What is claimed is:

1. A method for determining whether a fetus has a chromosomal aneuploidy, characterized by comprising:

(1) acquiring nucleic acid sequencing data from a pregnant woman sample, wherein the pregnant woman sample comprises a fetal free nucleic acid, and the nucleic acid sequencing data are composed of a plurality of sequencing reads;

(2) determining a fetal fraction of the pregnant woman sample and an estimated fraction by a predetermined chromosome based on the nucleic acid sequencing data, wherein the estimated fraction by a predetermined chromosome is determined based on a difference between a number of sequencing reads of the predetermined chromosome and a number of sequencing reads of a first comparison chromosome, the predetermined chromosome comprises a chromosome to be tested and a second comparison chromosome, and the first comparison chromosome comprises at least one autosome different from the predetermined chromosome;

(3) determining a first feature based on a difference between the estimated fraction by the chromosome to be tested and the estimated fraction by the second comparison chromosome, and determining a second feature based on a difference between the estimated fraction by the chromosome to be tested and the fetal fraction; and

(4) determining whether the fetus has an aneuploidy for the chromosome to be tested based on the first feature and the second feature by using corresponding data of a control sample, wherein the control sample comprises a positive sample and a negative sample, the positive sample has an aneuploidy for the chromosome to be tested, and the negative sample does not have an aneuploidy for the chromosome to be tested;

and the method is further characterized by the following (a) to (d):

(a) the first comparison chromosome, the second comparison chromosome and the chromosome to be tested are derived from the same sample;

(b) the number of sequencing reads of the first comparison chromosome is an average number of sequencing reads of a plurality of autosomes, and the plurality of autosomes comprise at least one autosome that is known to have no aneuploidy, and the estimated fraction is determined according to the following formula:

$$F_j = 2 \times \left|R_j - R_r\right| / R_r$$

wherein,

j represents the serial number of a chromosome the estimated fraction of which needs to be determined,

F_j represents the estimated fraction by the chromosome j,

R_r represents the average number of sequencing reads of the plurality of autosomes;

(c) the first feature is determined by the following formula:

$$X_1 = F_i - F_r$$

wherein,

X_1 represents the first feature,

i represents the serial number of the chromosome to be tested,

F_i represents the estimated fraction by the chromosome to be tested,

F_r represents the average value of the estimated fraction by the second comparison chromosome;

(d) the second feature is determined by the following formula:

$$X_2 = \frac{F_i}{F_a}$$

wherein,

X_2 represents the second feature,

i represents the serial number of the chromosome to be tested,

F_i represents the estimated fraction by the chromosome to be tested,

F_a represents the fetal fraction.

2. The method according to claim 1, characterized in that the pregnant woman sample comprises peripheral blood of a pregnant woman.

3. The method according to claim 1, characterized in that the nucleic acid sequencing data are obtained by paired-end sequencing, single-end sequencing, or single-molecule sequencing.

4. The method according to claim 1, characterized in that the fetal fraction is determined by the following steps:

(a) comparing the nucleic acid sequencing data from the pregnant woman sample with a reference sequence, so as to determine the number of sequencing reads that fall into a predetermined window; and

(b) determining the fetal fraction of the pregnant woman sample based on the number of sequencing reads that fall into the predetermined window.

5. The method according to claim 1, characterized in that in the step (2), the number of sequencing reads of the first comparison chromosome is an average number of sequencing reads of at least 15 autosomes,

optionally, the number of sequencing reads of the first comparison chromosome is an average number of sequencing reads of at least 20 autosomes,

optionally, the number of sequencing reads of the first comparison chromosome is an average number of sequencing reads of all autosomes.

6. The method according to claim 1, characterized in that in the step (2), the second comparison chromosome comprises a plurality of autosomes having no aneuploidy, and in the step (3), the first feature is determined based on a difference between the estimated fraction by the chromosome to be tested and an average value of the estimated fraction by the second comparison chromosome.

7. The method according to claim 6, characterized in that the second comparison chromosome comprises at least 10 autosomes.

8. The method according to claim 6, characterized in that the second comparison chromosome comprises 15 autosomes.

11. The method according to claim 8, characterized by further comprising:

determining the estimated fractions by a plurality of autosomes; and

selecting target autosomes from the sorted autosomes as the second comparison chromosome in a priority order from small to large.

10. The method according to any one of claims 1 to 9, characterized in that, before performing the step (4), the first feature and the second feature are standardized, so that the absolute values of the first feature and the second feature are independently between 0 to 1.

11. The method according to claim 1, characterized in that, in the step (4), the numbers of the positive sample and the negative sample have a ratio of not less than 1:4.

12. The method according to claim 1, characterized in that in the step (4), the numbers of the positive sample and the negative sample have a ratio of not exceeding 4:1.

13. The method according to claim 1, characterized in that in the step (4), the numbers of the positive sample and the negative sample have a ratio of 1:0.1 to 1:5.

14. The method according to claim 1, characterized in that in the step (4), the numbers of the

positive sample and the negative sample have a ratio of 1:0.25 to 1:4.

15. The method according to claim 1, characterized in that neither the positive sample nor the negative sample has aneuploidy for chromosomes other than the chromosome to be tested.

16. The method according to claim 1, characterized in that in the step (4), a two-dimensional feature vector of the pregnant woman sample and the control samples is determined based on the first feature and the second feature, a distance between samples is determined based on the two-dimensional feature vector, and the pregnant woman sample is classified as positive sample or negative sample, so as to determine whether the fetus has an aneuploidy for the chromosome to be tested.

17. The method according to claim 16, characterized in that the distance is an Euclidean distance, a Manhattan distance or a Chebyshev distance.

18. The method according to claim 16, characterized in that the step (4) further comprises:

(4-1) calculating distances between the pregnant woman sample and the control samples respectively;

(4-2) sorting the obtained distances, the sorting being based on the order from small to large;

(4-3) selecting a predetermined number of control samples in order from small to large based on the sorting;

(4-4) determining the number of positive samples and the number of negative samples in the predetermined number of control samples respectively;

(4-5) determining a classification result of the pregnant woman sample based on a majority decision method.

19. The method according to claim 18, characterized in that the predetermined number does not exceed 20.

20. The method according to claim 18, characterized in that the predetermined number is 3 to 10.

21. The method according to claim 18, characterized in that, in the step (4-2), the distances between the sample to be tested and the predetermined control samples are weighted in advance before the sorting is performed.

22. A device for determining whether a fetus has a chromosomal aneuploidy and for use in the method according to any one of claims 1 to 21, characterized by comprising:

a data acquisition module, which is configured to acquire nucleic acid sequencing data from a

pregnant woman sample, wherein the pregnant woman sample comprises a fetal free nucleic acid, and the nucleic acid sequencing data are composed of a plurality of sequencing reads;

a fetal fraction- estimated fraction determination module, which is configured to determine a fetal fraction of the pregnant woman sample and an estimated fraction by a predetermined chromosome based on the nucleic acid sequencing data, wherein the estimated fraction by a predetermined chromosome is determined based on a difference between a number of sequencing reads of the predetermined chromosome and a number of sequencing reads of a first comparison chromosome, the predetermined chromosome comprises a chromosome to be tested and a second comparison chromosome, and the first comparison chromosome comprises at least one autosome different from the predetermined chromosome;

a feature determination module, which is configured to determine a first feature based on a difference between the estimated fraction by the chromosome to be tested and the estimated fraction by the second comparison chromosome, and to determine a second feature based on a difference between the estimated fraction by the chromosome to be tested and the fetal fraction; and

an aneuploidy determination module, which is configured to determine whether the fetus of the pregnant woman has an aneuploidy for the chromosome to be tested based on the first feature and the second feature by using corresponding data of control sample, wherein the control sample comprises a positive sample and a negative sample, the positive sample has an aneuploidy for the chromosome to be tested, and the negative sample does not have an aneuploidy for the chromosome to be tested;

and the device is further characterized by the following (a) to (d):

(a) the first comparison chromosome, the second comparison chromosome and the chromosome to be tested are derived from the same sample;

(b) the fetal fraction- estimated fraction determination module comprises:

an estimated fraction calculation unit, which is configured to determine the estimated fraction according to the following formula:

$$F_j = 2 \times \left|R_j - R_r\right| / R_r$$

wherein,

j represents the serial number of a chromosome the estimated fraction of which needs to be determined,

F_j represents the estimated fraction by the chromosome j,

R_r represents the average number of sequencing reads of the plurality of autosomes, and

R_j represents the number of sequencing reads of the chromosome j;

(c) the feature determination module comprises:

a first feature determination unit, which is configured to determine the first feature by the following formula:

$$X_1 = F_i - F_r$$

wherein,

X_1 represents the first feature,

i represents the serial number of the chromosome to be tested,

F_i represents the estimated fraction by the chromosome to be tested,

F_r represents the average value of the estimated fraction by the second comparison chromosome;

(d) the feature determination module comprises:

a second feature determination unit, which is configured to determine the second feature by the following formula:

$$X_2 = \frac{F_i}{F_a}$$

wherein,

X_2 represents the second feature,

i represents the serial number of the chromosome to be tested,

F_i represents the estimated fraction by the chromosome to be tested,

F_a represents the fetal fraction.

23. The device according to claim 22, characterized in that the fetal fraction- estimated fraction determination module comprises:

an alignment unit, which is configured to align the nucleic acid sequencing data from the pregnant woman sample with a reference sequence, so as to determine the number of sequencing reads that fall into a predetermined window; and

a fetal fraction calculation unit, which is configured to determine the fetal fraction of the pregnant woman sample based on the number of sequencing reads that fall into the predetermined

window.

24. The device according to claim 22, characterized in that the fetal fraction- estimated fraction determination module comprises:

a second comparison chromosome determination unit, which is configured to sort the estimated fractions by a plurality of autosomes in a priority order from small to large, and select target autosomes from the sorted autosomes as the second comparison chromosome.

25. The device according to claim 22, characterized in that the feature determination module comprises:

a standardization processing unit, which is configured to perform standardization processing on the first feature and the second feature, so that the absolute values of the first feature and the second feature are independently between 0 to 1.

26. The device according to claim 22, characterized in that the aneuploidy determination module is configured to determine a two-dimensional feature vector of the pregnant woman sample and the control samples, to determine a distance between samples based on the two-dimensional feature vector, and to classify the pregnant woman sample as positive sample or negative sample, so as to determine whether the fetus has an aneuploidy for the chromosome to be tested.

27. The device according to claim 26, characterized in that the distance is an Euclidean distance, a Manhattan distance or a Chebyshev distance.

28. The device according to claim 22, characterized in that the aneuploidy determination module is configured to determine a classification result of the pregnant woman sample by using a *k*-nearest neighbor model.

29. The device according to claim 28, characterized in that the *k*-nearest neighbor model adopts a *k* value of not exceeding 20.

30. The device according to claim 28, characterized in that the *k*-nearest neighbor model adopts a *k* value of 3 to 10.

31. The device according to claim 28, characterized in that in the *k*-nearest neighbor model, the distance between samples is weighted.

32. A computer-readable storage medium for use in the method according to any one of claims 1 to 21, on which a computer program is stored, characterized in that, when the program is executed by a processor, the steps of the method according to any one of claims 1-25 are implemented.

33. An electronic device for use in the method according to any one of claims 1 to 21, characterized by comprising:

the computer-readable storage medium according to claim 32; and

one or more processors, which are configured to execute the program stored on the computer-readable storage medium.

34. A method for constructing a machine learning classification model, characterized by comprising:

(a) performing the following steps for each of a plurality of pregnant women samples:

acquiring nucleic acid sequencing data from the pregnant woman sample, wherein the pregnant woman sample comprise a fetal free nucleic acid, the nucleic acid sequencing data are composed of a plurality of sequencing reads, the pregnant woman sample comprises at least one positive sample and at least one negative sample, the positive sample has an aneuploidy for a chromosome to be tested, and the negative sample does not have an aneuploidy for the chromosome to be tested;

determining a fetal fraction of the pregnant woman sample and an estimated fraction by a predetermined chromosome based on the nucleic acid sequencing data, wherein the estimated fraction by a predetermined chromosome is determined based on a difference between a number of sequencing reads of the predetermined chromosome and a number of sequencing reads of a first comparison chromosome, the predetermined chromosome comprises a chromosome to be tested and a second comparison chromosome, and the first comparison chromosome comprises at least one autosome different from the predetermined chromosome; and

determining a first feature based on a difference between the estimated fraction by the chromosome to be tested and the estimated fraction by the second comparison chromosome, and determining a second feature based on a difference between the estimated fraction by the chromosome to be tested and the fetal fraction;

(b) performing a machine learning training by taking the plurality of pregnant women samples as samples and using the first features and the second features of the samples, so as to construct a machine learning classification model for determining whether the fetus has an aneuploidy;

and the method is further characterized by the following (a) to (d):

(a) the first comparison chromosome, the second comparison chromosome and the chromosome to be tested are derived from the same sample;

(b) the number of sequencing reads of the first comparison chromosome is an average number

of sequencing reads of a plurality of autosomes, and the plurality of autosomes comprise at least one autosome that is known to have no aneuploidy, and the estimated fraction is determined according to the following formula:

$$F_j = 2 \times \left|R_j - R_r\right| / R_r$$

wherein,

j represents the serial number of a chromosome the estimated fraction of which needs to be determined,

F_j represents the estimated fraction by the chromosome j,

R_r represents the average number of sequencing reads of the plurality of autosomes;

(c) the first feature is determined by the following formula:

$$X_1 = F_i - F_r$$

wherein,

X_1 represents the first feature,

i represents the serial number of the chromosome to be tested,

F_i represents the estimated fraction by the chromosome to be tested,

F_r represents the average value of the estimated fraction by the second comparison chromosome;

(d) the second feature is determined by the following formula:

$$X_2 = \frac{F_i}{F_a}$$

wherein,

X_2 represents the second feature,

i represents the serial number of the chromosome to be tested,

F_i represents the estimated fraction by the chromosome to be tested,

F_a represents the fetal fraction.

35. The method according to claim 34, characterized in that the machine learning classification model is a KNN model.

36. The method according to claim 35, characterized in that the KNN model adopts a Euclidean distance.

37. A device for constructing a machine learning classification model and for sue in the method according to any one of claims 34 to 36, characterized by comprising:

a feature acquisition module, which is configured to perform the following steps for each of a plurality of pregnant women samples:

acquiring nucleic acid sequencing data from the pregnant woman sample, wherein the pregnant woman sample comprise a fetal free nucleic acid, the nucleic acid sequencing data are composed of a plurality of sequencing reads, the pregnant woman sample comprises at least one positive sample and at least one negative sample, the positive sample has an aneuploidy for a chromosome to be tested, and the negative sample does not have an aneuploidy for the chromosome to be tested;

determining a fetal fraction of the pregnant woman sample and an estimated fraction by a predetermined chromosome based on the nucleic acid sequencing data, wherein the estimated fraction by a predetermined chromosome is determined based on a difference between a number of sequencing reads of the predetermined chromosome and a number of sequencing reads of a first comparison chromosome, the predetermined chromosome comprises a chromosome to be tested and a second comparison chromosome, and the first comparison chromosome comprises at least one autosome different from the predetermined chromosome; and

determining a second feature based on a difference between the estimated fraction by the chromosome to be tested and the fetal fraction, and determining a first feature based on a difference between the estimated fraction by the chromosome to be tested and the estimated fraction by the second comparison chromosome;

a training model, which is configured to perform a machine learning training by taking the plurality of pregnant women samples as samples, so as to construct a machine learning classification model for determining whether the fetus has an aneuploidy;

and the device is further characterized by the following (a) to (d):

(a) the first comparison chromosome, the second comparison chromosome and the chromosome to be tested are derived from the same sample;

(b) the number of sequencing reads of the first comparison chromosome is an average number of sequencing reads of a plurality of autosomes, and the plurality of autosomes comprise at least one autosome that is known to have no aneuploidy, and the estimated fraction is determined according to the following formula:

$$F_j = 2 \times \left| R_j - R_r \right| / R_r$$

wherein,

j represents the serial number of a chromosome the estimated fraction of which needs to be determined,

F_j represents the estimated fraction by the chromosome j,

R_r represents the average number of sequencing reads of the plurality of autosomes;

(c) the first feature is determined by the following formula:

$$X_1 = F_i - F_r$$

wherein,

X_1 represents the first feature,

i represents the serial number of the chromosome to be tested,

F_i represents the estimated fraction by the chromosome to be tested,

F_r represents the average value of the estimated fraction by the second comparison chromosome;

(d) the second feature is determined by the following formula:

$$X_2 = \frac{F_i}{F_a}$$

wherein,

X_2 represents the second feature,

i represents the serial number of the chromosome to be tested,

F_i represents the estimated fraction by the chromosome to be tested,

F_a represents the fetal fraction.

38. The device according to claim 37, characterized in that the machine learning classification model is a KNN model.

39. A computer-readable storage medium and for use in the method according to any one of claims 34 to 36, on which a computer program is stored, characterized in that when the program is executed by a processor, the steps of the method according to any one of claims 34 to 36 are implemented.