Chromosomal Structural Abnormality Diagnosis by Homologous Similarity

Juren Li*
jrlee@zju.edu.cn
Zhejiang University
Hangzhou, China

Fanzhe Fu*
ffanz@zju.edu.cn
Zhejiang University
Hangzhou, China

Ran Wei
ranwei@diagens.com
Hangzhou Diagens Biotechnology
Co., Ltd.
Hangzhou, China

Yifei Sun
yifeisun@zju.edu.cn
Zhejiang University
Hangzhou, China

Zeyu Lai
jerrylai@zju.edu.cn
Zhejiang University
Hangzhou, China

Ning Song
ningsong@diagens.com
Hangzhou Diagens Biotechnology
Co., Ltd.
Hangzhou, China

Xin Chen
xin.21@intl.zju.edu.cn
Zhejiang University
Hangzhou, China

Yang Yang†
yangy@zju.edu.cn
Zhejiang University
Hangzhou, China

ABSTRACT

Pathogenic chromosome abnormalities are very common among the general population. While numerical chromosome abnormalities can be quickly and precisely detected, structural chromosome abnormalities are far more complex and typically require considerable efforts by human experts for identification. This paper focuses on investigating the modeling of chromosome features and the identification of chromosomes with structural abnormalities. Most existing data-driven methods concentrate on a single chromosome and consider each chromosome independently, overlooking the crucial aspect of homologous chromosomes. In normal cases, homologous chromosomes share identical structures, with the exception that one of them is abnormal. Therefore, we propose an adaptive method to align homologous chromosomes and diagnose structural abnormalities through homologous similarity. Inspired by the process of human expert diagnosis, we incorporate information from multiple pairs of homologous chromosomes simultaneously, aiming to reduce noise disturbance and improve prediction performance. Extensive experiments on real-world datasets validate the effectiveness of our model compared to baselines.

CCS CONCEPTS

• **Applied computing → Life and medical sciences**; **Bioinformatics**.

---

*Both authors contributed equally to this research.
†Corresponding author.

KEYWORDS

Chromosome modeling, Homologous Similarity, Chromosome Structural Abnormality, Deep Neural Networks

1 INTRODUCTION

chromosomal abnormalities, which include missing, extra, or irregular portions of chromosomal DNA [23], are the root cause of various genetic and hereditary diseases. It is estimated that 6 of every 1,000 newborns have chromosomal abnormalities, which often result in *dysmorphism*, *malformations*, and *developmental disabilities* [52]. These chromosomal changes caused by abnormalities lead to serious consequences, e.g. 25% of all miscarriages and stillbirths, and 50%–60% of first-trimester miscarriages [18]. With the help of chromosomal abnormality detection, clinicians can identify more abnormalities that may result in birth defects and take further preventive measures, which is of great significance to people's livelihood and social development.

There are two types of chromosomal abnormalities: *numerical abnormality* and *structural abnormality* [10]. Numerical abnormalities involve the absence or addition of entire chromosomes [13]. For example, an extra copy of chromosome 21 will cause *Down syndrome*. Structural abnormality refers to the deletion, excess, inversion, or translocation of a certain segment of one chromosome to another chromosome [3]. For example, a partial deletion of chromosome 4 causes *Wolf–Hirschhorn syndrome*. Numerical abnormalities can be easily detected by assessing the quantity of chromosomes. However, structural abnormalities are more challenging as they are
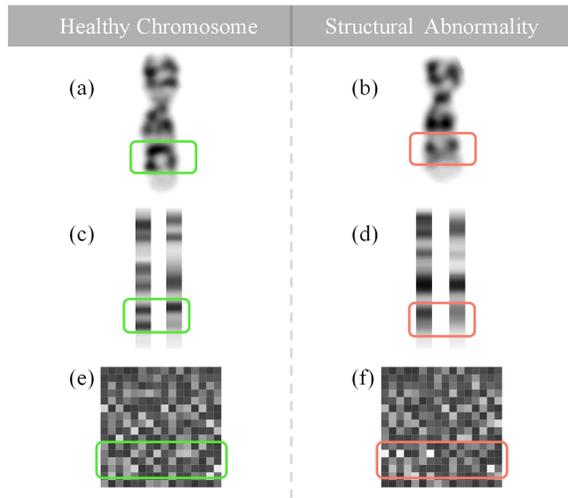
Figure 1: The homologous chromosomes with structural abnormality. (a) shows corresponding healthy region where structural abnormality occurred on (b). As for (c) and (d) are the data preprocessing we applied. (e) and (f) are the embedding of (c) and (d) by CMSBlock from HomNet model.

associated with chromosome mutation[1] [58], which makes it impossible to diagnose simply by the number of chromosomes as the numerical abnormality. The detection of structural abnormalities demands the expertise of professionals well-versed in karyotype[2] analysis. Besides, to enhance the accuracy of diagnosing, clinicians typically need to analyze chromosomes from more than 10 cells for one patient. Given that there are 46 chromosomes in each human cell [14], experts are required to visually inspect over 460 chromosome diagrams from the karyotypes. Admittedly, this inspection process is time-consuming, labor-intensive, and therefore expensive. Thus, we propose to design a deep learning-based model to automatically diagnose chromosomal structural abnormalities, aiming to expedite diagnosis, assist rapid decision-making, and reduce manual labor for clinicians, ultimately improving efficiency and accuracy in patient care.

Despite the superior performance achieved by deep learning in various applications, challenges persist in detecting chromosomal structural abnormalities. The first challenge arises from the diverse morphology of chromosomes. Structural abnormalities can occur on any of the 46 chromosomes, each with its own unique morphology, that is, different chromosomes may have different diagrams. Even within the same chromosome, variations exist due to their origins from different cells, potentially at different cell cycle [15], resulting in differing morphologies. Thus, it is necessary to design of a method capable of detecting chromosomal structural abnormalities across various chromosomal morphologies. The second challenge is the label imbalance. Compared to normal chromosomes, those with

structural abnormalities are relatively rare. Indeed, in our data, less than 4% of the chromosomes have structural abnormalities. While the imbalance problem has been addressed in many fields [26], to the best of our knowledge, it has not been well studied in the context of detecting chromosomal structural abnormalities. The third challenge lies in the inconsistent data distribution. Chromosomes obtained from different hospitals may exhibit variations in their chromosomal diagrams due to differences in equipment and environmental conditions. Additionally, the distribution of type of structural abnormalities varies from hospital to hospital, resulting in a different distribution of data from different hospitals. These factors make the data distribution of chromosomes inconsistent and make it challenging to diagnose the structural abnormalities.

To address the aforementioned challenges, we propose a domain knowledge based framework, *HomNet*, to model chromosomes and then diagnose structural abnormalities. To tackle the first challenge, we aim to capture common characteristics of structural abnormalities across chromosomes with varying morphologies. For a pair of homologous chromosomes[3], their diagrams exhibit high similarity, since they share the same chromosome type and in the same cell. However, when a pair of homologous chromosomes has a structural abnormality, their diagrams typically manifest differences in the region where the abnormality occurs, as illustrated in Figure 1. Therefore, we propose to diagnose structural abnormalities by capturing the differences between homologous chromosomes based on the homologous similarity. Addressing the label imbalance problem, where there are only a few abnormal samples, we utilize a large number of normal chromosomes from healthy individuals to construct artificial structural abnormalities based on the occurrence of real-world structural abnormalities. Then, a self-supervised task is designed to train the *HomNet* to detect pairs of homologous chromosomes with artificial structural abnormalities, enabling *HomNet* to diagnose structural anomalies using homology similarity. To mitigate the problem of inconsistent data distribution across different hospitals, we employ pre-training strategy with self-supervised tasks to pre-train *HomNet* and then fine-tune it on different hospital datasets. Fine-tuning enables the model to adapt to different data distributions. Thus, the problem of inconsistent distribution can be overcome. In addition, we collected data of 61350 chromosomes diagrams from 102 real-world patients with structural abnormalities and data of 458876 chromosomes diagrams from 1581 real-world healthy people. The data is anonymized to protect the privacy. The ample number of chromosome data from healthy people enables *HomNet* to capture common characteristics of chromosomes and develop the capability to diagnose structural abnormalities.

HomNet integrates an end-to-end pipeline with AutoVision [54], an AI analysis system for karyotypes. And it has undergone multicenter clinical trials in three Grade A tertiary hospitals in China. Overall, the contribution of this article can be summarized as follows:

- To the best of our knowledge, we are the first to study data-driven methods for identifying chromosomal structural abnormalities.

---

[1]Chromosome mutation means a change in a chromosomal segment, involving more than one gene [41].

[2]Karyotype is the overall visual representation of the entire set of chromosomes within the cells of an individual organism, including information about their sizes, quantities, and shapes [24].

---

[3]Homologous chromosomes refer to a pair of chromosomes in a biological organism's cell that share similar gene sequences and structural features. Typically, one chromosome in the pair is inherited from the father, while the other is inherited from the mother. [8]
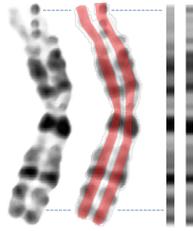
Figure 2: The chromosomal sequences are the gray mean sequences of the left and right parts of the chromosome diagram.

- Inspired by domain knowledge, we introduce *HomNet*, designed to diagnose chromosomal structural abnormalities through homologous similarity.
- We propose a self-supervised learning strategy to train *HomNet*, enabling it to capture chromosome features and address the challenge of inconsistent distribution across different hospitals.
- We conduct extensive experiments on real-world dataset involving multiple chromosomes with structural abnormalities. Our results validate the effectiveness of *HomNet* on chromosomal structural abnormality diagnosis.

2 PRELIMINARIES

To clearly illustrate our work, this section first introduces the dataset and the preprocessing steps. Detailed and formal definitions of the dataset and the problem formulation will be provided in this section.

2.1 Dataset and Preprocessing

Each chromosome data in the dataset contains two sequences sampled from the chromosome diagram. As shown in Figure 1 (a) and (b), the chromosome diagrams indicate a unidirectional flow of information, suggesting that the information density aligns with the chromosome's direction. Typically, image data exhibits information density along two perpendicular directions. Therefore, we propose converting the original data into chromosomal sequences for our dataset.

As shown in Figure 2, the chromosomal sequences represent the gray mean sequences derived from the left and right parts of the corresponding chromosome diagram. In a chromosome diagram, sister chromatids cross over the left and right sides of the chromosome. A sister chromatid refers to the identical copies (chromatids) formed by the DNA replication of a chromosome [39]. In theory, both of them would look the same in chromosome diagram. However, the left and right sides of the chromosome diagram are not completely symmetrical due to the bending of the chromosome and uneven staining of stains. Thus, we need to sample both the left and right sequences. For convenience, we use the term "chromosome" to represent the sampled chromosomal sequence.

In addition to data from diagrams, the corresponding chromosome type and band level[4] can be obtained. Human beings have

---

[4]The bands on chromosomes correspond to distinct segments of the DNA molecule, and the band level is determined by the number of bands in a set of distinct chromosomes [15].

22 pairs of autosomes (common to both men and women) and 1 pair of sex chromosomes (XX for women, XY for men), resulting in 24 types of chromosomes. Different cells may be in different cell cycles [1], which makes their chromosomes different in shape. In other words, the same chromosomes in different cells may have different band levels. We collected chromosomal data from four hospitals, denoted as Hos#1, Hos#2, Hos#3 and Hos#4 to meet privacy and data security requirements. The respective patient counts for these hospitals are 24, 27, 30, and 21. For each patient, we gathered multiple chromosomes from different cells, including both normal and abnormal chromosomes. Each patient exhibited only one type of chromosomal structural abnormality. After excluding unidentifiable chromosomes due to mutual occlusion between chromosomes and excessive bending, each hospital's dataset contained 14196, 14840, 16920, and 13496 pairs of normal homologous chromosomes, along with 428, 457, 590, and 378 pairs of homologous chromosomes with structural abnormalities. The proportion of abnormal chromosomes in each dataset was less than 4%. Additionally, we also collected chromosomal data from 1581 healthy individuals at Hos#1, resulting in a total of 458876 pairs of homologous chromosomes.

The subsequent table provides breakdown of patient samples across each hospital:

Table 1: Hospital Sample Data

Hospital	Patient	Normal sample	Abnormal sample
Hos#1	24	14196	428
Hos#2	27	14840	457
Hos#3	30	16920	590
Hos#4	21	13496	378

2.2 Problem definition

Inspired by the diagnostic process of human experts, we find that considering the information of multiple-pairs homologous chromosomes at the same time can make the judgment more reliable. This also reduces the noise disturbance of considering only one pair of homologous chromosomes. Thus we combine multi-pair homologous chromosomes to form a *bag*, denoted by X. Each homologous chromosome in X is denoted by H_i. Then we get $X = \{H_1, H_2, \ldots, H_m\}$, where m is the number of homologous pairs. All the homologous chromosomes $H_i \in X$ are from different cells of the same patient, and their chromosome type are the same, denoted by c. c is one-hot vector with the length of 24. As mentioned above, the band levels of chromosomes of cells may be different. In the dataset, band level is denoted by b, a one-hot vector with the length of 4. Because there are four types of band level in our dataset: 300, 400, 550 and 700, which also covers most of the common band types. Each H_i consists of two chromosomes, where homologous chromosome 1 and 2 are denoted by $H_{i,1} \in \mathbb{R}^{2 \times d}$ and $H_{i,2} \in \mathbb{R}^{2 \times d}$, respectively. d is the length of the chromosome sequence.

The chromosomal structural abnormalities detection task is: given a query tuple (X, c, b) containing the chromosomes bag X and the corresponding information (chromosome type c and band level b), determining whether this query tuple (X, c, b) is from chromosomes with structural abnormalities.

3 METHODOLOGY

In this section, we introduce the proposed model *HomNet* for diagnosing chromosomal structural abnormalities by homologous similarity.

3.1 Overview

The framework of *HomNet* is depicted in Figure 3, comprising three main components: *CMSBlock*, *HomBlock*, and *BagBlock*. To capture common characteristics of structural abnormalities across chromosomes with varying morphologies, we design *HomNet* based on homologous similarity. We generate five types of artificial abnormal chromosome sequences as negative samples based on real-world structural abnormalities. Subsequently, we devise a self-supervised training strategy to utilize the chromosomes of healthy individuals and the artificially generated chromosomes. To address the data inconsistent distribution problem across different hospitals, we propose fine-tuning the model for each specific hospital. For each query tuple (X, c, b), *CMSBlock* extracts the distinguishable features $R''_{i,j}$ from each chromosome in X, where $j \in 1, 2$. Subsequently, *HomBlock* aligns each pair of homologous chromosomes and captures the differences between them based on homologous similarity, denoted as H'_i. Finally, *BagBlock* considers all the H'_i to diagnose the structural abnormalities.

More specifically, *CMSBlock* first captures the features of the chromosomes. During this process, *CMSBlock* utilizes prior information band level b and chromosome type c to help *HomNet* to extract the features. Generally, there is at most one chromosome with structural abnormality in any pair of homologous chromosomes. Thus, we propose to detect the chromosomal structural abnormalities by homologous similarity. However, it is challenging to capture the differences since the shape of each chromosome varies and the diagram of chromosome pair may be flipped over, *i.e.*, they are not perfectly aligned. To align each part, *HomBlock* uses an inter-attention based strategy to capture the difference H'_i between the representations $H'_{i,1}$ and $H'_{i,2}$ of homologous chromosome 1 and 2. As mentioned above, in order to reduce noise disturbance and increase judgment reliability, *HomNet* considers multiple pairs of homologous chromosomes simultaneously in *BagBlock*. After getting all the difference representation $(H'_1, H'_2, \ldots, H'_m)$ of the Bag X, *BagBlock* makes a prediction $\hat{Y}$ according to these representations.

3.2 Self-supervised Learning with Artificial Abnormalities

Due to the significant imbalance between the number of chromosomes with structural abnormalities and those without, it becomes challenging to identify distinctive features for abnormality detection [25]. On the bright side, we possess a substantial amount of chromosome data from healthy individuals. Leveraging this data alongside artificially generated structural abnormalities, we pretrain the model. To simulate various types of structural abnormalities observed in real-world scenarios, we create five types of artificial chromosomes with structural abnormalities, as detailed in the Appendix. Our self-supervised learning task involves classifying normal and artificial abnormal chromosome pairs to capture diverse structural abnormality manifestations. To prevent *HomNet*

from solely relying on the presence of artificial data when assessing abnormal chromosome pairs, we also incorporate chromosome pairs from various types as abnormal samples. Given the length discrepancies among different chromosome types, we categorize the chromosomes into seven groups based on their length [33], ensuring each group of chromosomes is similar in length. Consequently, we exclusively form abnormal chromosome pairs using chromosomes from the same group within a cell. This approach enables our model to capture chromosome features and understand fundamental structural abnormality characteristics. Subsequently, we fine-tune the model using chromosomes from different hospitals.

3.3 Modeling the Feature of Chromosomes

Before capturing the differences between homologous chromosomes, we model features of chromosomes. Each chromosome, $H_{i,j}$, is from a chromosome diagram with or without structural abnormality. We propose *CMSBlock* to extract the features of each chromosome. As Figure 3 shows, the *CMSBlock* first merges two sequences into one, and get $H^{mg}_{i,j}$. Then, *CMSBlock* utilizes $H^{mg}_{i,j}$, band type b and chromosome type c to obtain the chromosomal representation $R''_{i,j}$.

Merge two chromosome sequences. The chromosome data we obtain consists of two sequences: $H_{i,j} \in \mathbb{R}^{2 \times d}$. As previously mentioned, these sequences represent the mean sequences of the left and right portions of the chromosome diagrams, respectively. Given the partial differences between the left and right sequences within $H_{i,j}$, both containing crucial chromosome information, a simple merging approach using average, maximum, or minimum values would result in substantial information loss. Therefore, we opt to merge these sequences using a learnable method. Specifically, we employ a convolution operation, and get the representation $H^{mg}_{i,j}$.

$$H^{mg}_{i,j} = Conv(H_{i,j}) \tag{1}$$

Model the feature of regions. Subsequently, *CMSBlock* extracts the features of each chromosome, obtaining the feature representation $R''_{i,j}$. As depicted in Figure 2, the chromosome diagram displays multiple grayscale bands along the chromosomes, which are referred to as regions [33]. When a structural abnormality occurs, the region where the abnormality happens will correspondingly change. For instance, in the case of a chromosomal deletion [33], the corresponding region will also be missing. Therefore, *HomNet* detects structural abnormalities by analyzing the differences in regions between homologous chromosomes. Therefore, we divide $H^{mg}_{i,j}$ into multiple subsequences, each representing a region of the chromosome. Additionally, leveraging the prior knowledge that chromosome types and band levels can aid in modeling the chromosomes, we fuse the feature of chromosome type c and band level b_i into each subsequence.

$$R_{i,j} = Div(W_{info}(c \oplus b_i)) + Div(H^{mg}_{i,j}) \tag{2}$$

where $\oplus$ denotes the concatenation operation and Div represents the segmentation operation. For $H^{mg}_{i,j}$, the Div operation is carried out as per Eq 1, using convolution with a kernel size of $(2, k^{mg})$, a stride of $(2, k^{mg})$, and with l_r output channels. Consequently, the output after the convolution produces a matrix where each row,
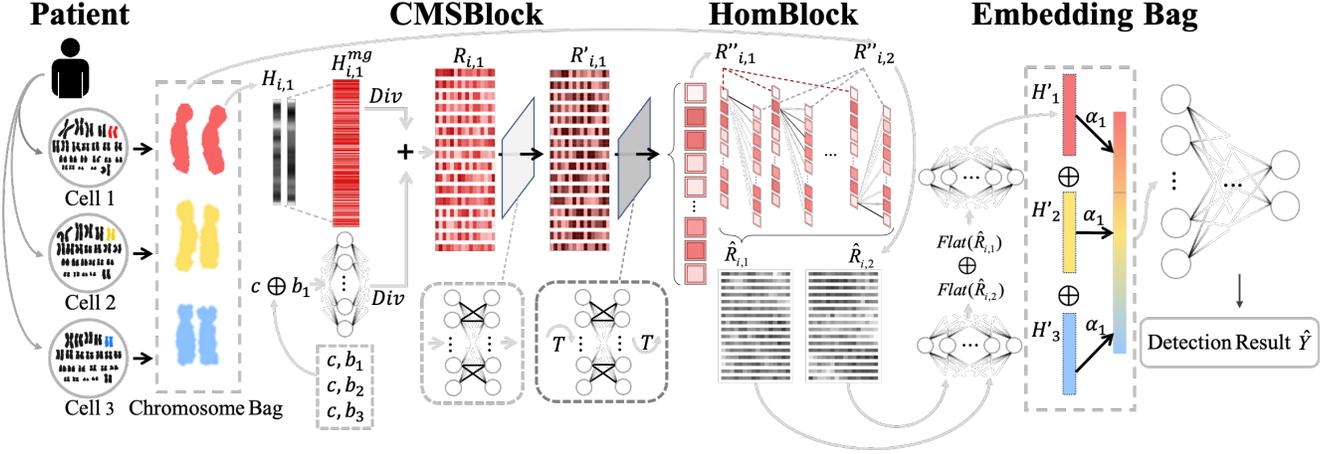
Figure 3: The framework of *HomNet*. The arrow heads refer to the dataflow of a pair of homologous chromosomes.

having a length of l_r, corresponds to a region of length k^{mg}. As for $W_{info}(c \oplus b_i)$, the *Div* operation is implemented through reshaping, where W_{info} denotes a learnable parameter matrix. Thus, the matrix of regions representation $R_{i,j} \in \mathbb{R}^{n_r \times l_r}$, where n_r represents the number of regions depending on k^{mg} and the length of $H_{i,j}$.

For each region representation in $R_{i,j}$, we use the same method to extract similar texture features, which is also called weight sharing:

$$R'_{i,j} = \sigma(R_{i,j}W_{R1})W_{R2} + R_{i,j} \qquad (3)$$

where W_{R1} and W_{R2} are learnable parameter matrices. Eq 3 extracts the features of each chromosome region, also referred to as local features.

However, there are also some types of structural abnormalities where a region of a chromosome can be found in a similar region of another normal chromosome. For example, in chromosome duplication abnormality [32], the duplicated region originates from a part of the normal chromosomes. Consequently, *HomNet* may struggle to capture the differences between homologous chromosomes, as their corresponding regions are similar. Despite this, a particular structural abnormality, the Robertsonian translocation, indicates that an entire chromosome has attached to another [56]. This results in data with this structural abnormality exhibiting more texture features due to the addition of another chromosome. Thus, capturing the global features of the chromosome will help the model better detect such abnormalities. By transposing $R'_{i,j}$, *CMSBlock* can capture the cross-region features, also referred to as global features.

$$R''_{i,j} = (\sigma(R'_{i,j}{}^T W_{R3})W_{R4})^T + R'_{i,j} \qquad (4)$$

After feature extraction, we obtain the segmented features representation of chromosomes, $R''_{i,j}$.

3.4 Capture the Differences between Homologous Chromosomes

Based on expert detection processes, *HomNet* diagnoses structural abnormalities through homologous similarity. As depicted in Figure 1, diagrams of homologous chromosome pairs with structural

abnormalities differ, while those of normal homologous chromosomes are similar. However, most existing anomaly detection methods concentrate on individual objects and do not leverage the distinctions between homologous chromosomes with structural abnormalities. Therefore, we introduce *HomBlock* to capture the differences between homologous chromosomes, based on their feature representations, $R''_{i,1}$ and $R''_{i,2}$.

Adaptive region alignment in homologous pairs. As mentioned above, obtaining the difference of homologous chromosomes involves comparing the distinctions between the feature representations of corresponding regions in a pair of homologous chromosomes. However, due to the free oscillation of chromosomes in cells, and potential inversion of chromosome diagrams, the representations $R''_{i,1}$ and $R''_{i,2}$, which encode chromosomal structural information, are not well aligned. Mere subtraction or stacking of these representations without proper alignment introduces bias. Therefore, *HomBlock* employs an adaptive alignment method to capture homologous differences by aligning each region between homologous chromosomes. *HomBlock* adaptively aligns each row of $R''_{i,1}$ with each row of $R''_{i,2}$. Here, $r^e_{i,j}$ represents the e_{th} row of $R''_{i,j}$ and also the e_{th} region in the chromosome.

HomBlock initially computes the similarity between corresponding regions. For instance, considering the e_{th} region of homologous chromosome 1, $r^e_{i,1}$, we calculate the similarity between all regions of homologous chromosome 2. Subsequently, we utilize the similarity score as the alignment coefficient, allowing $r^e_{i,1}$ to focus more on regions with high similarity. This process can be achieved through an attention mechanism.

$$a^{e,z}_{i,1} = (W_q r^e_{i,1}) \cdot (W_k r^z_{i,2})/\sqrt{d_a} \qquad (5)$$

Capture the differences in each region. Eq 5 denotes the similarity between the e_{th} region of homologous chromosome 1 and the z_{th} region of homologous chromosome 2. Here, d_a denotes the dimension of $(W_q r^e_{i,1})$ and $(W_k r^z_{i,2})$. Once the similarity score is obtained, *HomBlock* adaptively aligns the regions based on the similarity score. This alignment process allows the capture of differences between regions, providing crucial information for *HomNet*

to detect structural abnormalities.

$$\tilde{r}_{i,1}^{e} = \frac{\sum_{z=1}^{n_r} a_{i,1}^{e,z}(W_v r_{i,2}^{z})}{\sum_{z=1}^{n_r} a_{i,1}^{e,z}} \qquad (6)$$

$\tilde{r}_{i,1}^{e}$ represents the aligned result of the e_{th} region in homologous chromosome 1. Subsequently, *HomBlock* captures the difference between $r_{i,1}^{e}$ and the aligned result $\tilde{r}_{i,1}^{e}$. In our approach, we employ a multi-headed attention mechanism for alignment Specifically, $\tilde{r}_{i,1}^{e,1}$ and $\tilde{r}_{i,1}^{e,2}$ denote the alignment results from the first and second attention heads, respectively. Thus, the differences between extracted regions can be expressed as:

$$\hat{r}_{i,1}^{e} = W_{diff}(r_{i,1}^{e} + W_{head}(\tilde{r}_{i,1}^{e,1} \oplus \tilde{r}_{i,1}^{e,2} \oplus \cdots \oplus \tilde{r}_{i,1}^{e,n_h})) \qquad (7)$$

where n_h is the number of heads. $\hat{r}_{i,1}^{e}$ represents the difference feature representation of the e_{th} region. The collection of all obtained $\hat{r}_{i,j}^{e}$ forms the difference matrices $\hat{R}_{i,j}$. Consequently, we integrate these difference features of the regions to obtain the difference representation H_i' of the homologous chromosome pair:

$$H_i' = \sigma(W_{hom}(Flat(\hat{R}_{i,1}) \oplus Flat(\hat{R}_{i,2}))) \qquad (8)$$

where $Flat()$ is to flatten matrix to vector and W_{hom} is a learnable parameter matrix.

Resolve data distribution inconsistency by multi-pair homologous chromosomes. As discussed in Section 2.1, the data may be susceptible to noise disturbance. Relying solely on one pair of homologous chromosomes could potentially lead to less accurate predictions. Hence, we introduced *BagBlock* to enable a comprehensive prediction by incorporating data from multiple pairs of homologous chromosomes. This approach draws inspiration from the diagnostic process of human experts. Similar to how human experts diagnose chromosomal structural abnormalities by considering multiple chromosome diagrams, our method aims to improve the reliability of diagnostic results. By simultaneously integrating information from multiple pairs of homologous chromosomes, *HomNet* not only enhances the reliability of detection results but also mitigates the impact of noise disturbance. Structural abnormalities in some chromosome diagrams are obvious, while others may be obscure. This is why experts need to combine chromosome information from multiple cells to make a diagnosis. In this case, the contribution of the difference representation H_i^{prime} to the identification of the structural abnormality varies. Simply adding or concatenating them directly may not yield optimal results. Therefore, we compute an aggregate weight α_i for the abnormal representation H_i^{prime}. The abnormal representations H_i' are then summed based on the aggregate weight α_i, and *HomNet* makes predictions.

$$\alpha_i = Sigmoid(MLP(H_i')) \qquad (9)$$

$$\hat{Y} = \sigma(W_{bag}(\Sigma_{i=1}^{m} \alpha H_i')) \qquad (10)$$

where W_{bag} are learnable parameter matrices and $\hat{Y}$ is the prediction result. Ultimately, we employ cross-entropy loss as the objective function to optimize *HomNet*.

$$\mathcal{L} = L_{CE}(\hat{Y}, Y) \qquad (11)$$

where Y is the ground truth.

4 EXPERIMENT

4.1 Experiment Setting

Datasets . As mentioned in Section 2.1, we use the dataset sampled from healthy people for self-supervised learning. Additionally, we also use the artificial samples with and without structural abnormalities. After self-supervised training, we fine tune *HomNet* on each dataset from patients with chromosomal structural abnormalities from *Hos*#1, *Hos*#2, *Hos*#3 and *Hos*#4. The number of homologous chromosome pair in a bag is set as 5, that is, m is set as 5. Due to the privacy and sensitivity of chromosome data with structural abnormalities, almost all public datasets contain only normal chromosomes. Therefore, we constructed artificial structural abnormal chromosome data on the public chromosome dataset Pki-3 [42]. Experiments are also performed on this dataset for supplemental reference. More details about the dataset can be found in the Appendix.

Baseline. We compare our proposed model, *HomNet*, with several baseline models. Specifically, we initially compare our model with a classical machine learning method: Logistic Regression (LR) based on feature engineering. Additionally, we conduct experiments using commonly employed unsupervised methods for anomaly detection, including OC-SVM [51], Deep-SVDD [44], and an AutoEncoder (AE) based method [61]. Furthermore, we compare our model with supervised methods, such as MLP-Mixer [57], Vit [11], Xception [9], ResNet [19], DensNet [22], and MobileNet [21]. In addition, we also conduct an experiment with a recent contrastive learning-based method for detecting structural anomalies [5]. This method employs a Siamese architecture and allows for interchangeable backbone models, hence it is abbreviated as "Siamese".

Training. For the pre-training stage, we utilize normal samples and samples with artificial abnormalities. During this stage, we implement Adam as the optimizer, setting learning rate at 0.001 with a batch size of 512. A validation set, which does not overlap with the training set, is also used to identify the optimal pre-trained model. We then incorporate an early stopping mechanism to monitor model convergence closely. Specifically, if there's no observed enhancements in the validation set over 10 consecutive epochs, the training is discontinued. For the fine-tuning stage, we freeze the parameters of CMSBlock and the first layer of HomBlock. This strategy helps transfer learned knowledge from the pre-training stage. The Adam optimizer, with a learning rate of 0.00001, is once again used here. We also employ the same early stopping mechanism here.

Evaluation metrics and implementation details. We evaluate the prediction preformance of *HomNet* and baseline methods with F1 score and AUC-ROC score, capturing both the precision-recall balance and the model's overall ranking ability across varying decision thresholds. Besides, we use Adam [28] optimizer with a learning rate of 0.001 for self-supervised learning and that of 0.00001 for fine-tuning, and the batch size is set as 512. During the fine-tuning, we freeze the parameters of *CMSBlock* and the first difference extraction layer in *HomBlock*, namely the first attention layer. More details on the baseline and experimental setup can be found in the Appendix. The code is available at https://github.com/JuRenGithub/HomNet.

Table 2: Experimental results on four real-world hospital datasets.

Hospital Method	Hos#1		Hos#2		Hos#3		Hos#4	
	AUC	F1	AUC	F1	AUC	F1	AUC	F1
LR	72.43	19.51	75.23	20.54	81.72	23.47	75.56	19.10
OC-SVM	49.8	11.0	47.9	9.23	65.6	13.97	17.38	17.38
Deep − SVDD	46.91	10.28	59.31	12.56	64.83	14.67	51.13	13.01
AE	60.27	14.78	63.74	21.14	65.82	17.63	61.41	21.50
MLP − Mixer	90.76	50.02	92.83	56.27	85.24	39.49	91.20	41.65
Vit	84.42	24.81	88.08	35.84	83.80	34.12	89.31	27.84
ResNet	86.84	24.26	85.09	26.72	80.98	32.37	90.0	34.09
Xception	89.38	39.62	88.95	40.15	79.39	23.0	94.24	40.08
MobileNet	86.08	28.32	86.57	35.29	80.57	28.97	88.78	34.37
DensNet	87.59	38.21	87.95	25.04	81.69	35.85	91.66	41.64
Siamese	58.73	12.84	58.74	12.85	49.2	11.6	70.97	12.22
HomNet	**91.64**	**52.91**	**95.73**	**66.30**	**91.64**	**52.65**	**98.25**	**66.32**

Table 3: Experimental results on Pki-3 with artificial structural abnormalities.

Method	LR	OC-SVM	Deep-SVDD	AE	MLP-Mixer	Vit
AUC	60.82	45.72	48.51	49.73	50.22	40.51
F1	53.79	49.38	37.59	15.66	43.53	51.89
Method	ResNet	Xception	MobileNet	DensNet	Siamese	*HomNet*
AUC	51.68	48.88	70.31	50.58	74.09	**75.71**
F1	38.28	38.15	58.97	31.87	59.95	**64.04**

4.2 Performance Evaluation

We present the performance of all methods in chromosomal structural abnormalities detection across four hospital datasets in Table 2. The results for *Hos*#4 outperform those of other hospitals, while the outcomes for *Hos*#1 and *Hos*#3 are marginally less favorable. The varying performances across different hospitals highlight the inconsistency in the distribution of chromosome data, emphasizing the need to individually fine-tune the data from each hospital based on these experimental findings.

Overall, *HomNet* demonstrates superior performance across all four hospitals. On the one hand, the manual feature-based logistic regression method only exhibits limited capability in detecting chromosomal structural abnormalities. On the other hand, the traditional anomaly detection method OC-SVM performs poorly, facing challenges in detecting anomalies in high-dimensional data [62], especially in the context of the 4×512 dimension of homologous chromosomes data. Similarly, the performance of the deep anomaly detection models Deep-SVDD and the AE-based model lags behind that of logistic regression, as they primarily capture characteristics of normal chromosomes during training, without effectively leveraging the information from structural abnormalities. The Siamese method, utilizing contrastive learning, fails to yield optimal results as it does not incorporate a dedicated feature extraction backbone network tailored to chromosome characteristics. In the case of the MLP-Mixer network, which use the strategy of dividing patches resembling chromosome regions, results in suboptimal performance

across most experiments. In contrast, *HomNet* leverages the homologous similarity of chromosomes for diagnosing structural abnormalities, while also effectively modeling the chromosome regions. *HomNet* also considers information from multiple pairs of homologous chromosomes from multiple cells, thereby enabling it to obtain more reliable diagnostic results. Therefore, it performs the best in both AUC-ROC and F1 score.

The experimental results for all baseline methods and *HomNet* on the public dataset Pik-3 with artificial structural abnormalities are presented in Table 3. Once again, *HomNet* emerges as the top performer. Notably, in comparison to experiments conducted on datasets with real abnormalities, Siamese demonstrates improved performance on Pik-3 and almost rivals *HomNet*. This discrepancy may arise from a distribution shift between artificial and real structural abnormalities. This observation further emphasizes the importance of fine-tuning on datasets with real structural abnormalitis after pre-training with artificial abnormalities.

Table 4: In-deep analysis on the difference of *HomNet*. The "w/o Pair" row indicates that only one chromosome is used instead of homologous chromosomes pair, but the bag strategy using chromosome data from multiple cells is still used. The "w/o Bag" row indicates that only one pair of homologous chromosomes is used. "align→MLP" and "align→CNN" rows indicate that we replace the alignment part of *HomNet* with MLP and convolution neural networks. The last five rows changes the #merge range in the *CMSBlock* with difference length.

Hospital Variant	Hos#1		Hos#2		Hos#3		Hos#4	
	AUC	F1	AUC	F1	AUC	F1	AUC	F1
w/o Pair	78.13	17.45	79.42	16.42	75.31	19.48	91.63	16.51
align→MLP	87.82	46.98	86.19	39.20	71.40	28.17	97.32	58.65
align→CNN	78.12	17.42	79.43	16.38	75.33	19.82	91.65	16.45
w/o Bag	88.91	42.05	88.13	45.28	88.23	47.45	93.91	46.30
range#8	90.41	52.72	94.03	62.81	91.52	51.91	98.19	**66.76**
range#16	91.43	46.12	92.34	58.85	91.12	50.78	97.04	58.00
range#32	**91.64**	**52.91**	**95.73**	**66.30**	**91.64**	**52.65**	**98.25**	66.32
range#48	58.32	13.56	67.63	17.82	60.51	13.73	74.04	15.07
range#64	50.61	11.03	54.64	12.06	59.60	12.13	49.62	9.45

4.3 Ablation Study

HomNet utilizes the homologous similarity between aligned homologous chromosomes (Pair strategy) to diagnose structural abnormalities. Additionally, *HomNet* considers multiple pairs of homologous chromosome information (Bag strategy) to improve prediction reliability and reduce noise disturbance, rather than just a single pair of homologous chromosomes. To investigate the impact of the Pair strategy, Bag strategy and other part of *HomNet* on the performance, we conduct ablation studies in this section.

Effect of homologous similarity of homologous chromosomes (Pair strategy) . Inspired by domain knowledge indicating that homologous chromosomes with structural abnormalities differ from each other, we designed *HomNet* to diagnose structural abnormalities by comparing the differences between a pair of homologous chromosomes. To investigate whether comparing the differences based on homologous similarity is effective, we detect structural

abnormalities using data from just a single chromosome in a homologous pair while retaining the Bag strategy, which involves considering multiple chromosomes simultaneously. According to the results shown in Table 4, without comparing homologous chromosomes, the model's performance decreased the most. Therefore, comparing homologous chromosomes based on similarity is the most crucial aspect of the model for detecting structural abnormalities.

Effect of alignment . According to Table 2, the performance of methods based on supervised learning is considerably worse than that of *HomNet*. One of the main reasons is that they do not align homologous chromosomes but directly extract features and then make judgments. *HomNet* extracts the difference between regions of homologous chromosomes using an adaptive alignment based on attention mechanism. We compare the model with replacements for attention using MLP and Convolutional Neural Networks (CNNs). According to the result shown in Table 4 , in the absence of an aligned model, the performance degradation is large and MLP outperforms CNNs.

Effect of multiple pairs (Bag strategy) of homologous chromosomes . When human experts diagnose chromosomes with structural abnormalities, they observe information from multiple pairs of homologous chromosomes to make more reliable judgments. The proposed model *HomNet* also takes advantage of this idea by considering information from multiple pairs of homologous chromosomes simultaneously. To intuitively demonstrate the effectiveness of multiple pairs, we use information from only one pair of homologous chromosomes. According to the result shown in Table 4 , it is important to utilize information from multiple pairs of homologous chromosomes simultaneously.

The appropriate #merge range . As mentioned in Section 3, we divide $H_{i,j}^{mg}$ into multiple subsequences, each representing a region of the chromosome with a length of #merge range. In order to explore the most appropriate #merge range, we conduct experiments with various ranges . In *HomNet*, we set the #merge range to 32 . According to the result shown in Table 4 , When #merge range is too small, such as 8 or 16, it may not be sufficient to cover a region, resulting in a slight degradation in *HomNet*'s performance. For *Hos*#4, the F1 score of the *HomNet* improves by 0.44% when #merge range is set as 8. However, it should be noted that the #merge range is one quarter of the original *HomNet*, so the number of subsequences will increase fourfold, requiring more training time and GPU memory space. When #merge range is set too large, such as 48 or 64, they may cover more than one region, causing information loss or compression, and thus leading to a decrease in performance. Therefore, when #merge range is too large, the performance decreases the most. Overall, an #merge range of 32 for the *HomNet* is the most appropriate setting.

4.4 Running Time Comparison

We evaluated the running time of HomNet in comparison with other models by testing the completion of 100 batches and measuring the convergence time on the pre-training dataset. The following results, presented in seconds, are obtained:

While HomNet is not the fastest, it requires a comparatively lesser amount of time. And it is more time-efficient than Siamese,

Method	100 batches	Convergence
LR	0.08	3543
OC-SVM	41.16	23812
Deep-SVDD	11.72	5067
AE	16.51	11398
MLP-Mixer	1.62	1549
Vit	2.45	3832
Xception	9.33	9772
MobileNet	1.80	5948
DenseNet	13.49	10998
Siamese	18.18	10229
HomNet	5.39	2468

Table 5: Running Time Comparison

another model for diagnosing chromosomal structural abnormalities. For clinical diagnostic applications (as presented in Section 5), HomNet is capable of diagnosing a patient in under 5 milliseconds, which comfortably meets the basic speed requirements.
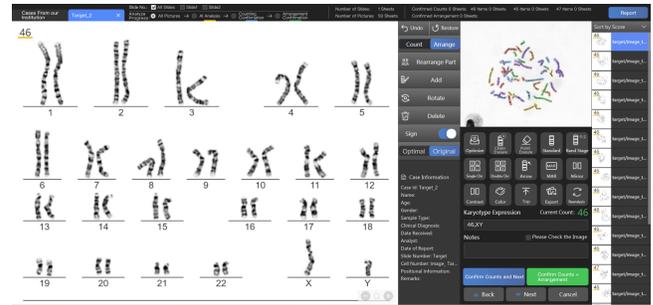
5 APPLICATION

Figure 4: AutoVision, equipped with *HomNet*, is an advanced intelligent chromosome karyotype analysis system, which registered in NMPA (National Medical Products Administration)[5] and obtained CE mark[6].

For practical clinical diagnostic applications, *HomNet* integrates an end-to-end pipeline with AutoVision [54], an AI analysis system for karyotypes developed by Hangzhou Diagens Biotechnology Co., Ltd., China. To begin with, physicians upload their patients' karyotypes to AutoVision, which has been enhanced with the integration of HomNet. Subsequently, the enhanced AutoVision, now incorporating HomNet, optimizes and organizes the karyotypes, yielding chromosome diagrams and band level identifications within an impressive 3 seconds, as depicted in Figure Figure 4. The chromosome diagrams are then converted into sequence data, as detailed in Section 2.1. This sequence data, along with the chromosome type and band levels, serves as input for *HomNet*, which diagnoses the presence of structural abnormalities. The final diagnosis results are

---

[6]NMPA is a national bureau responsible for drug supervision under the State Council of China.
[6]The CE mark on a product indicates that the manufacturer or importer of that product affirms its compliance with the relevant EU legislation.

directly output to the physicians. Notably, *HomNet* can diagnose chromosomal structural abnormalities using data from a single patient, encompassing 115 pairs of homologous chromosomes from five cells, in less than 5 milliseconds.

AutoVision, integrated with *HomNet*, has undergone multicenter clinical trials in three Grade A tertiary hospitals in East, Central, and West China. Presently, 1518 cases have been enrolled, with 499 sample analyses completed. Preliminary results demonstrate that AutoVision exhibits 75% sensitivity[7], 99% specificity[8], and a diagnostic accuracy of 98.4% in automated analysis, showcasing its precision and efficiency in clinical applications. The product was honored with the first prize in technological achievements by the China Birth Defects Intervention Assistance Fund.

6 RELATED WORK

For clarity, we first review methods related to anomaly detection, then introduce works about detecting chromosomes with structural abnormalities.

Anomaly detection is an important topic in diverse research fields and application domains. Anomalies are also referred to as outliers , deviants or abnormalities [2]. Chromosomes with structural abnormalities that are different from normal chromosomes can also be called anomalies. Traditional anomaly detection methods include Principal Component Analysis (PCA) based methods [20][53], Nearest Neighbor (NN) based methods [29][6], and the One-Class Support Vector Machine (OC-SVM) [51] that learns a decision function for novelty detection, that is, classifying new data as similar or different one to the training set. These methods are unsupervised. Because in the anomaly detection setting, even if abnormalities are available, it is usually insufficient to make a supervised approach ffective [43].

As deep learning [30][50] has made substantial progress in dealing with many machine learning problems, some deep learning-based anomaly detection methods have also been proposed. A substantial amount of deep learning methods on anomaly detection choose to use autoencoder (AE) and its variants [40][37]. Deep Support Vector Data Description (Deep-SVDD) [44], a variant of one-class neural network methods, trains a neural network and minimizes the volume of a hypersphere that maps the normal instances closely to the center of sphere. Besides, there are some other popular deep one-class classification based approaches [12][46][35, 45]. Method based on one deep generative models [49][48][59] and self-supervised methods [16][55] are also applied in anomaly detection.

In recent years, some anomaly detection methods have been applied in the medical field [47][7][27]. These methods employed state-of-the-state techniques in anomaly detection and improved performance compared to traditional methods. But only a few works focus on chromosomal structural abnormality detection. [5] proposes to use the method of contrastive learning to detect chromosomal structural abnormalities using Siamese network. Most of these methods do not take advantage of the homologous similarity between homologous chromosomes.

7 CONCLUSION

In this paper, we study the problem of chromosomal structural abnormalities diagnosis. Inspired by the homologous similarity of chromosomes and the diagnostic process of doctors, we propose a novel model, *HomNet*, to model features of chromosomes and diagnose chromosomal structural abnormalities by homologous similarity. *HomNet* simultaneously considers the information of multiple pairs of homologous chromosomes and detects structural abnormalities by capturing the differences between homologous chromosomes. Additionally, we artificially construct chromosomal samples with structural abnormalities according to different types of real-world structural abnormalities. And self-supervised learning has been adopted to enable *HomNet* to fully learn chromosomes with data of normal chromosomes and artificial structural abnormalities. Then, we fine-tune *HomNet* with data from different hospital to address the challenge of inconsistent distribution. Experimental results show that *HomNet* achieved significantly better performance than other baselines on real-world dataset and public datset with artificial abnormalities. In addition, ablation studies demonstrate the importance of each part of *HomNet*.

---

[7]Sensitivity (true positive rate) is the probability of a positive test result, conditioned on the individual truly being positive.

[8]Specificity (true negative rate) is the probability of a negative test result, conditioned on the individual truly being negative.

ACKNOWLEDGMENTS

REFERENCES

[1] Michael Abercrombie, Clarence James Hickman, and Minnie L Johnson. 2017. *A dictionary of biology*. Routledge.

[2] Charu C Aggarwal. 2017. An introduction to outlier analysis. In *Outlier analysis*. Springer, 1–34.

[3] Genetic Alliance, New York-Mid-Atlantic Consortium for Genetic, Newborn Screening Services, et al. 2009. Understanding genetics: a New York, mid-Atlantic guide for patients and health professionals. (2009).

[4] Nilesh K Banavali. 2013. Partial base flipping is sufficient for strand slippage near DNA duplex termini. *Journal of the American Chemical Society* 135, 22 (2013), 8274–8282.

[5] Mohammed El Amine Bechar, Jean-Marie Guyader, Marwa El Bouz, Nathalie Douet-Guilbert, Ayman Al Falou, and Marie-Bérengère Troadec. 2023. Highly performing automatic detection of structural chromosomal abnormalities using siamese architecture. *Journal of Molecular Biology* 435, 8 (2023), 168045.

[6] Markus M Breunig, Hans-Peter Kriegel, Raymond T Ng, and Jörg Sander. 2000. LOF: identifying density-based local outliers. In *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*. 93–104.

[7] Chensi Cao, Feng Liu, Hai Tan, Deshou Song, Wenjie Shu, Weizhong Li, Yiming Zhou, Xiaochen Bo, and Zhi Xie. 2018. Deep learning and its applications in biomedicine. *Genomics, proteomics & bioinformatics* 16, 1 (2018), 17–32.

[8] Sean B Carroll, John Doebley, Anthony JF Griffiths, and Susan R Wessler. 2015. *Introduction to genetic analysis*. Springer.

[9] François Chollet. 2017. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1251–1258.

[10] Queremel Milani DA and Tadi P. 2023. Genetics, Chromosome Abnormalities. [Updated 2023 Apr 24]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK557691/.

[11] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929* (2020).

[12] Sarah M Erfani, Sutharshan Rajasegarar, Shanika Karunasekera, and Christopher Leckie. 2016. High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning. *Pattern Recognition* 58 (2016), 121–134.

[13] RJ MKinlay Gardner, Grant R Sutherland, and Lisa G Shaffer. 2012. *Chromosome abnormalities and genetic counseling*. Number 61. OUP USA.

[14] Stanley M Gartler. 2006. The chromosome number in humans: a brief history. *Nature Reviews Genetics* 7, 8 (2006), 655–660.

[15] Katherine B Geiersbach, Anna E Gardner, Andrew Wilson, Shashirekha Shetty, Hélène Bruyère, James Zabawski, Debra F Saxe, Rebecca Gaulin, Cynthia Williamson, and Daniel L Van Dyke. 2014. Subjectivity in chromosome band–level estimation: a multicenter study. *Genetics in Medicine* 16, 2 (2014), 170–175.

[16] Izhak Golan and Ran El-Yaniv. 2018. Deep anomaly detection using geometric transformations. *Advances in neural information processing systems* 31 (2018).

[17] Leland Hartwell, Michael L Goldberg, Janice A Fischer, Leroy E Hood, and Charles F Aquadro. 2018. *Genetics: from genes to genomes*. McGraw-Hill Education.

[18] T Hassold, N Chen, J Funkhouser, T Jooss, B Manuel, J Matsuura, A Matsuyama, C Wilson, JA Yamane, and PA Jacobs. 1980. A cytogenetic study of 1000 spontaneous abortions. *Annals of human genetics* 44, 2 (1980), 151–164.

[19] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.

[20] Heiko Hoffmann. 2007. Kernel PCA for novelty detection. *Pattern recognition* 40, 3 (2007), 863–874.

[21] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. 2017. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861* (2017).

[22] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. 2017. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 4700–4708.

[23] National Human Genome Research Institute. 2020. *Chromosome Abnormalities Fact Sheet*. https://www.genome.gov/about-genomics/fact-sheets/Chromosome-Abnormalities-Fact-Sheet

[24] National Human Genome Research Institute. 2024. *KARYOTYPE*. https://www.genome.gov/genetics-glossary/Karyotype

[25] Nathalie Japkowicz and Shaju Stephen. 2002. The class imbalance problem: A systematic study. *Intelligent data analysis* 6, 5 (2002), 429–449.

[26] Justin M Johnson and Taghi M Khoshgoftaar. 2019. Survey on deep learning with class imbalance. *Journal of Big Data* 6, 1 (2019), 1–54.

[27] Samir Khan and Takehisa Yairi. 2018. A review on the application of deep learning in system health management. *Mechanical Systems and Signal Processing* 107 (2018), 241–265.

[28] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[29] Edwin M Knorr, Raymond T Ng, and Vladimir Tucakov. 2000. Distance-based outliers: algorithms and applications. *The VLDB Journal* 8, 3 (2000), 237–253.

[30] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *nature* 521, 7553 (2015), 436–444.

[31] R. Lewis. 2005. Human Genetics: Concepts and Applications. *WCB/McGraw-Hill* (2005).

[32] Wen-Hsiung Li, Jing Yang, and Xun Gu. 2005. Expression divergence between duplicate genes. *TRENDS in Genetics* 21, 11 (2005), 602–607.

[33] Jean McGowan-Jordan, Ros J Hastings, and Sarah Moore. 2020. *ISCN 2020: An International System for Human Cytogenomic Nomenclature (2020). Reprint Of: Cytogenetic and Genome Research 2020, Vol. 160, No. 7-8*. Karger, S.

[34] Aubrey Milunsky and Jeff M Milunsky. 2015. *Genetic disorders and the fetus: diagnosis, prevention, and treatment*. John Wiley & Sons.

[35] Poojan Oza and Vishal M Patel. 2018. One-class convolutional neural network. *IEEE Signal Processing Letters* 26, 2 (2018), 277–281.

[36] Theophilus S Painter. 1933. A new method for the study of chromosome rearrangements and the plotting of chromosome maps. *Science* 78, 2034 (1933), 585–586.

[37] Nick Pawlowski, Matthew CH Lee, Martin Rajchl, Steven McDonagh, Enzo Ferrante, Konstantinos Kamnitsas, Sam Cooke, Susan Stevenson, Aneesh Khetani, Tom Newman, et al. 2018. Unsupervised lesion detection in brain CT using bayesian convolutional autoencoders. (2018).

[38] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.

[39] Jan-Michael Peters and Tomoko Nishiyama. 2012. Sister chromatid cohesion. *Cold Spring Harbor Perspectives in Biology* 4, 11 (2012), a011130.

[40] Emanuele Principi, Fabio Vesperini, Stefano Squartini, and Francesco Piazza. 2017. Acoustic novelty detection with adversarial autoencoders. In *2017 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 3324–3330.

[41] Rigomar Rieger, Arnd Michaelis, and Melvin M Green. 2012. *Glossary of genetics and cytogenetics: classical and molecular*. Springer Science & Business Media.

[42] G. Ritter and G. Schreib. 2000. Profile and feature extraction from chromosomes. In *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, Vol. 2. 287–290 vol.2. https://doi.org/10.1109/ICPR.2000.906069

[43] Lukas Ruff, Jacob R Kauffmann, Robert A Vandermeulen, Grégoire Montavon, Wojciech Samek, Marius Kloft, Thomas G Dietterich, and Klaus-Robert Müller. 2021. A unifying review of deep and shallow anomaly detection. *Proc. IEEE* 109, 5 (2021), 756–795.

[44] Lukas Ruff, Robert Vandermeulen, Nico Goernitz, Lucas Deecke, Shoaib Ahmed Siddiqui, Alexander Binder, Emmanuel Müller, and Marius Kloft. 2018. Deep one-class classification. In *International conference on machine learning*. PMLR, 4393–4402.

[45] Lukas Ruff, Robert A Vandermeulen, Nico Görnitz, Alexander Binder, Emmanuel Müller, Klaus-Robert Müller, and Marius Kloft. 2019. Deep semi-supervised anomaly detection. *arXiv preprint arXiv:1906.02694* (2019).

[46] Mohammad Sabokrou, Mohammad Khalooei, Mahmood Fathy, and Ehsan Adeli. 2018. Adversarially learned one-class classifier for novelty detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3379–3388.

[47] Daisuke Sato, Shouhei Hanaoka, Yukihiro Nomura, Tomomi Takenaga, Soichiro Miki, Takeharu Yoshikawa, Naoto Hayashi, and Osamu Abe. 2018. A primitive study on unsupervised anomaly detection with an autoencoder in emergency head CT volumes. In *Medical Imaging 2018: Computer-Aided Diagnosis*, Vol. 10575. SPIE, 388–393.

[48] Thomas Schlegl, Philipp Seeböck, Sebastian M Waldstein, Georg Langs, and Ursula Schmidt-Erfurth. 2019. f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks. *Medical image analysis* 54 (2019), 30–44.

[49] Thomas Schlegl, Philipp Seeböck, Sebastian M Waldstein, Ursula Schmidt-Erfurth, and Georg Langs. 2017. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In *International conference on information processing in medical imaging*. Springer, 146–157.

[50] Jürgen Schmidhuber. 2015. Deep learning in neural networks: An overview. *Neural networks* 61 (2015), 85–117.

[51] Bernhard Schölkopf, John C Platt, John Shawe-Taylor, Alex J Smola, and Robert C Williamson. 2001. Estimating the support of a high-dimensional distribution. *Neural computation* 13, 7 (2001), 1443–1471.

[52] Lisa G Shaffer and James R Lupski. 2000. Molecular mechanisms for constitutional chromosomal rearrangements in humans. *Annual review of genetics* 34, 1 (2000), 297–329.

[53] Mei-Ling Shyu, Shu-Ching Chen, Kanoksri Sarinnapakorn, and LiWu Chang. 2003. *A novel anomaly detection scheme based on principal component classifier.* Technical Report. Miami Univ Coral Gables Fl Dept of Electrical and Computer Engineering.

[54] Ning Song, Chaoyu Wu, and MA Weiqi. 2021. Chromosome recognition method based on deep learning. US Patent App. 17/272,254.

[55] Jihoon Tack, Sangwoo Mo, Jongheon Jeong, and Jinwoo Shin. 2020. Csi: Novelty detection via contrastive learning on distributionally shifted instances. *Advances in neural information processing systems* 33 (2020), 11839–11852.

[56] E Therman, B Susman, and C Denniston. 1989. The nonrandom participation of human acrocentric chromosomes in Robertsonian translocations. *Annals of human genetics* 53, 1 (1989), 49–65.

[57] Ilya Tolstikhin, Neil Houlsby, Alexander Kolesnikov, Lucas Beyer, Xiaohua Zhai, Thomas Unterthiner, Jessica Yung, Daniel Keysers, Jakob Uszkoreit, Mario Lucic, et al. 2021. Mlp-mixer: An all-mlp architecture for vision. *arXiv preprint arXiv:2105.01601* (2021).

[58] Gail H. Vance. 2020. 30 - Cytogenetics/cytogenomics. In *Rodak's Hematology (Sixth Edition)* (sixth edition ed.), Elaine M. Keohane, Catherine N. Otto, and Jeanine M. Walenga (Eds.). Elsevier, St. Louis (MO), 525–539. https://doi.org/10.1016/B978-0-323-53045-3.00039-8

[59] Houssam Zenati, Manon Romain, Chuan-Sheng Foo, Bruno Lecouat, and Vijay Chandrasekhar. 2018. Adversarially learned anomaly detection. In *2018 IEEE International conference on data mining (ICDM)*. IEEE, 727–736.

[60] Jianzhi Zhang. 2003. Evolution by gene duplication: an update. *Trends in ecology & evolution* 18, 6 (2003), 292–298.

[61] Yue Zhao, Zain Nasrullah, and Zheng Li. 2019. PyOD: A Python Toolbox for Scalable Outlier Detection. *Journal of Machine Learning Research* 20, 96 (2019), 1–7. http://jmlr.org/papers/v20/19-011.html

[62] Arthur Zimek, Erich Schubert, and Hans-Peter Kriegel. 2012. A survey on unsupervised outlier detection in high-dimensional numerical data. *Statistical Analysis and Data Mining: The ASA Data Science Journal* 5, 5 (2012), 363–387.

A APPENDIX

A.1 Artificial abnormality

According to the real world chromosomes with structural abnormalities in our datasets, we construct five types of artificial structural abnormality chromosome sequences:

- **Deleted.** Part of the sampled chromosome sequences is deleted, which makes the length of the chromosome sequence shorter than before. The corresponding types of real world chromosome structural abnormality include deletion: a portion of the chromosome is missing or has been deleted [31], and the donor chromosomes in insertion abnormalities: a portion of one chromosome has been deleted from its normal place and inserted into another chromosome [4].
- **Added with other fragments.** The sampled chromosome sequences are added with a non-self segment, which makes the length of the chromosome sequences longer than before. The corresponding types of real world chromosome structural abnormalities include the receptor chromosomes of insertion abnormality and some unbalanced translocations: a portion of one chromosome has been transferred to another chromosome [34].
- **Duplicated with self fragments.** The sampled chromosome sequences are added with a segment of its own, which also makes the chromosome sequences longer. The corresponding type of real world chromosome structural abnormality is duplication: a portion of the chromosome has been duplicated [60].
- **Replaced.** One segment of a pair of sampled chromosome sequences, the left part and right part, is replaced by another, which doesn't change the length. If it is replaced by a segment of itself, the corresponding type of real world chromosome structural abnormality is inversion: A portion of the chromosome has broken off, turned upside down, and reattached [36]. If it is replaced by a segment from another pair of sampled chromosome sequences, the corresponding type of real world chromosome structural abnormality is balanced translocations.
- **Robertsonian translocation.** The chromosome sampling sequence was symmetrical and repeated, which make the length doubled. The corresponding type of real world chromosome abnormality is robertsonian translocation: an entire chromosome has attached to another at the centromere [17].

A.2 Datasets

As mentioned in Section 2.1, we use the dataset sampled from both healthy people and patients. We split the datasets according to the desensitized patient IDs. There is no overlap between the patients in the training set and test set. In the self-supervised learning stage,

the split ratio of training set and test set is about 9: 1. And chromosome data with artificial structural abnormalities can avoid data imbalance. Therefore, the ratio of normal chromosomes bag to bag of chromosomes with artificial structural abnormalities is balanced. In the fine tuning stage, we split the dataset into training set and test set with a ratio of 1: 4. The ratio of chromosome with and without structural abnormalities is less than 4%. Each type of chromosome data for each patient corresponds to multiple query tuples (X, c, b). For each type of chromosome of each person, there are multiple chromosome diagrams from different cell. In the dataset, the average number of homologous chromosomes pair of each type of chromosome from one person is 6 for the healthy person dataset of Hos#1, because some of the chromosome photos are incomplete due to occlusion and some other reasons. And the average number is 13 for the patients datasets of Hos#1, Hos#2, Hos#3 and Hos#4. In our experiment, in order to cover as many patients as possible in the datasets, we let each bag, X, contain five pairs of homologous chromosome. In other words, m, the number of homologous pairs, is set as 5. Additionally, for the same type of chromosome data of the same patient, we randomly combined multiple bags.

A.3 Experiment setting

For feature engineering-based LR, we carefully extract features of one chromosome including mean, variance, on-peak and off-peak of the sampled sequence data, and features between the chromosomes in a homologous pair including DTW distance, the correlation coefficient and covariance. For anomaly detection methods, such as OC-SVM, Deep-SVDD and AE-based method, we use the data of normal chromosomes from healthy people as the training set. For the supervised-learning based method, we change the operation of dividing the image into multiple patches in the first step into the operation of dividing the sampled sequence into multiple subsequences or transfer the Conv2D into Conv1D to handle the sequential chromosome data. For LR and OC-SVM, We used the sk-learn [38] library for our experiments. For Deep-SVDD and AE-based method, we use the PyOD [61] libary for our experiments. The codes of MobileNet, Xception, DensNet and ResNet refer to https://github.com/weiaicunzai/pytorch-cifar100. The code of MLP-Mixer refers to https://github.com/google-research/vision_transformer. The code of Vit refers to https://github.com/lucidrains/vit-pytorch. The code of Siamese refers to https://github.com/MEABECHAR/ChromosomeSiameseAD. Besides, we use Adam [28] optimizer with a learning rate of 0.001 for self-supervised learning and that of 0.00001 for fine-tuning, and the batch size is set as 512. During the fine-tuning, we freeze the parameters of *CMSBlock* and the first difference extraction layer in *HomBlock*, namely the first attention layer.