sites and proceed to subsequent detailed analysis. This design ensures the reliability of detection results while providing high-quality candidate targets for downstream analysis. Experiments demonstrate that this threshold setting can effectively cover over 90% of true variation regions while maintaining high accuracy.

Each 800×800 RGB image corresponds to a Hi-C submatrix obtained via sliding window scanning (Section 2.2). If an image is generated from a window with the starting genomic bin Sstart (calculated using Equations 5–9 in Section 2.2), the genomic bin corresponding to pixel (x, y) in the image is S_start+(x×W)/800 (for the x-axis) and S_start+(y×W)/800 (for the y-axis), where W represents the window size (160 for intra-chromosomal regions and 200 for inter-chromosomal regions).

To optimize bounding-box regression and confidence estimation in the improved RT-DETR network, a combined loss function was employed for the candidate SV region detection stage.First, the regression branch adopts the GIoU loss, which alleviates the misalignment between predicted boxes B and ground-truth boxes B^{gt}. The GIoU loss is defined as follows:

$$L_{GIoU} = 1 - \left(IoU - \frac{|C - (B \cup B^{gt})|}{|C|} \right) \quad (10)$$

Here, C denotes the smallest enclosing box of B and B^{gt}. For the confidence classification branch, Focal Loss was used to address class imbalance. Its formulation is:

$$L_{Focal} = - \alpha_t (1 - p_t)^{\gamma} \log(p_t) \quad (11)$$

Where α =0.25,γ= 2,and p_t represents the predicted confidence for the target class. The total detection loss is expressed as:

$$L_{Detection} = 0.5 \times L_{GIoU} + 0.5 \times L_{Focal} \quad (12)$$

This balanced combination ensures stable regression performance and enhanced robustness against hard negative samples.

2.4 Filtering candidate SV regions

By analyzing changes in the topological structure of contact matrices near breakpoints, the system distinguishes between different types of genomic rearrangement events. As shown in Figure 3, for structural variations occurring within a single chromosome, we observed that deletion events form characteristic interaction signal interruption patterns (category a). At the same time, repetitions with different orientations exhibit specific interaction enhancement features (categories b–d correspond to forward, reverse, and tandem repetitions, respectively). Balanced translocations (Category e) and inversion events (Category f) exhibit unique spatial configuration changes in the Hi-C matrix, while small-scale structural variations (Categories g–h) show subtle alterations in local patterns. When variations involve two different chromosomes, unbalanced translocations can be classified into four typical patterns (Categories a–d) based on spatial configuration differences, while balanced translocations are further divided into two types (Categories e–f) based on distinct characteristics. This classification scheme comprehensively considers multidimensional information, such as the length characteristics and spatial orientation of the variation fragments.

As shown in Figure 2 c), after initially screening out regions that may contain structural variations, we cropped the corresponding submatrices from the original Hi-C contact matrix. We performed principal component analysis (PCA) on their row vectors and column vectors [24]. By analyzing the sign change positions of the first principal component, we identified the points with the most significant mutation signals as SV breakpoints. We expanded each candidate SV region by 10 bins in each of the four directions around the breakpoint, encoding each region as a 20×20 contact frequency matrix. These matrices contain classifiable feature information near the breakpoint. Each row is treated as a token, which is mapped to a high-dimensional representation via a linear embedding layer and then input into the Transformer encoder. the decision head is implemented as a standalone TransformerSVClassifier, tailored for the 9-class SV prediction task (eight SV subtypes plus background). It first linearly embeds each 20×20 local feature vector into a 64-dimensional space, followed by a two-layer Transformer encoder (four attention heads, feed-forward dimension of 128) that integrates both local and global contextual information. A sequence-averaged aggregation layer is then applied, and a linear classification head generates log-softmax-normalized class logits.

For the Transformer-based SV classification stage, weighted cross-entropy loss was used to mitigate the imbalance across the eight SV subtypes.The class weights were computed according to:

$$w_c = \frac{N_{total}}{N_c \times C} \quad (13)$$

Where N_{total} is the total number of training samples, N_c denotes the sample count of class c, and □=9 is the total number of SV categories. The weighted cross-entropy loss is defined as:

$$L_{Classification} = -\sum_{c=1}^{9} w_c \cdot y_c \cdot \log(p_c) \quad (14)$$

In this equation, y_c is the one-hot encoded label of class c, and p_c is the softmax probability for that class. This design encourages the model to place greater emphasis on rare SV types while maintaining strong performance on common categories.

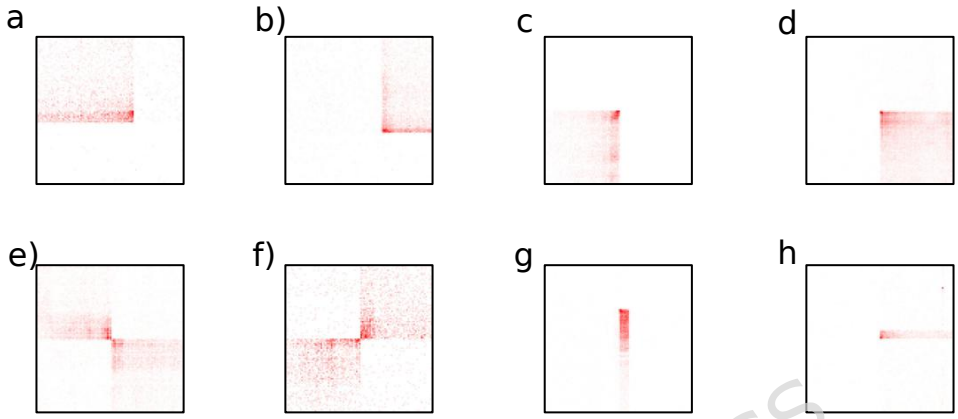

Fig 3. Schematic diagram of breakpoints in different types of structural variations. Figure a) represents deletion events, Figures b-d) represent duplication events in different orientations, e) is a balanced translocation, f) represents inversion events, and g) and h) are small-scale structural variations.

2.5 Statistics and Reproducibility

Our dataset was constructed using seven representative cancer cell lines: K562, T47D, SK-N-MC, Caki2, LNCaP, NCI-H460, and HelaS3. Based on the high-confidence structural variation (SV) list provided in Reference[25], we rigorously screened the data to identify SV sites with typical characteristics. To ensure the deep learning model receives sufficient training data, we designed a data augmentation scheme. First, centred on each SV breakpoint, we expanded 100 bins in each of the four directions (up, down, left, and right), extracting a 200×200 submatrix as the base sample. Subsequently, we further expanded the samples by translating the breakpoint positions to different locations within the submatrix. This 200×200 submatrix is specifically prepared for constructing the training set with over 27,000 annotated images, which differs from the 200×200 window used in the sliding window scanning step for converting preprocessed Hi-C contact matrices into image representations (Section 2.2).

Additionally, we rotated each image by 90°, 180°, and 270° to enhance the model's robustness to directional changes. Using this strategy, each original variation site generated approximately 100 derived samples, ultimately constructing a high-quality training set containing over 27,000 annotated images. To optimize model performance, pre-trained ResNet50 network parameters were introduced into the network backbone, and fine-tuning training was conducted over 30 epochs to achieve rapid

convergence. In the design of the feature filtering module, a 20×20 local contact matrix near the breakpoint was selected as the positive sample, At the same time, the misjudgment areas in the model prediction process were collected as negative samples. By performing mirror flipping and other operations on the positive samples, data diversity was further enhanced, effectively improving the model's ability to recognize complex variant features.

To optimize the training process and ensure model convergence and generalization, the following key training parameters were adopted: the optimizer used was AdamW with a learning rate of 1×10^{-4}, weight decay of 5×10^{-5}, and momentum parameters betas=(0.9, 0.999); the total number of training epochs was set to 100, including a 3-epoch warm-up phase where the learning rate linearly increased from 1×10^{-6} to 1×10^{-4}, followed by a linear decay of the learning rate to 1×10^{-5} in epochs 4–30; and the batch size was configured as 4 to balance training efficiency and GPU memory constraints.

Results

We selected four existing methods (HiC_breakfinder, HiNT-TL, HiSVision, HiSV) and compared their performance across six cancer cell lines (K562, T47D, Caki2, NCI-H460, SK-N-MC, LNCaP). The gold standard dataset in this study was obtained by screening a set of high-confidence SV sites, which were derived from [26] and validated by at least two platforms (whole-genome sequencing, optical mapping, and Hi-C sequencing). Translocations with a variation length greater than 1 Mb (including both intra-chromosomal and inter-chromosomal variations) were retained from the high-confidence structural variation sites as our gold standard set. The evaluation framework for performance comparison is as follows: recall, precision, and F1-score are used as evaluation metrics for each method. After detecting the final structural variation, if the coordinates of the variation breakpoint differ by less than 2 bins from the corresponding accurate variation breakpoint coordinates, it is deemed a correct prediction.

To systematically evaluate the performance and generalization ability of the proposed method in the task of structural variation detection, we designed two sets of cross-experimental scenarios to analyze the model's transferability across different cell line combinations, based on seven cancer cell lines (HelaS3, K562, T47D, SK-N-MC, Caki2, LNCaP, NCI-H460). In each experiment, three cell lines were used, with HelaS3 serving as a supplementary dataset for training, while the model's performance was evaluated across the remaining three cell lines. When evaluating the performance of different methods in structural variation detection tasks, some methods have limitations in parameter settings and applicable scenarios. The HiSV method typically requires manual parameter adjustment across different cell lines to adapt to sample characteristics, such as significance thresholds and regularization coefficients. Therefore, when assessing HiSV's performance, this study ran HiSV with different parameters for different cell line datasets and retained the best performance as the final result. Second, the HiNT-TL method is inherently only applicable to detecting interchromosomal translocations (inter SV). It identifies translocation events based on the non-uniformity of interaction frequencies between

chromosome pairs. Therefore, HiNT-TL was excluded when evaluating the performance of intrachromosomal variation detection.

3.1 Benchmarking in Caki2, LNCaP, and NCI-H460 cell lines

In the first set of experiments, we selected K562 (leukemia), T47D (breast cancer), and SK-N-MC (neuroblastoma) as training samples. Then we tested the model on three cell lines—Caki2, LNCaP, and NCI-H460—derived from the urinary and pulmonary systems, respectively, to evaluate the model's generalization ability across cancer types and different tissue backgrounds. The results showed that our proposed method demonstrated good stability and high F1 scores in detecting both interchromosomal and intrachromosomal structural variations. Especially in complex samples like NCI-H460, the accuracy of identifying balanced translocations and inversions was significantly higher than that of other methods. This is mainly because the model has the ability to perform dual reconstruction modeling of variation image patterns and local principal component trends.

Among the methods used for comparison, HiC_breakfinder relies on global statistical modeling and maintains a high recall rate, but it often produces a large number of false positives and has low accuracy, especially in the Caki2 sample, where the accuracy drops significantly; HiNT-TL is suitable for detecting large-scale interchromosomal SVs, but its performance is not satisfactory when dealing with intrachromosomal SVs and complex inversions; HiSV achieves moderate performance under unsupervised conditions, with strong recall capability but unclear classification, leading to false positives; HiSVision, leveraging its DETR+LSTM architecture, performs exceptionally well in interchromosomal SV detection, particularly demonstrating high precision in detecting specific balanced translocations, though it remains limited in intrachromosomal SV detection and boundary determination. As shown in Table 1, VarHiCNet achieves competitive performance in inter-chromosomal SV detection across Caki2, LNCaP, and NCI-H460 cell lines. Specifically, it matches HiSVision’s F1-score (0.4570) in Caki2 and (0.7272) in NCI-H460, while outperforming HiSV, HiNT-TL, and HiC_breakfinder in LNCaP with an F1-score of 0.6666 by balancing recall and precision effectively. For intra-chromosomal SV detection (Table 2), VarHiCNet demonstrates superior overall performance: it achieves the highest F1-score (0.7500) in Caki2 by maintaining both high recall (0.8181) and precision (0.6923), and shows stable performance in LNCaP (F1=0.6897) and NCI-H460 (F1=0.6666), outperforming comparative methods in balancing detection accuracy and reliability.Overall, our method achieves superior F1 scores compared to HiSVision and HiC_breakfinder in both intra-chromosomal and inter-chromosomal scenarios, demonstrating its excellent structural adaptability and category discrimination capabilities.

Table 1 Performance comparison of SV callers (inter-chromosomal SVs) on Caki2,LNCaP and NCI-H460

	Caki2			LNCaP			NCI-H460		
	Recall	Precision	F1-Score	Recall	Precision	F1-Score	Recall	Precision	F1-Score

HISV	0.3636	0.4000	0.3809	0.4444	0.2500	0.3199	0.5714	0.8000	0.6666
HINT-TL	0.4545	0.2941	0.3571	0.3333	0.2000	0.2499	0.2857	0.1052	0.1537
HiSVision	0.3636	0.6153	0.4570	0.6666	0.8571	0.7499	0.5714	1	0.7272
Hic_breakfinder	0.6363	0.3414	0.4443	0.8888	0.6153	0.7271	1	0.5384	0.6999
VarHiCNet	0.3636	0.6153	0.4570	0.5555	0.8333	0.6666	0.5714	1	0.7272

Table 2 Performance comparison of SV callers (intra-chromosomal SVs) on Caki2,LNCaP and NCI-H460

	Caki2			**LNCaP**			**NCI-H460**		
	Recall	Precision	F1-Score	Recall	Precision	F1-Score	Recall	Precision	F1-Score
HISV	0.4545	0.2777	0.3447	0.5714	0.8000	0.6666	0.3333	0.5000	0.3999
HINT-TL	/	/	/	/	/	/	/	/	/
HiSVision	0.5454	0.6000	0.5714	0.5714	0.8000	0.6667	0.3333	0.5000	0.3999
Hic_breakfinder	0.8181	0.3913	0.5293	1	0.4375	0.6086	1	0.7500	0.8571
VarHiCNet	0.8181	0.6923	0.7500	0.7142	0.6667	0.6897	0.6666	0.6666	0.6666

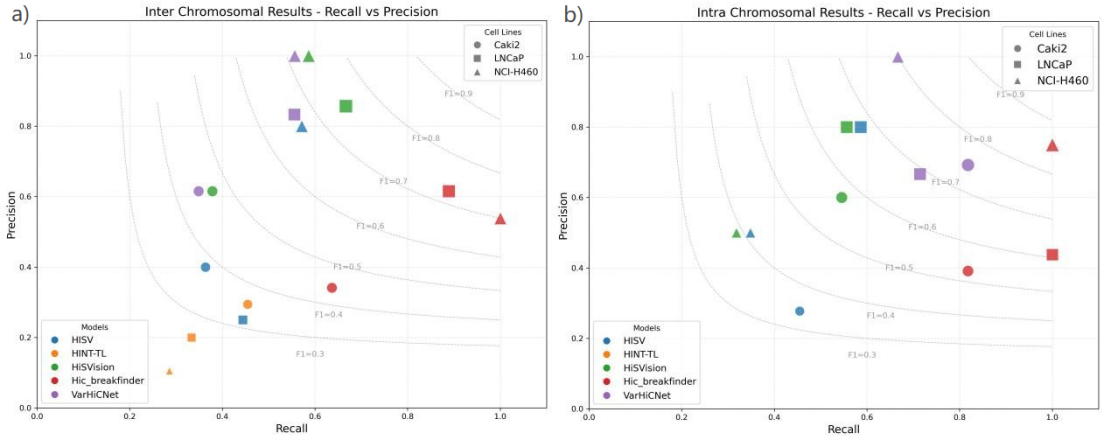

Fig.4 Recall-precision comparison scatter plot of structural variation detection tools in Caki2, LNCaP, and NCI-H460 cell lines. Figure a) shows the results between chromosomes, while Figure b) shows the results within chromosomes.

3.2 Benchmarking in K562, T47D, and SK-N-MC cell lines

In the second set of experiments, we reversed the training and testing sets, selecting Caki2, LNCaP, and NCI-H460 as the training set and evaluating the model's cross-organization generalization ability in K562, T47D, and SK-N-MC. This set of experiments further validated the robustness of our method: in T47D and SK-N-MC samples, our method demonstrated high recall and precision in the detection of intra-chromosomal SVs, with F1 scores significantly leading the pack. Particularly in the SK-N-MC sample, our model demonstrates outstanding recognition capabilities for complex inversions and long-segment balanced translocations, which is closely related to the detection model's advantage in capturing symmetrical morphological patterns.

When comparing multiple methods, HiSVision maintains a leading position in inter-chromosomal SV detection, with high precision in multiple samples; however, its recall rate for small-scale intrachromosomal SVs is relatively low, resulting in detection omissions. HiC_breakfinder consistently achieves a high overall recall rate, but its precision drops sharply in K562 samples with high background noise. HiNT-TL demonstrates limited overall performance except for specific significant inter-chromosomal SV detections; HiSV tends to misjudge in images with blurred boundaries or non-typical variations, resulting in consistently low accuracy rates. In this task set, our proposed method achieves overall leading F1 scores across three test samples, particularly excelling in maintaining both high recall rates and low false positive rates compared to other methods.

Table 3 Performance comparison of SV callers (inter-chromosomal SVs) on K562,T47D and SK-N-MC

	K562			T47D			SK-N-MC		
	Recall	Precision	F1-Score	Recall	Precision	F1-Score	Recall	Precision	F1-Score
HISV	0.1818	0.3333	0.2352	0.4137	0.9230	0.5713	0.8888	0.421	0.5713
HINT-TL	0.2727	0.2195	0.2432	0.2758	0.1702	0.2104	0.2222	0.6666	0.3333
HiSVision	0.3333	0.6111	0.4314	0.5862	0.8947	0.7083	1	1	1
Hic_breakfinder	0.4240	0.3589	0.3880	0.6200	0.5800	0.5990	1	0.9	0.9473
VarHiCNet	0.3939	0.6500	0.4889	0.5517	0.7619	0.6429	1	1	1

Table 4 Performance comparison of SV callers (intra-chromosomal SVs) on K562,T47D and SK-N-MC

	K562			T47D			SK-N-MC		
	Recall	Precision	F1-Score	Recall	Precision	F1-Score	Recall	Precision	F1-Score
HISV	0.5000	0.3750	0.4285	0.4347	0.7142	0.5404	0.8888	0.421	0.5713
HINT-TL	/	/	/	/	/	/	/	/	/
HiSVision	0.2500	0.4000	0.3076	0.3636	0.6428	0.4864	0.6250	0.8333	0.7142
Hic_breakfinder	0.4583	0.4780	0.4679	0.5454	0.7058	0.6499	0.6250	0.8333	0.7142
VarHiCNet	0.4583	0.5882	**0.5152**	0.5000	0.8462	0.6284	0.6250	0.8333	0.7142

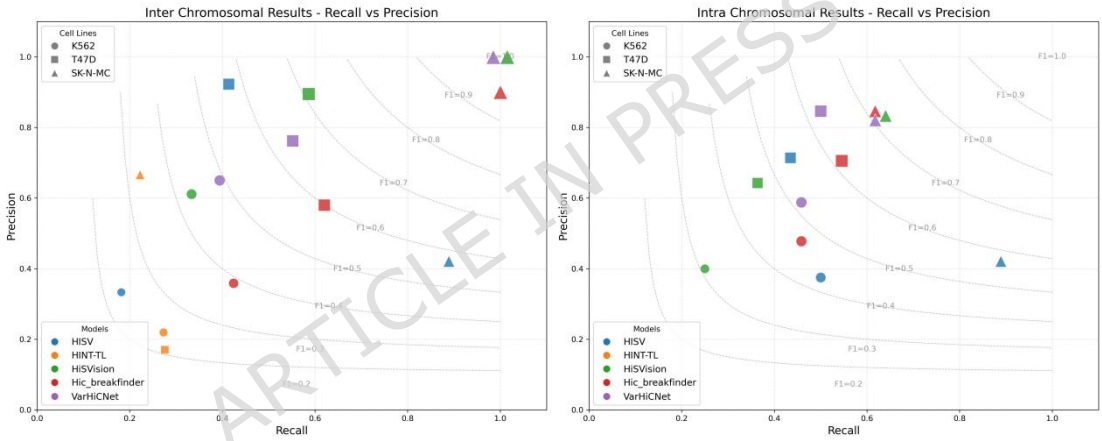

Fig.5 Recall-precision comparison scatter plot of structural variation detection tools in K562, T47D, and SK-N-MC cell lines. Figure a) shows the results between chromosomes, while Figure b) shows the results within chromosomes.

In PacBio sequencing, GCphase was compared with WhatsHap, HapCUT2, and LongPhase by using the whole-genome sequencing data of the human HG002 sample provided by GIAB (PacBio CCS 15 kb_20 kb chemistry2) at coverage depths of 20x, 30x, and 50x. Similar to the Nanopore experiments, the evaluation was conducted by using the "compare" method of WhatsHap. The VCF files containing phasing information obtained from GCphase and the other three methods were compared against the standard set of the human HG002 sample provided by GIAB. This comparison resulted in the final evaluation results (Tables 3/4). In terms of the number of phased SNPs, there was still no significant difference among the four methods. In the number of blocks, there is little difference among the four methods, but GCphase has the smallest number of blocks among the three depths (50x, 30x, 20x). In terms of the Hamming distance, no significant difference is observed among the four methods. However,

GCphase consistently outperforms the other methods in two coverage depths (30x, 20x) and ranks in the second position in the 50x coverage depth. In terms of accuracy, due to the higher accuracy of PacBio sequencing data compared to that of the Nanopore sequencing data, the switch error performance of each program is better in PacBio sequencing data than that in Nanopore sequencing data. However, similar to the Nanopore sequencing data, GCphase still outperforms the other methods in all three coverage depths, with switch error rates of approximately 0.15%.

Since EagleC's training set includes the 6 cell lines in our test set, direct comparison on these real datasets would risk data leakage, so it was not included in the original real-dataset baseline comparison.

3.3 Supplementary Experiments

We conducted an extended set of supplementary experiments, with detailed results reported in Supplementary Tables S1–S17. These experiments systematically examine the effects of image resolution and Hi-C bin size on both inter- and intra-chromosomal SV detection, validate the selection of sliding-window configurations, and analyze the sensitivity of model performance to different confidence thresholds. In addition, extensive ablation studies are performed across multiple cell lines to quantify the individual contributions of the Feature Fusion and GSPP modules. We further evaluate the generalization capability of VarHiCNet through leave-one-out experiments on six cancer cell lines and benchmark its performance against existing methods on simulated datasets. Collectively, these supplementary analyses provide strong empirical support for the design choices adopted in this study and demonstrate the robustness and generalizability of VarHiCNet across diverse experimental settings.

Discussion

The VarHiCNet method proposed in this study transforms the detection of structural variations in Hi-C data into an object detection problem in images. By integrating an improved RT-DETR network, a feature fusion mechanism, and a Transformer encoder, it achieves precise identification of multiple types of structural variations. Experimental results demonstrate that VarHiCNet exhibits high stability and accuracy across six cancer cell lines, particularly outperforming existing methods in identifying balanced translocations and inversions within complex samples such as NCI-H460 and SK-N-MC.

However, we also observed several limitations of VarHiCNet. A statistical analysis of misdetection cases across the six cell lines indicates that the model errors are mainly concentrated in two types of structural variations. The first is small-scale inversions (500 kb–1 Mb), which exhibit weak signal patterns in Hi-C contact matrices and are therefore easily confounded with local interaction fluctuations arising from normal chromatin folding; this issue is particularly pronounced in the NCI-H460 and SK-N-MC cell lines, whose genomes contain abundant short-segment repetitive regions. The second involves complex backward tandem duplications (e.g., nested duplications), whose overlapping interaction patterns are difficult to effectively capture using the current approach.

These findings indicate that deep learning models are capable of capturing complex spatial conformation features within Hi-C matrices, thereby effectively compensating for the limitations of traditional sequencing and statistical methods in detecting multi-breakpoint rearrangements and complex structural variations. Comparative experiments with other methods further underscore the advantages of this approach.While HiC_breakfinder demonstrates good recall, its high false positive rate compromises reliability; HiNT-TL is applicable to large-scale interchromosomal translocation detection but struggles with intrachromosomal variations and complex events; HiSV and HiSVision, though incorporating deep learning frameworks to enhance cross-sample applicability, remain limited in boundary precision and small-scale variation detection. In contrast, VarHiCNet significantly strengthens feature representation through multi-scale feature fusion and a spatially adaptive fusion mechanism. Leveraging the long-range dependency modeling capabilities of Transformers, it simultaneously captures both global signals and local mutation features, thereby demonstrating superior generalization across diverse variation types.

Nevertheless, several limitations remain. First, VarHiCNet training relies heavily on large volumes of high-quality annotated data. Although data augmentation partially alleviates sample scarcity, its performance still requires further validation on larger single-cell Hi-C datasets or cross-species datasets. Second, the current classification module primarily depends on 20×20 matrix features from local breakpoint regions. Future work could explore incorporating broader contextual information, which may further improve the recognition of complex variation patterns.

Conclusion

VarHiCNet innovatively transforms structural variation detection in Hi-C data into an image object detection problem, establishing an end-to-end detection and classification framework. Through an improved RT-DETR network, a multi-scale feature fusion mechanism, and a Transformer encoder, VarHiCNet achieves high-precision and stable structural variation identification across multiple cancer cell lines. Experimental results demonstrate that the method excels in detecting balanced translocations, inversions, and complex structural variations, and overall outperforms existing mainstream tools. This study provides a novel technical pathway for structural variation detection using Hi-C data and offers important insights into the role of three-dimensional genomic architecture in tumorigenesis and progression.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

The high-confidence SV list used in this study is sourced from Reference [25] and can be accessed via the following link: https://github.com/000425/data

Competing interests

The authors declare that they have no competing interests.

Funding

This research was supported by the Henan Provincial Department of Science and Technology Research Project (Grant No. 242102210097, 242102210110).

Authors' contributions

JQS, HJW and JFW participated in the design of the study and the analysis of the experimental results. JFW and JQS performed the implementation. HJW and HXZ prepared the tables and figures. JQS and JWL summarized the results of the study and checked the format of the manuscript. All authors read and approved the final manuscript.

Acknowledgements

References

1. Logsdon, Glennis A., et al. "Complex genetic variation in nearly complete human genomes." Nature (2025): 1-12.
2. Hwang, Michael S., et al. "Targeting loss of heterozygosity for cancer-specific immunotherapy." Proceedings of the national academy of sciences 118.12 (2021): e2022410118.
3. Nezami, M., S. Hager, and R. Shirazi. "Transforming Growth Factor and the Role of Epigenetic Aberrancies in Oncogenic Amplifications: A New Perspective in Preventive and Therapeutic Arena." Journal of Cancer Therapy 14.9 (2023): 390-407.
4. Sreenivasan, Varun KA, Verónica Yumiceba, and Malte Spielmann. "Structural variants in the 3D genome as drivers of disease." Nature Reviews Genetics (2025): 1-19.
5. Su, Xinglei, et al. "Challenges and prospects in utilizing technologies for gene fusion analysis in cancer diagnostics." Med-X 2.1 (2024): 14.
6. Micheal, Dave. "Personalized Medicine in the Genomic Age: Utilizing Whole Genome Sequencing for Specific Therapies." (2025).
7. He, Sang, et al. "Systematic benchmarking of tools for structural variation detection using short- and long-read sequencing data in pigs." iScience 28.3 (2025).
8. Alves D, Bernardi F A, Lima V C, et al. Promoting Comprehensive Care for People With Rare Diseases in a Tertiary Care Setting in Brazil: Protocol for a Mixed Methods Implementation Study[J]. JMIR Research Protocols, 2025, 14(1): e68949.
9. Moustakli, Efthalia, et al. "Long-Read Sequencing and Structural Variant Detection: Unlocking the Hidden Genome in Rare Genetic Disorders." Diagnostics 15.14 (2025): 1803.
10. Liu, Ran, et al. "Hi-C, a chromatin 3D structure technique advancing the functional genomics of immune cells." Frontiers in Genetics 15 (2024): 1377238.
11. Dixon, Jesse R., et al. "Integrative detection and analysis of structural variation in cancer genomes." Nature genetics 50.10 (2018): 1388-1398.
12. Wang, Su, et al. "HiNT: a computational method for detecting copy number variations and translocations from Hi-C data." Genome biology 21.1 (2020): 73.
13. Zhai, Haixia, et al. "HiSVision: A Method for Detecting Large-Scale Structural Variations Based on Hi-C Data and Detection Transformer." Interdisciplinary Sciences: Computational Life Sciences (2024): 1-9.
14. Li, Junping, Lin Gao, and Yusen Ye. "HiSV: a control-free method for structural variation detection from Hi-C data." PLOS Computational Biology 19.1 (2023): e1010760.
15. Luo, Junwei, et al. "BreakNet: detecting deletions using long reads and a deep learning approach." BMC bioinformatics 22.1 (2021): 577.
16. Gao, Runtian, et al. "INSnet: a method for detecting insertions based on deep learning network." BMC bioinformatics 24.1 (2023): 80.

17. Zhao, Yian, et al. "Detrs beat yolos on real-time object detection." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2024.
18. Han, Kai, et al. "Transformer in transformer." Advances in neural information processing systems 34 (2021): 15908-15919.
19. Henderi, Henderi, Tri Wahyuningsih, and Efana Rahwanto. "Comparison of Min-Max normalization and Z-Score Normalization in the K-nearest neighbor (kNN) Algorithm to Test the Accuracy of Types of Breast Cancer." International Journal of Informatics and Information Systems 4.1 (2021): 13-20.
20. Schöpflin, Robert, et al. "Integration of Hi-C with short and long-read genome sequencing reveals the structure of germline rearranged genomes." Nature Communications 13.1 (2022): 6470.
21. Álvarez-Fidalgo, David, and Francisco Ortin. "CLAVE: A deep learning model for source code authorship verification with contrastive learning and transformer encoders." Information Processing & Management 62.3 (2025): 104005.
22. Koonce, Brett. "ResNet 50." Convolutional neural networks with swift for tensorflow: image recognition and dataset categorization. Berkeley, CA: Apress, 2021. 63-72.
23. Liu, Shu, et al. "Path aggregation network for instance segmentation." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.
24. Maćkiewicz, Andrzej, and Waldemar Ratajczak. "Principal components analysis (PCA)." Computers & Geosciences 19.3 (1993): 303-342.
25. Wang, Xiaotao, Yu Luan, and Feng Yue. "EagleC: A deep-learning framework for detecting a full range of structural variations from bulk and single-cell contact maps." Science Advances 8.24 (2022): eabn9215.
26. Dixon, Jesse R., et al. "Integrative detection and analysis of structural variation in cancer genomes." Nature genetics 50.10 (2018): 1388-1398.