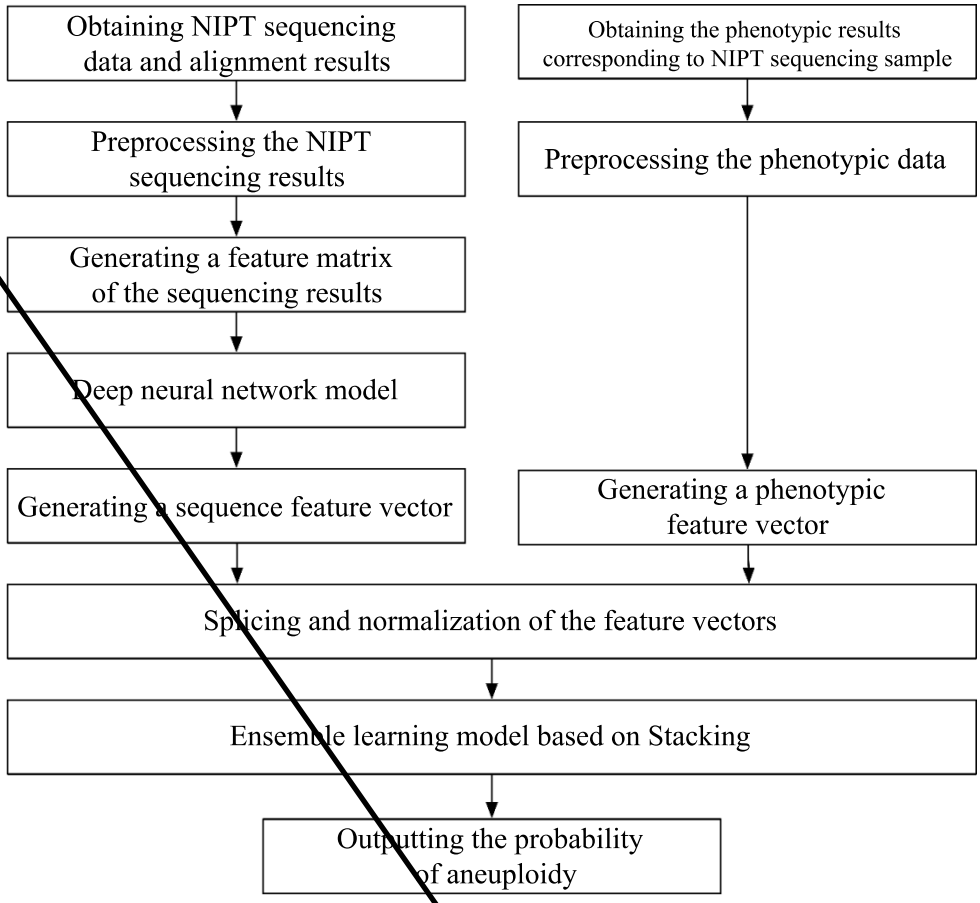
Obtaining NIPT sequencing data and alignment results
Obtaining the phenotypic results corresponding to NIPT sequencing sample
Preprocessing the NIPT sequencing results
Preprocessing the phenotypic data
Generating a feature matrix of the sequencing results
Deep neural network model
Generating a sequence feature vector
Generating a phenotypic feature vector
Splicing and normalization of the feature vectors
Ensemble learning model based on Stacking
Outputting the probability of aneuploidy

FIG. 1

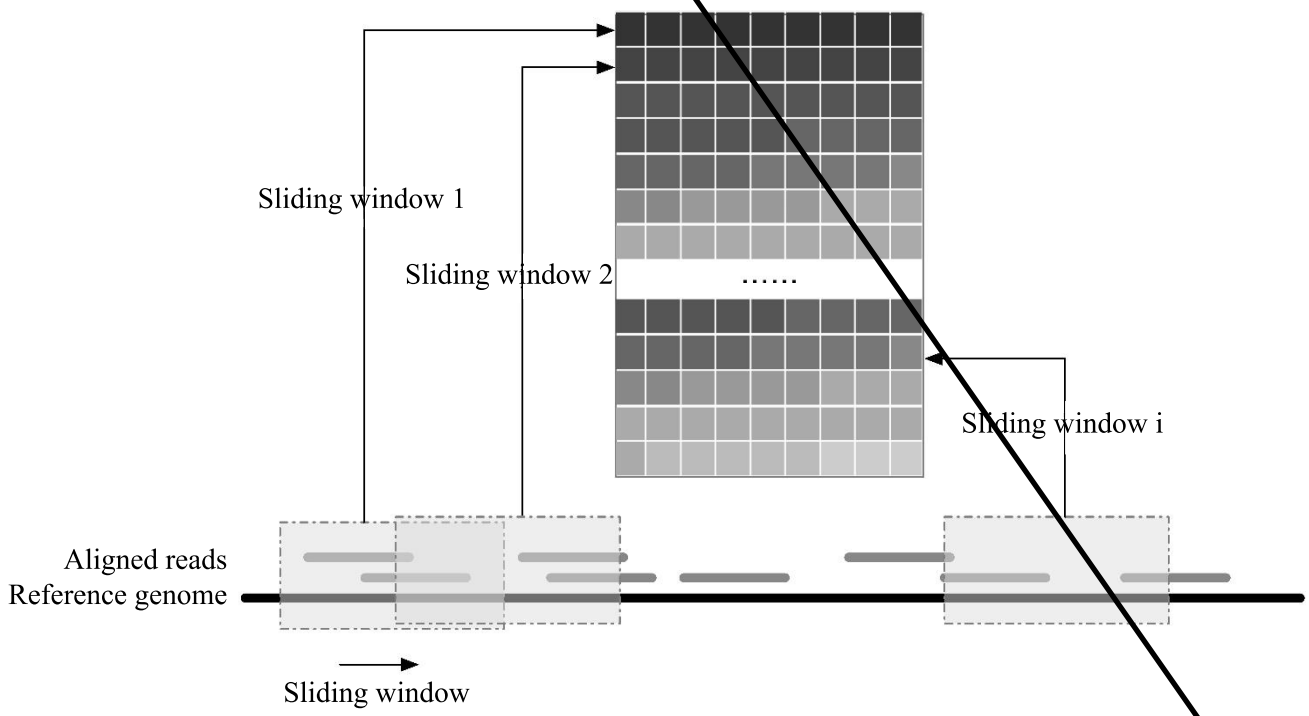
Sliding window 1
Sliding window 2
……
Sliding window i
Aligned reads
Reference genome
Sliding window

FIG. 2

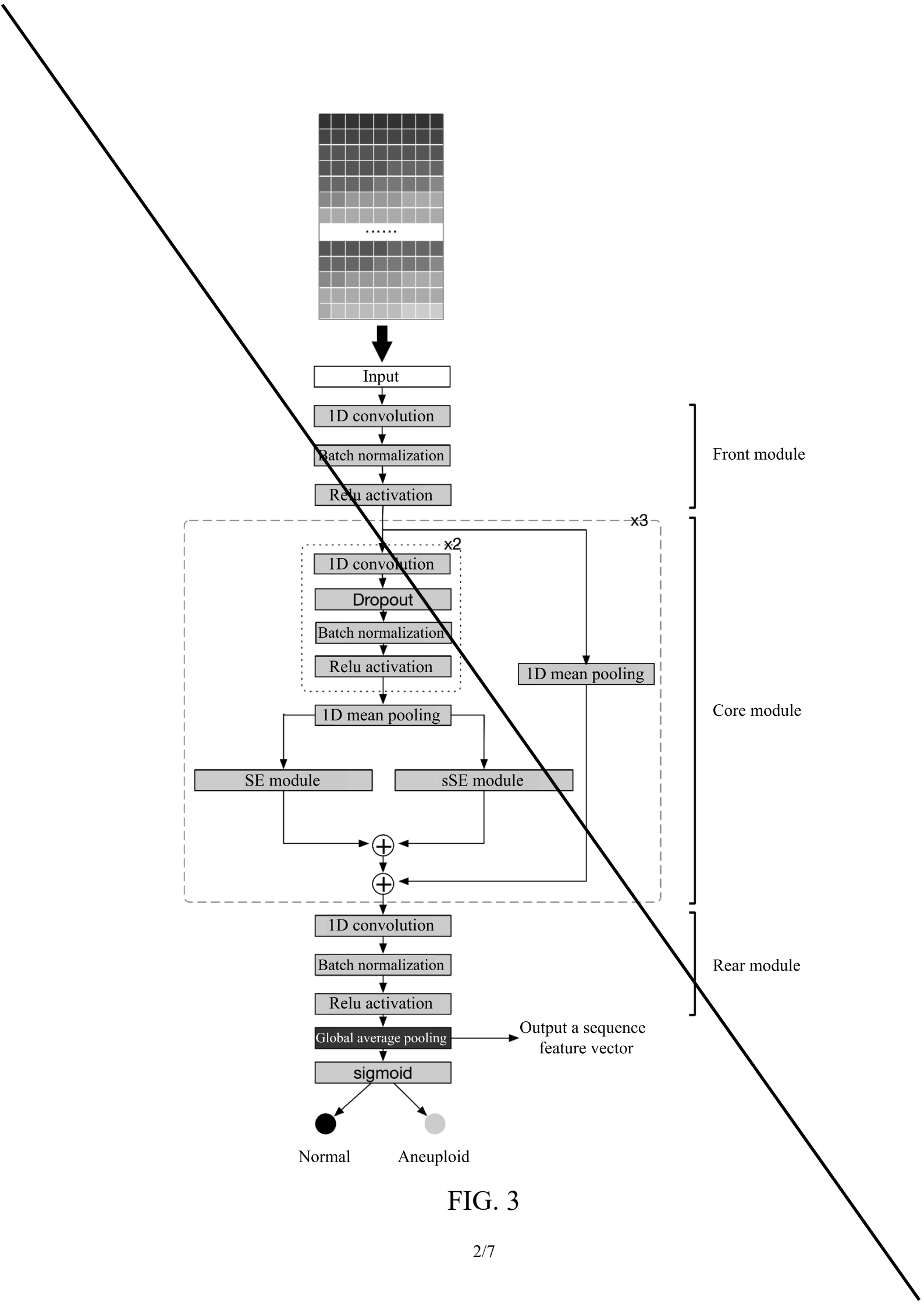
.....
Input
1D convolution
Batch normalization
Relu activation
Front module
x3
x2
1D convolution
Dropout
Batch normalization
Relu activation
1D mean pooling
1D mean pooling
SE module
sSE module
Core module
1D convolution
Batch normalization
Relu activation
Global average pooling
Output a sequence feature vector
sigmoid
Rear module
Normal
Aneuploid

FIG. 3

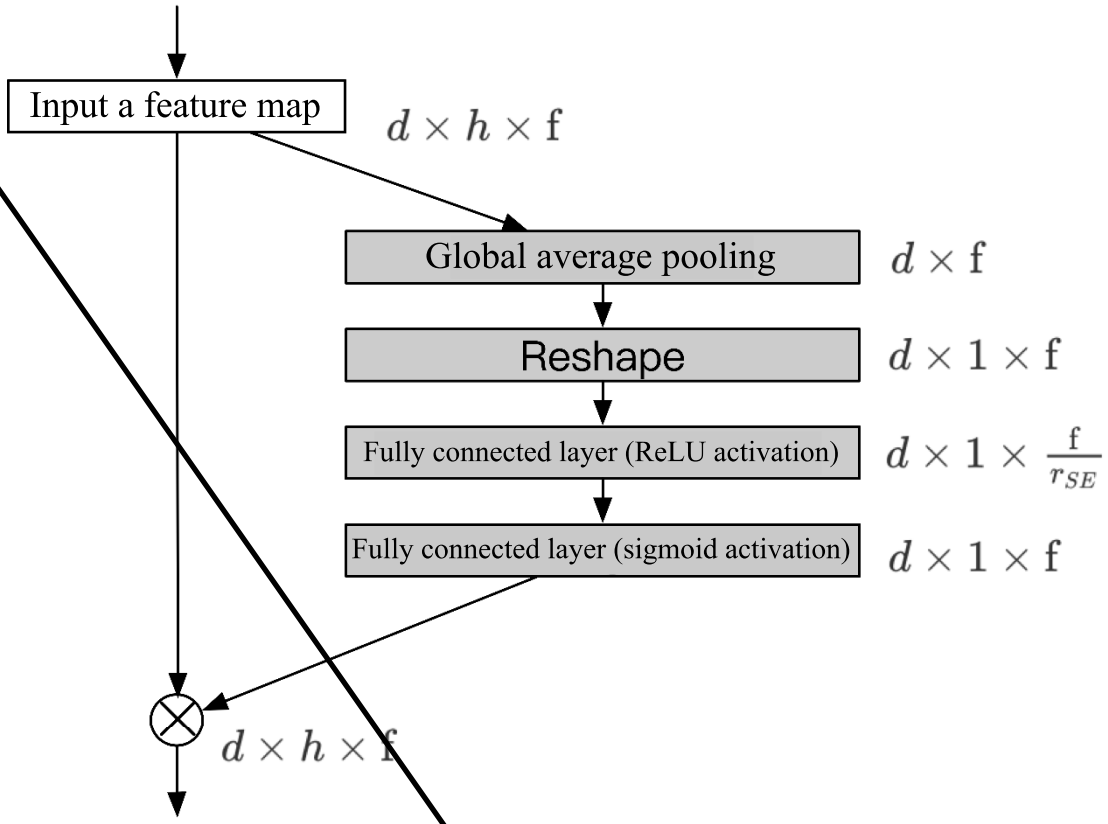
Input a feature map
$d \times h \times \mathrm{f}$
Global average pooling
$d \times \mathrm{f}$
Reshape
$d \times 1 \times \mathrm{f}$
Fully connected layer (ReLU activation)
$d \times 1 \times \frac{\mathrm{f}}{r_{SE}}$
Fully connected layer (sigmoid activation)
$d \times 1 \times \mathrm{f}$
$d \times h \times \mathrm{f}$

FIG. 4

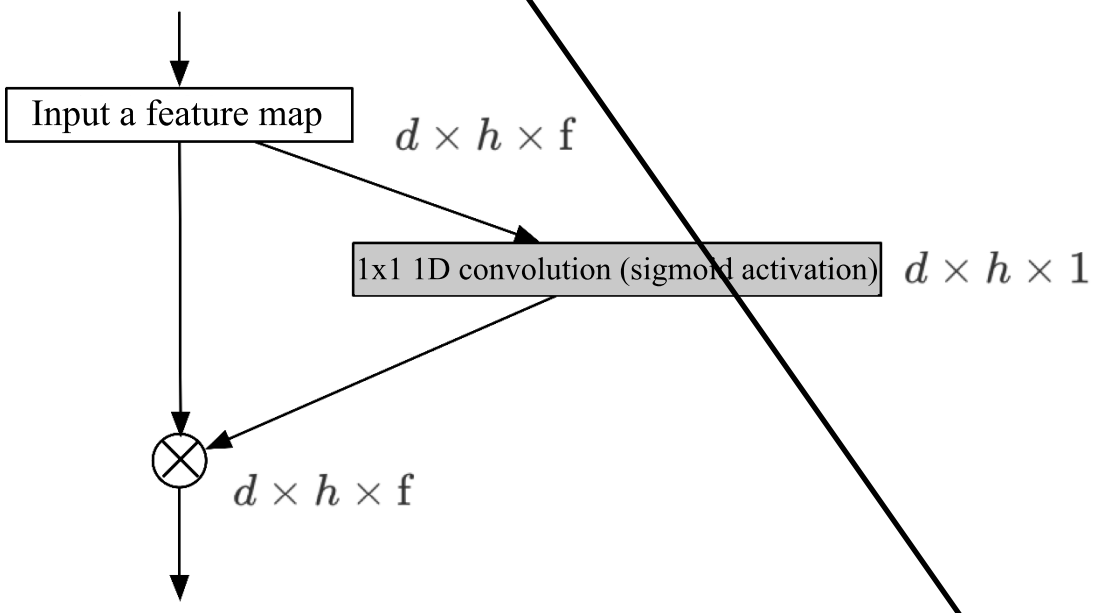
Input a feature map
$d \times h \times \mathrm{f}$
1x1 1D convolution (sigmoid activation)
$d \times h \times 1$
$d \times h \times \mathrm{f}$

FIG. 5

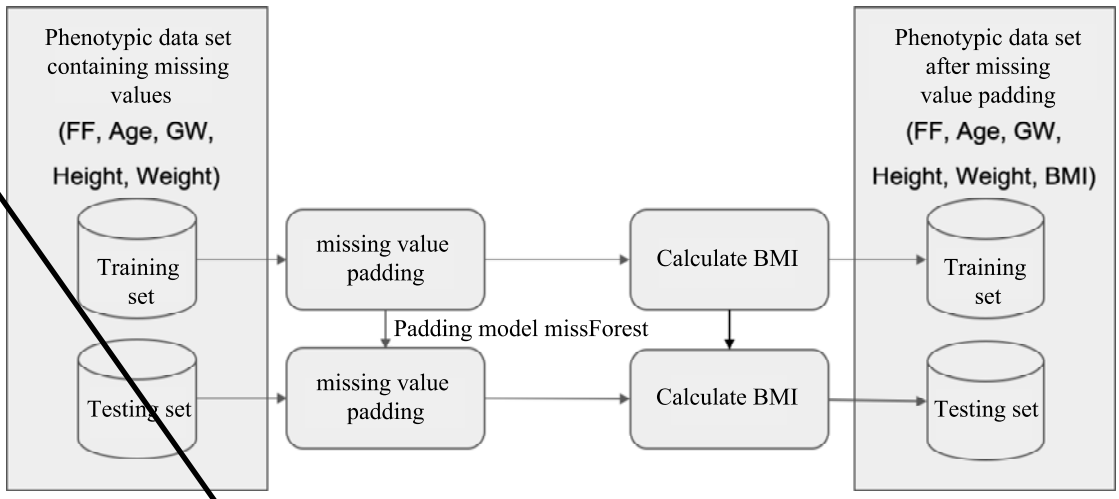
Phenotypic data set containing missing values
(FF, Age, GW, Height, Weight)
Training set
Testing set
missing value padding
missing value padding
Padding model missForest
Calculate BMI
Calculate BMI
Phenotypic data set after missing value padding
(FF, Age, GW, Height, Weight, BMI)
Training set
Testing set

FIG. 6

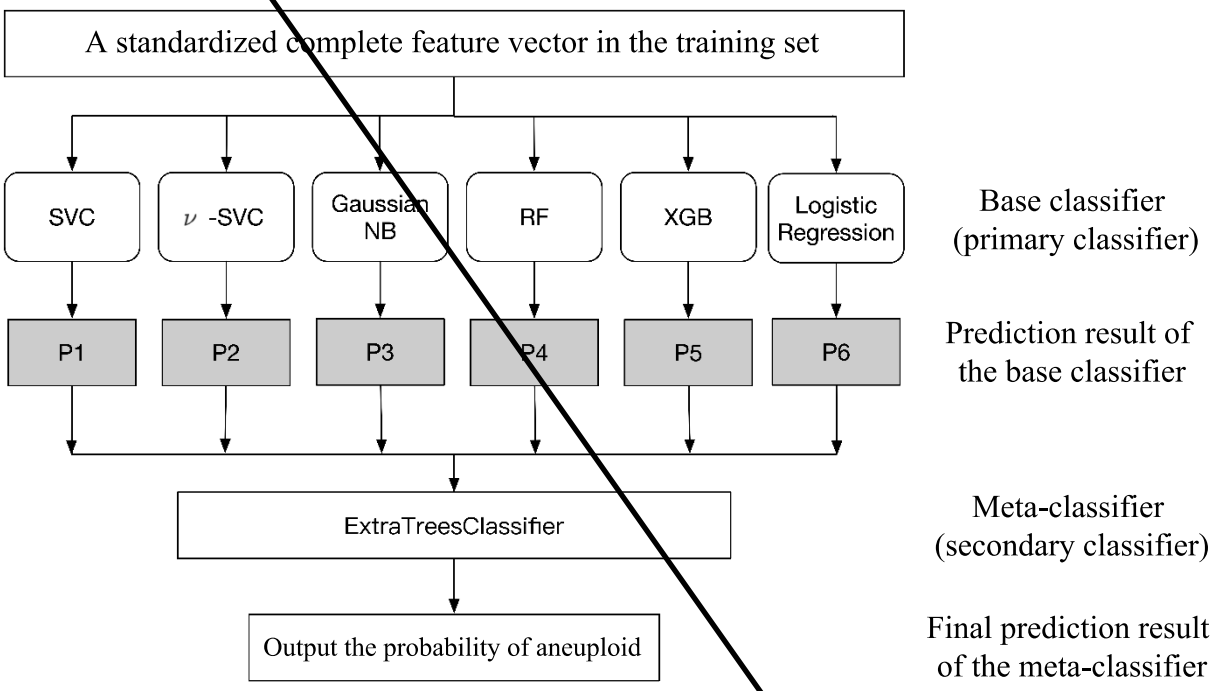
A standardized complete feature vector in the training set
SVC
ν -SVC
Gaussian NB
RF
XGB
Logistic Regression
P1
P2
P3
P4
P5
P6
ExtraTreesClassifier
Output the probability of aneuploid
Base classifier (primary classifier)
Prediction result of the base classifier
Meta-classifier (secondary classifier)
Final prediction result of the meta-classifier

FIG. 7

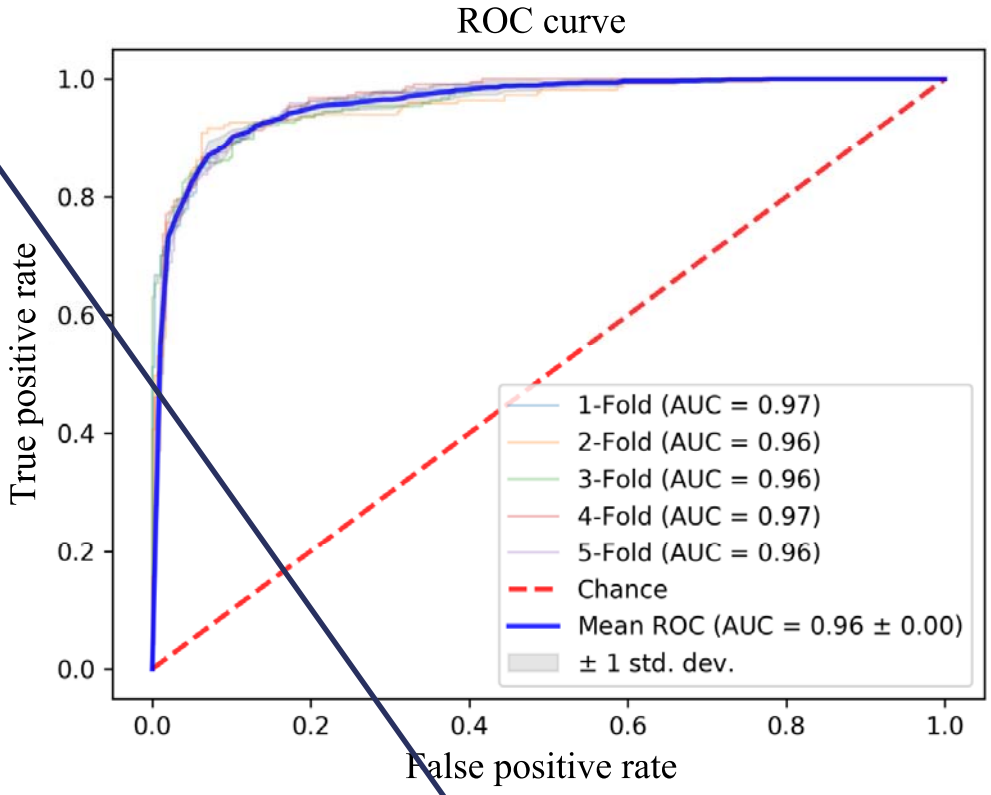
ROC curve
True positive rate
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0
1-Fold (AUC = 0.97)
2-Fold (AUC = 0.96)
3-Fold (AUC = 0.96)
4-Fold (AUC = 0.97)
5-Fold (AUC = 0.96)
Chance
Mean ROC (AUC = 0.96 ± 0.00)
± 1 std. dev.

FIG. 8

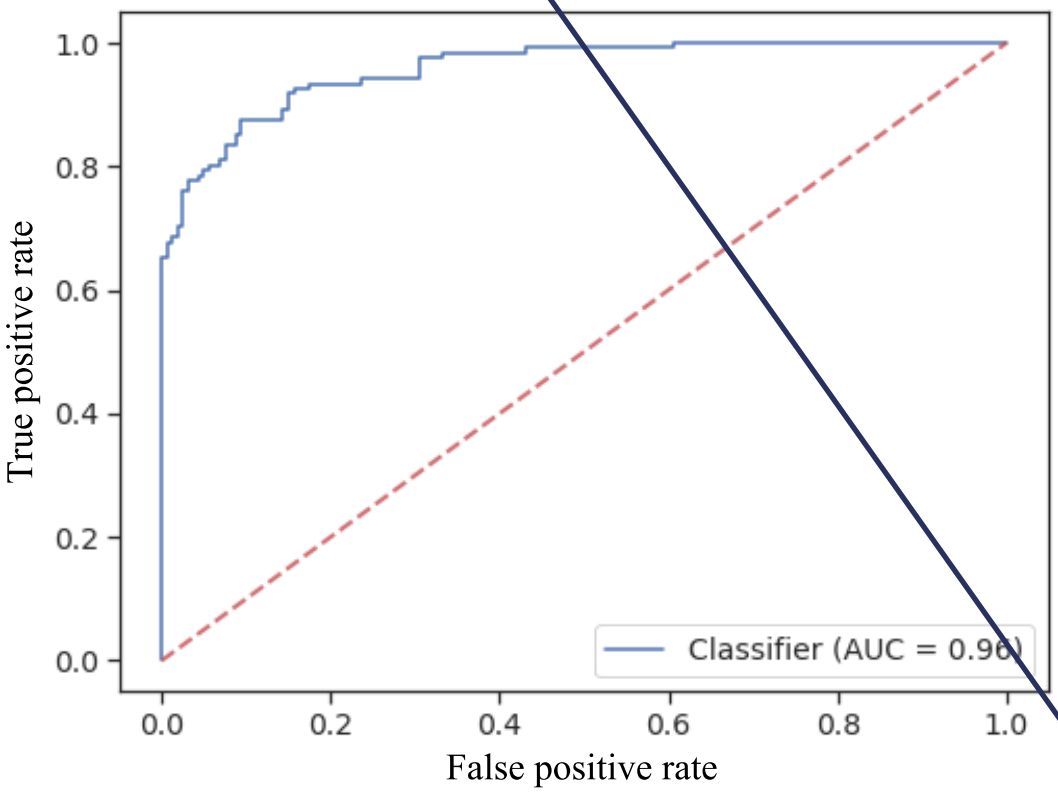
True positive rate
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0
Classifier (AUC = 0.96)

FIG. 9

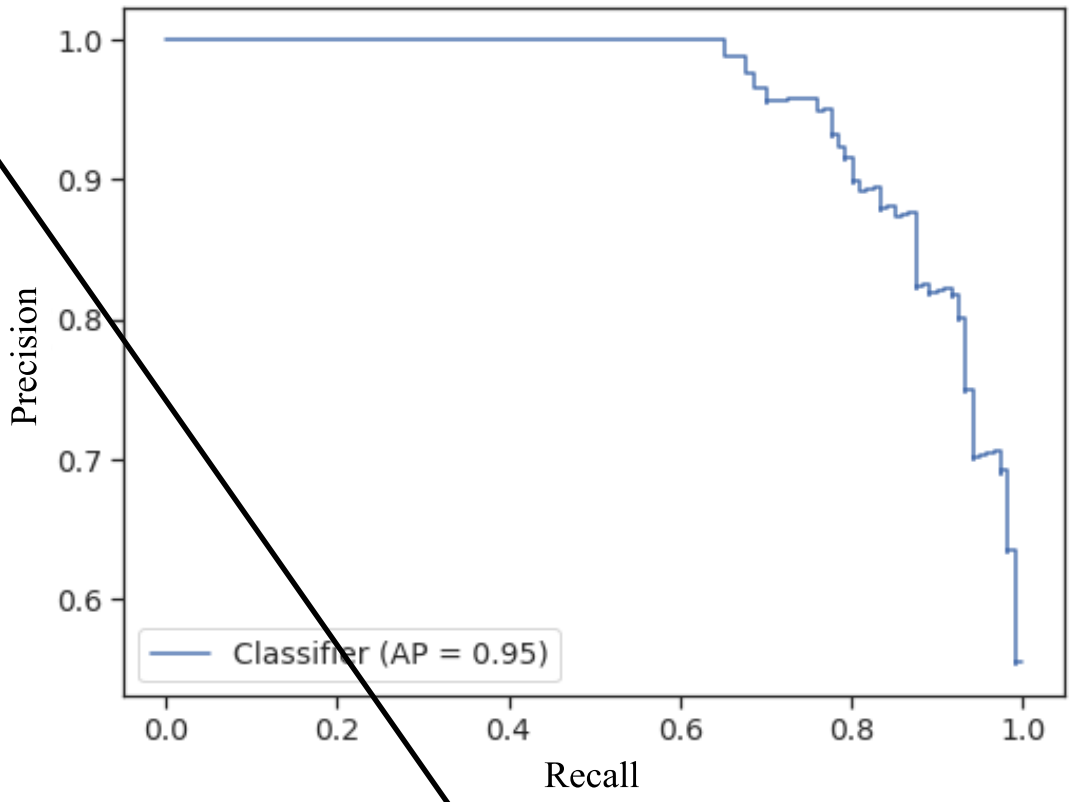
1.0
0.9
0.8
0.7
0.6
Precision
Classifier (AP = 0.95)
0.0
0.2
0.4
0.6
0.8
1.0
Recall

FIG. 10

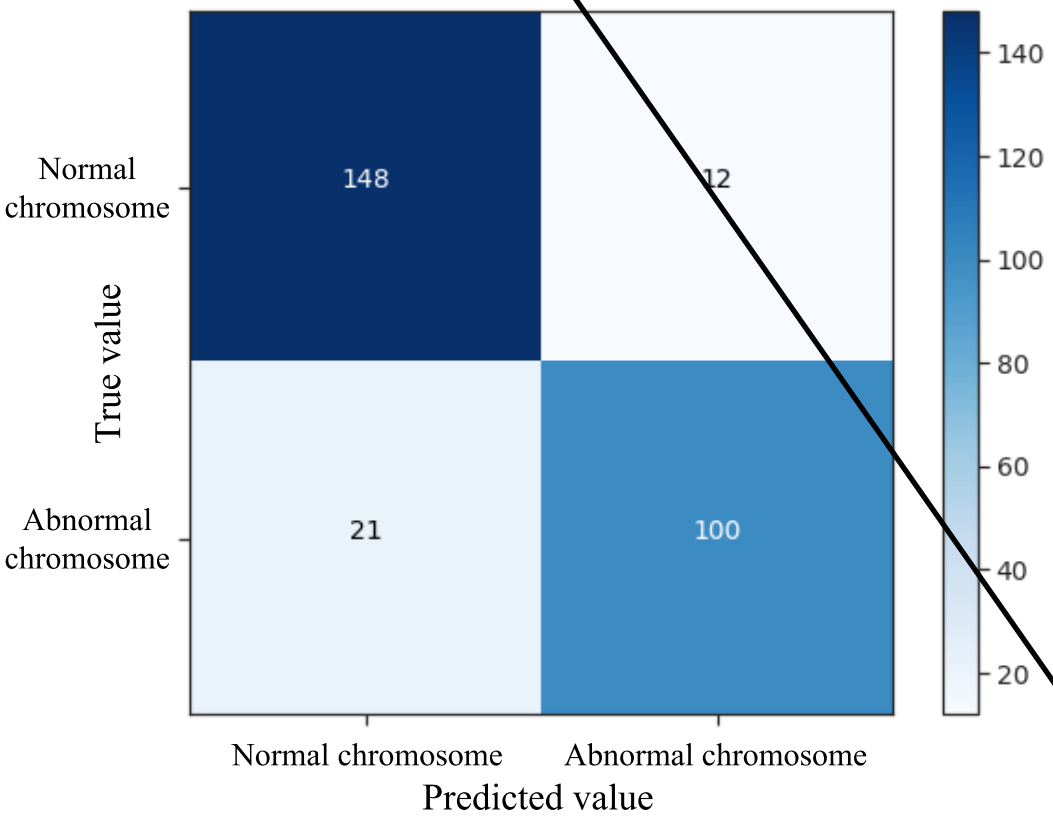
Normal chromosome
Abnormal chromosome
True value
148
12
21
100
140
120
100
80
60
40
20
Normal chromosome
Abnormal chromosome
Predicted value

FIG. 11

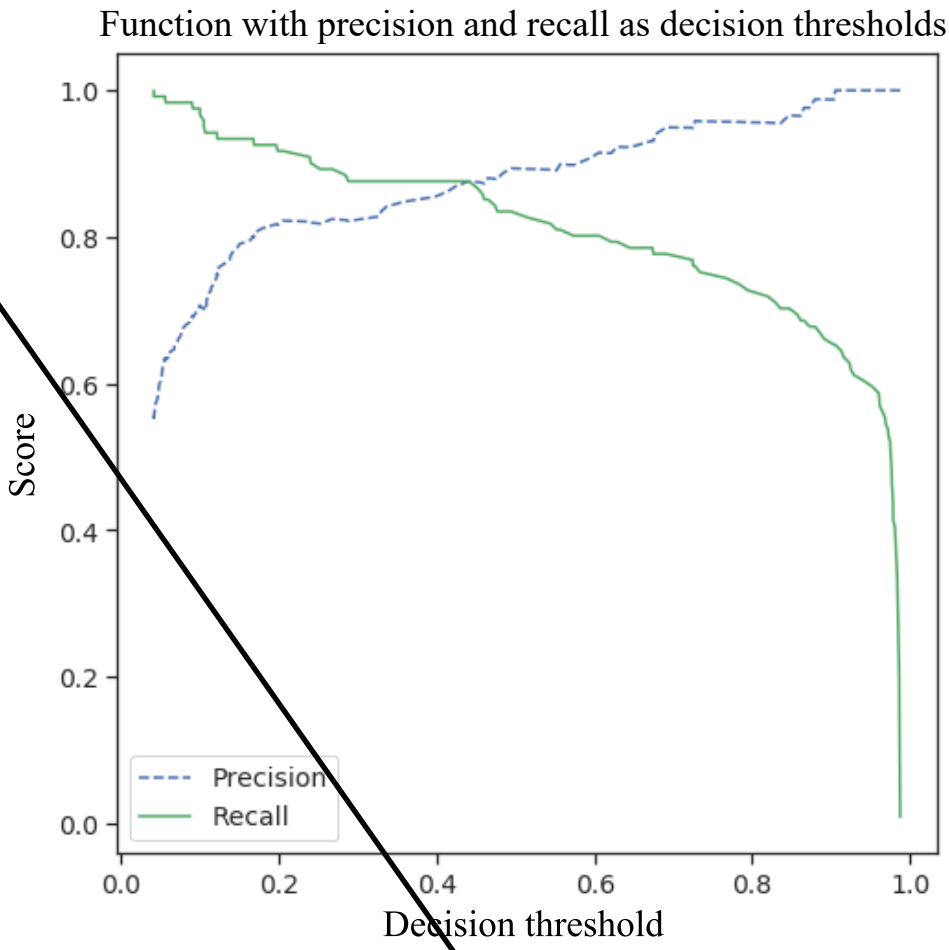
Function with precision and recall as decision thresholds
Score
Decision threshold
Precision
Recall
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

FIG. 12

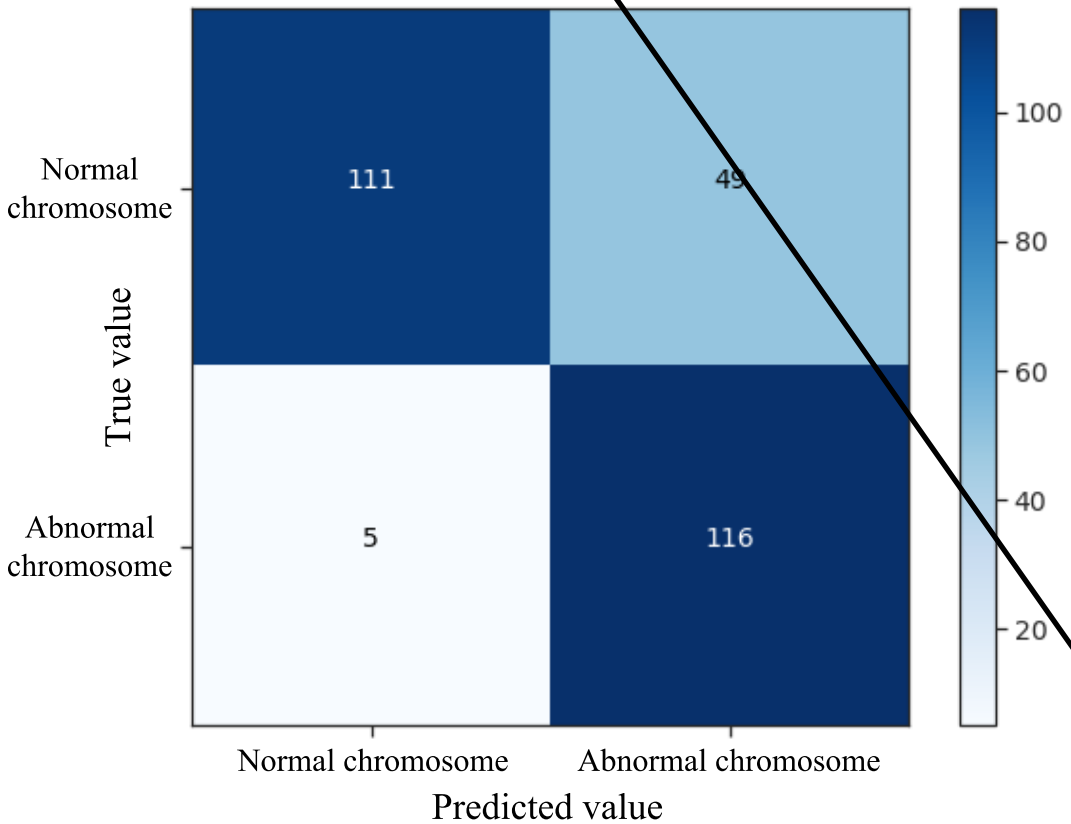
Normal chromosome
Abnormal chromosome
True value
111
49
5
116
Normal chromosome
Abnormal chromosome
Predicted value
100
80
60
40
20

FIG. 13

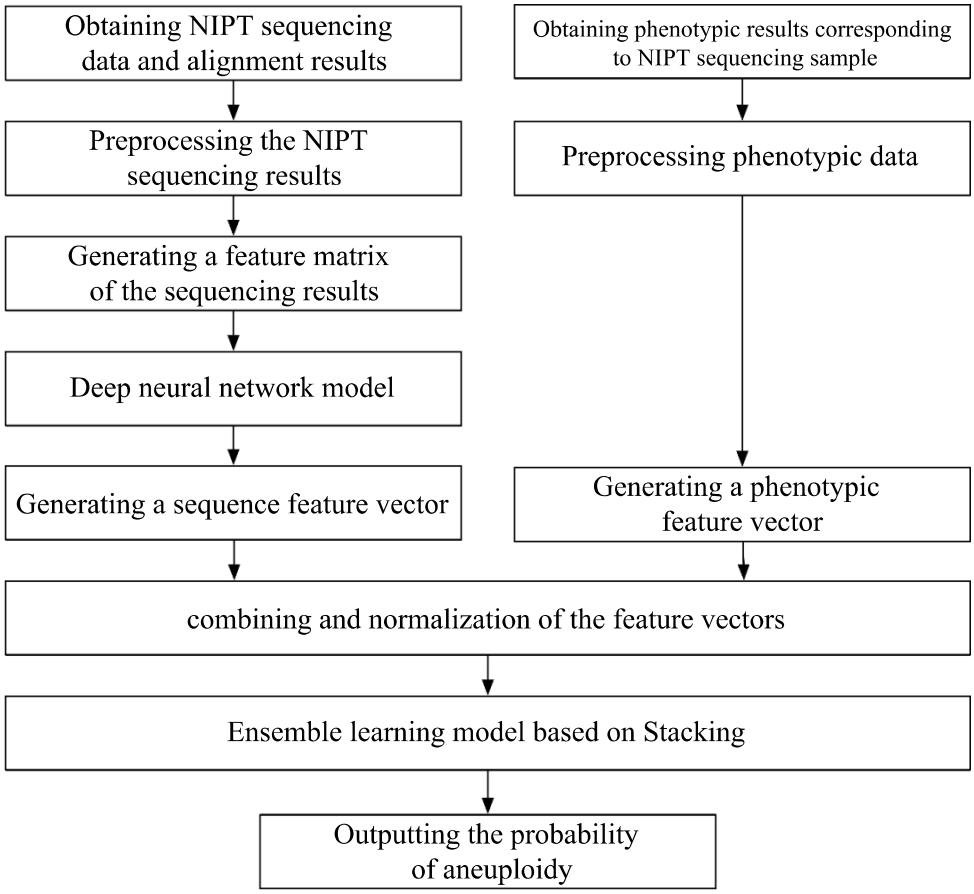
Obtaining NIPT sequencing data and alignment results
Obtaining phenotypic results corresponding to NIPT sequencing sample
Preprocessing the NIPT sequencing results
Preprocessing phenotypic data
Generating a feature matrix of the sequencing results
Deep neural network model
Generating a sequence feature vector
Generating a phenotypic feature vector
combining and normalization of the feature vectors
Ensemble learning model based on Stacking
Outputting the probability of aneuploidy

FIG. 1

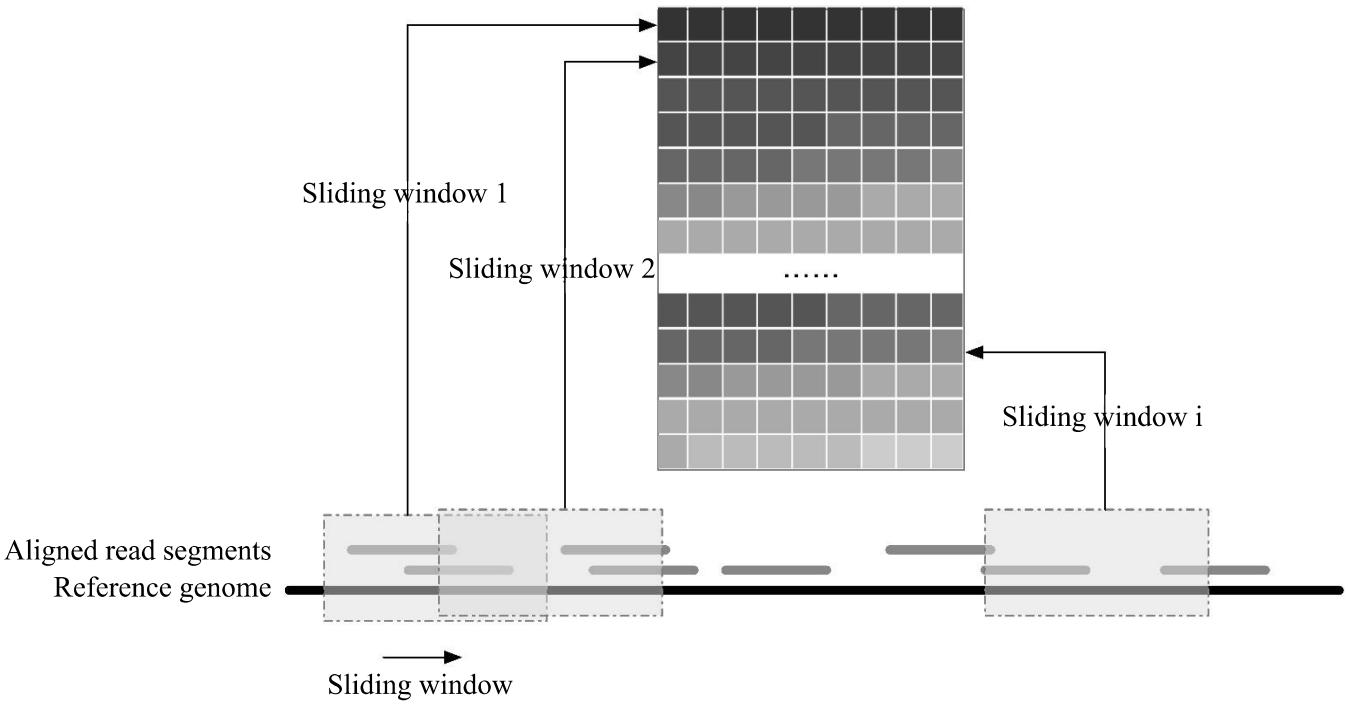
Sliding window 1
Sliding window 2
……
Sliding window i
Aligned read segments
Reference genome
Sliding window

FIG. 2

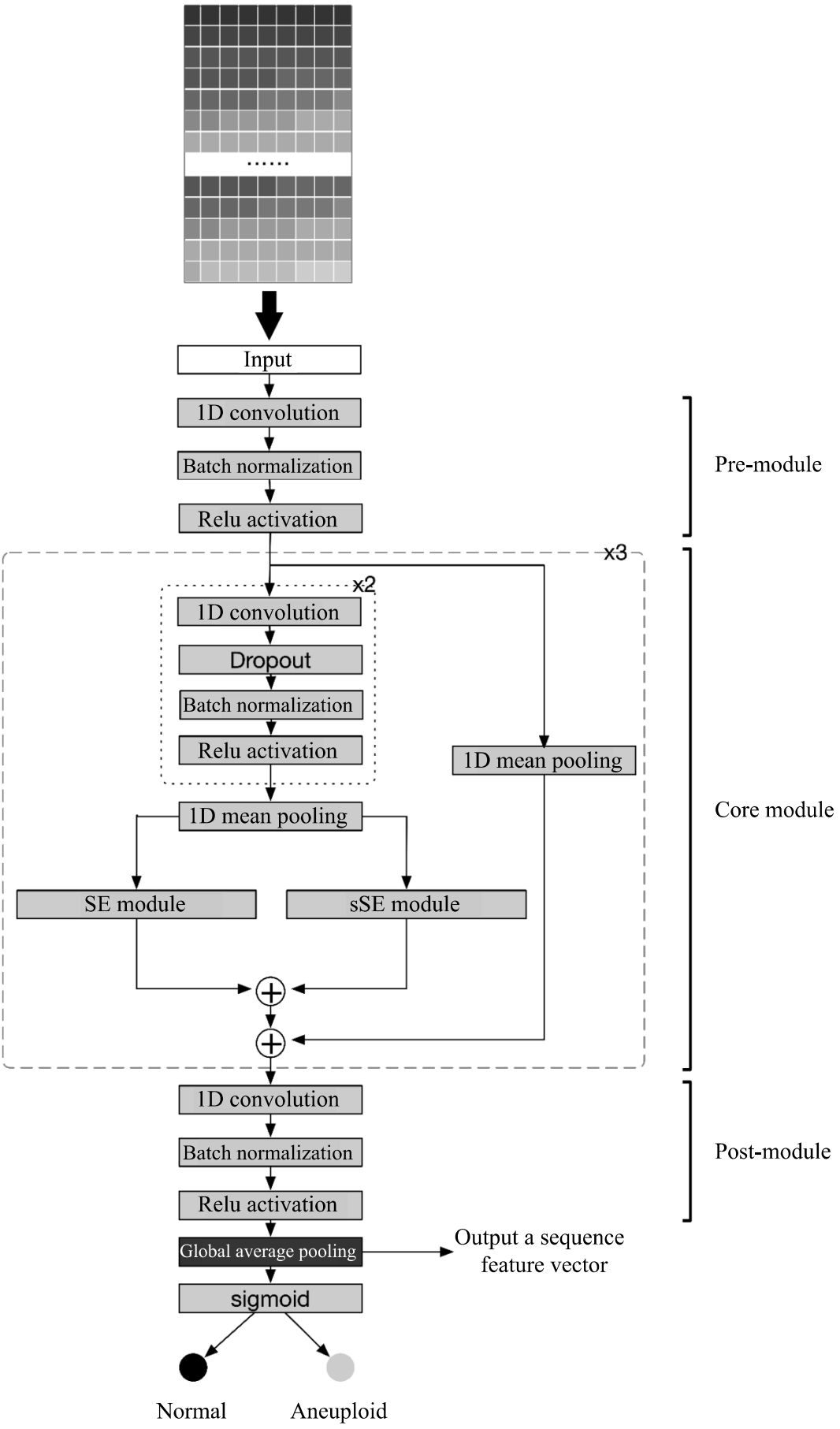
.....
Input
1D convolution
Batch normalization
Relu activation
Pre-module
x3
x2
1D convolution
Dropout
Batch normalization
Relu activation
1D mean pooling
1D mean pooling
SE module
sSE module
Core module
1D convolution
Batch normalization
Relu activation
Global average pooling
Output a sequence feature vector
sigmoid
Post-module
Normal
Aneuploid

FIG. 3

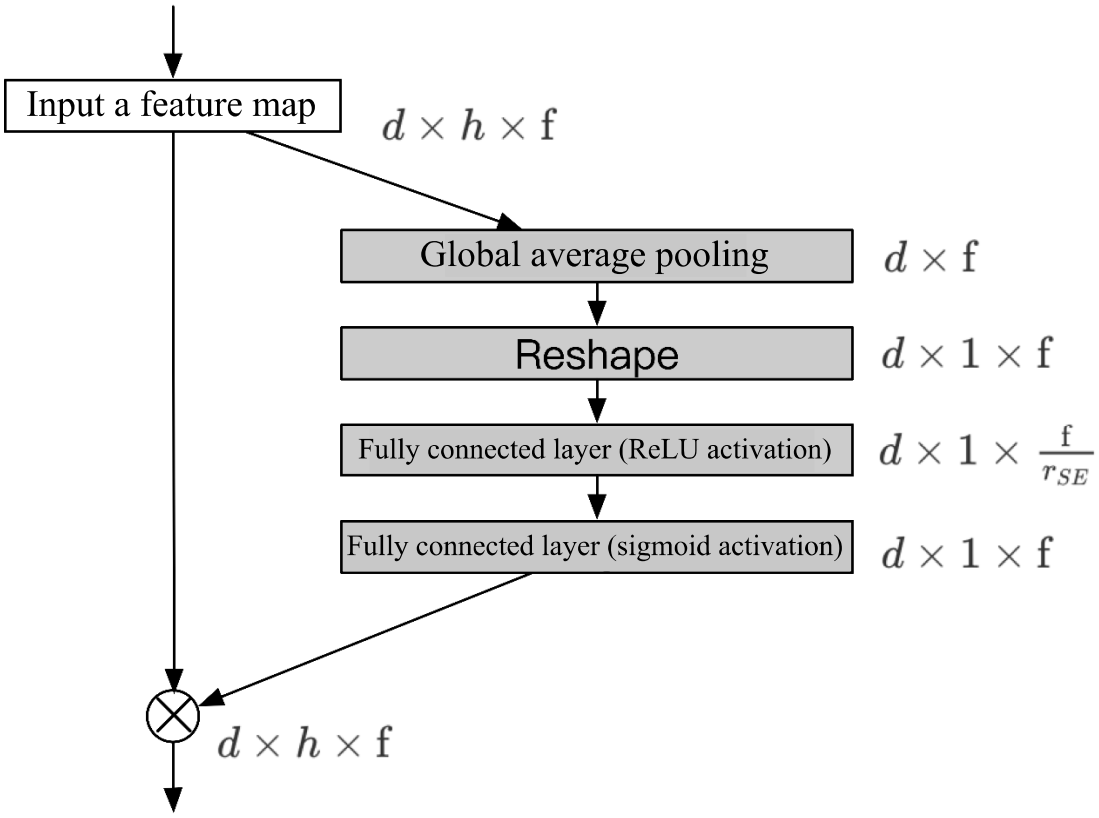
Input a feature map
$d \times h \times \mathrm{f}$
Global average pooling
$d \times \mathrm{f}$
Reshape
$d \times 1 \times \mathrm{f}$
Fully connected layer (ReLU activation)
$d \times 1 \times \frac{\mathrm{f}}{r_{SE}}$
Fully connected layer (sigmoid activation)
$d \times 1 \times \mathrm{f}$
$d \times h \times \mathrm{f}$

FIG. 4

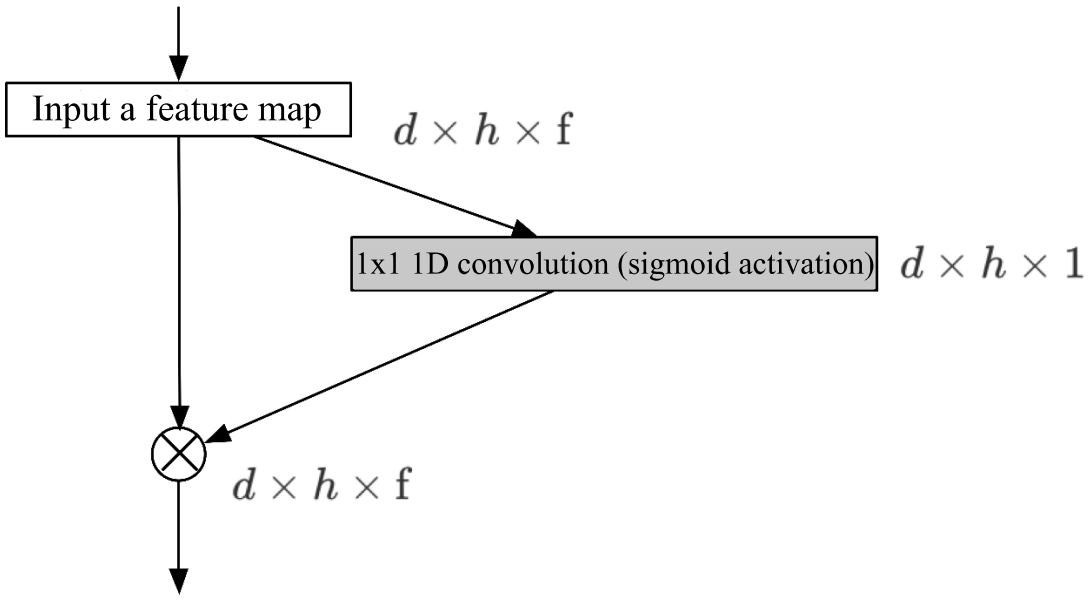
Input a feature map
$d \times h \times \mathrm{f}$
1x1 1D convolution (sigmoid activation)
$d \times h \times 1$
$d \times h \times \mathrm{f}$

FIG. 5

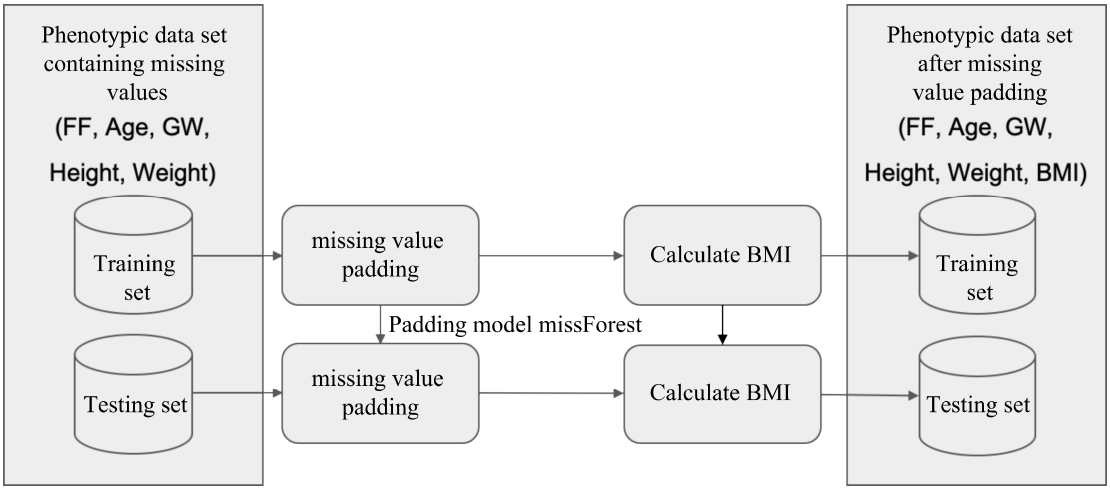
Phenotypic data set containing missing values
(FF, Age, GW, Height, Weight)
Training set
Testing set
missing value padding
missing value padding
Padding model missForest
Calculate BMI
Calculate BMI
Phenotypic data set after missing value padding
(FF, Age, GW, Height, Weight, BMI)
Training set
Testing set

FIG. 6

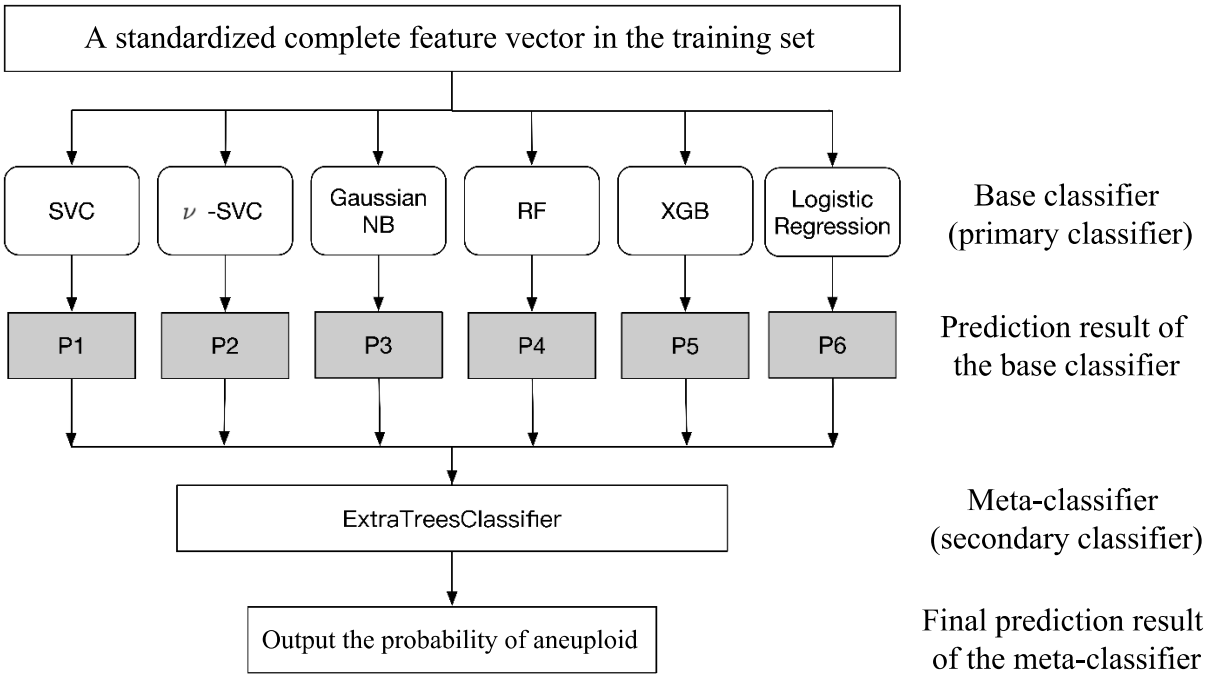
A standardized complete feature vector in the training set
SVC
ν -SVC
Gaussian NB
RF
XGB
Logistic Regression
P1
P2
P3
P4
P5
P6
ExtraTreesClassifier
Output the probability of aneuploid
Base classifier (primary classifier)
Prediction result of the base classifier
Meta-classifier (secondary classifier)
Final prediction result of the meta-classifier

FIG. 7

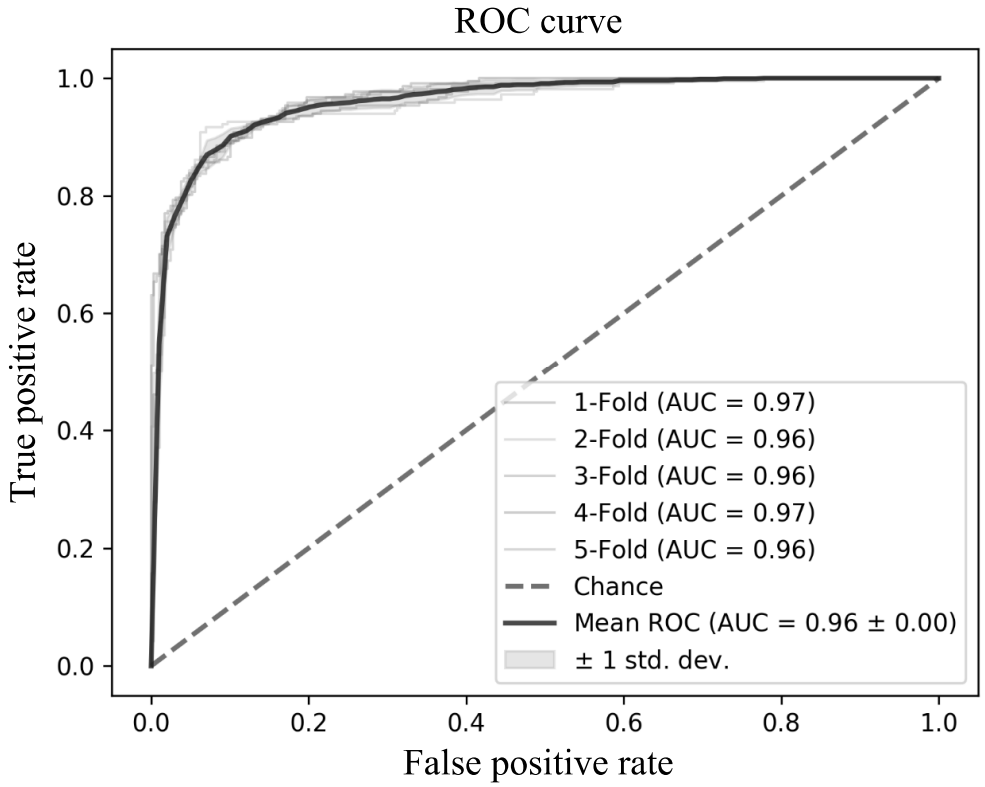
ROC curve
True positive rate
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0
1-Fold (AUC = 0.97)
2-Fold (AUC = 0.96)
3-Fold (AUC = 0.96)
4-Fold (AUC = 0.97)
5-Fold (AUC = 0.96)
Chance
Mean ROC (AUC = 0.96 ± 0.00)
± 1 std. dev.

FIG. 8

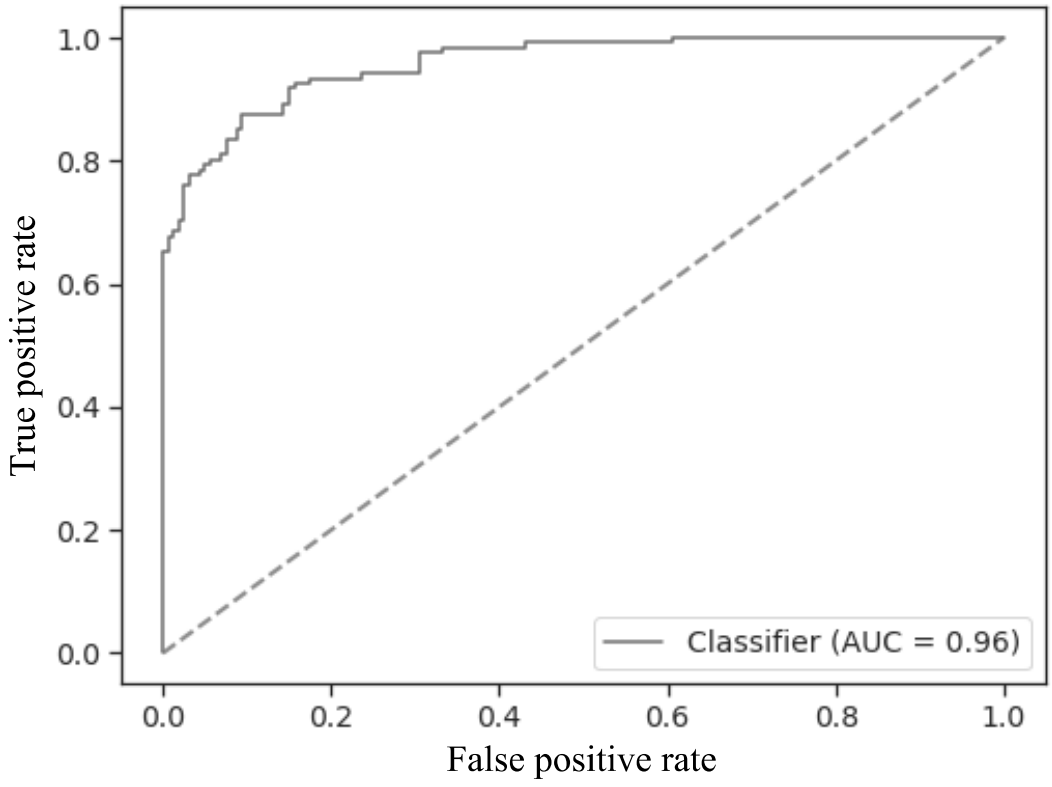
True positive rate
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0
Classifier (AUC = 0.96)

FIG. 9

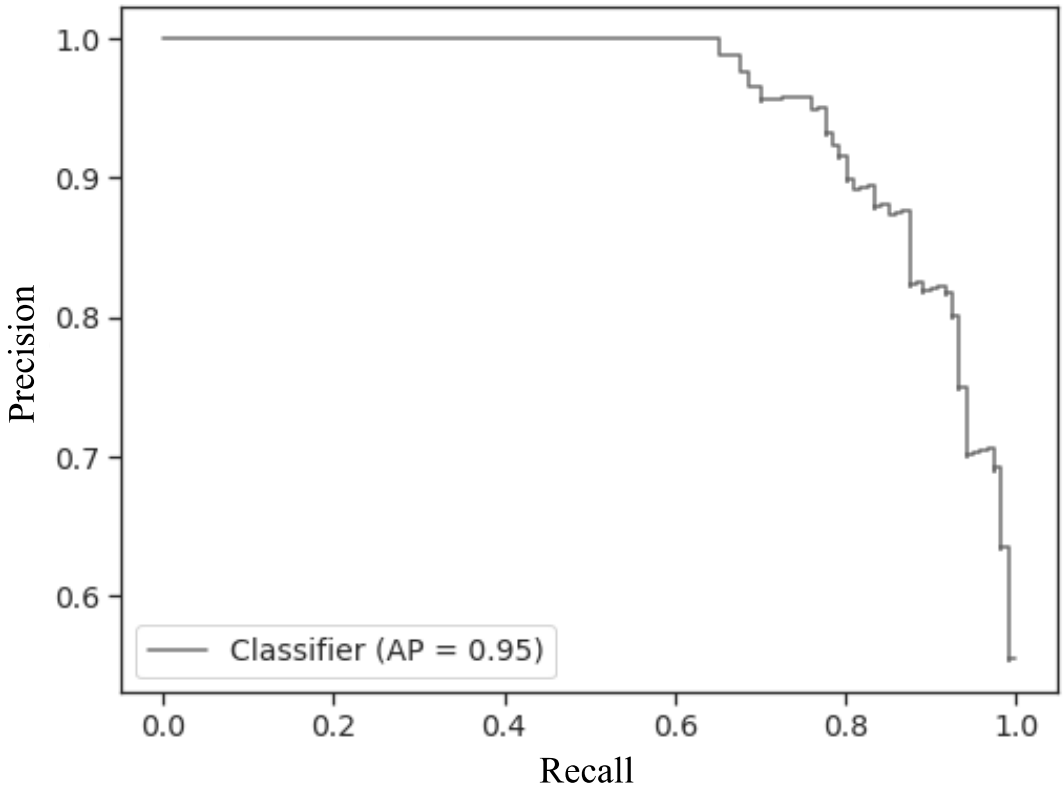
1.0
0.9
0.8
0.7
0.6
Precision
Classifier (AP = 0.95)
0.0
0.2
0.4
0.6
0.8
1.0
Recall

FIG. 10

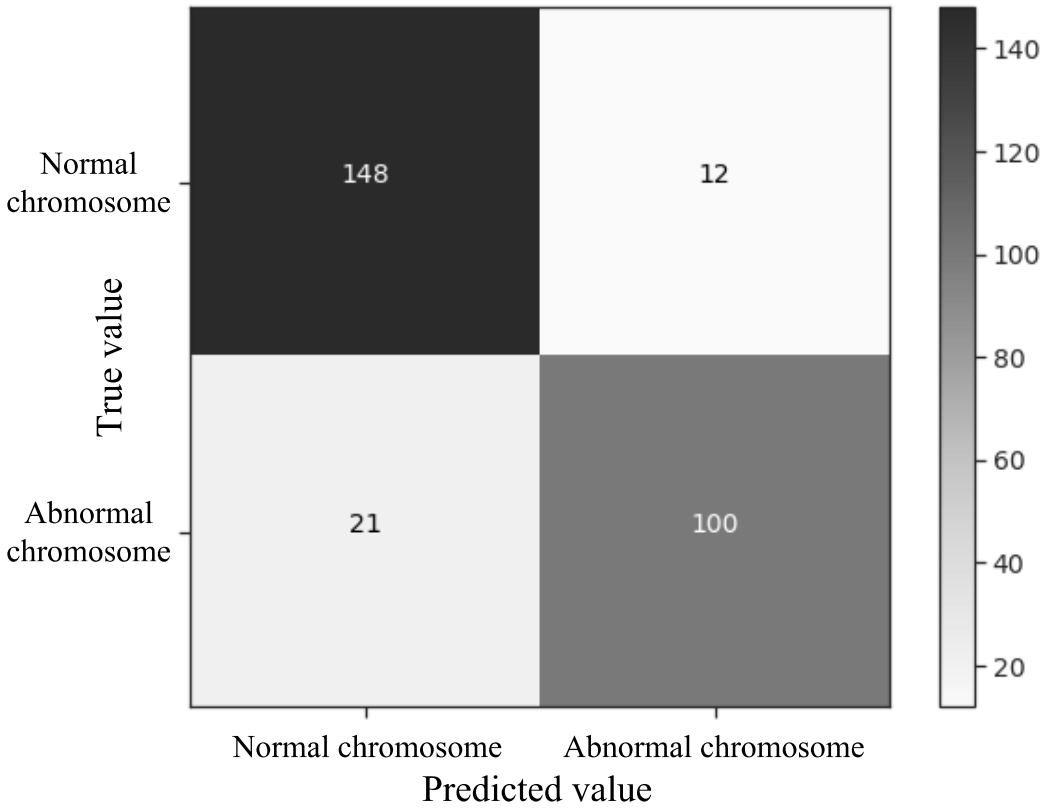
Normal chromosome
Abnormal chromosome
True value
148
12
21
100
140
120
100
80
60
40
20
Normal chromosome
Abnormal chromosome
Predicted value

FIG. 11

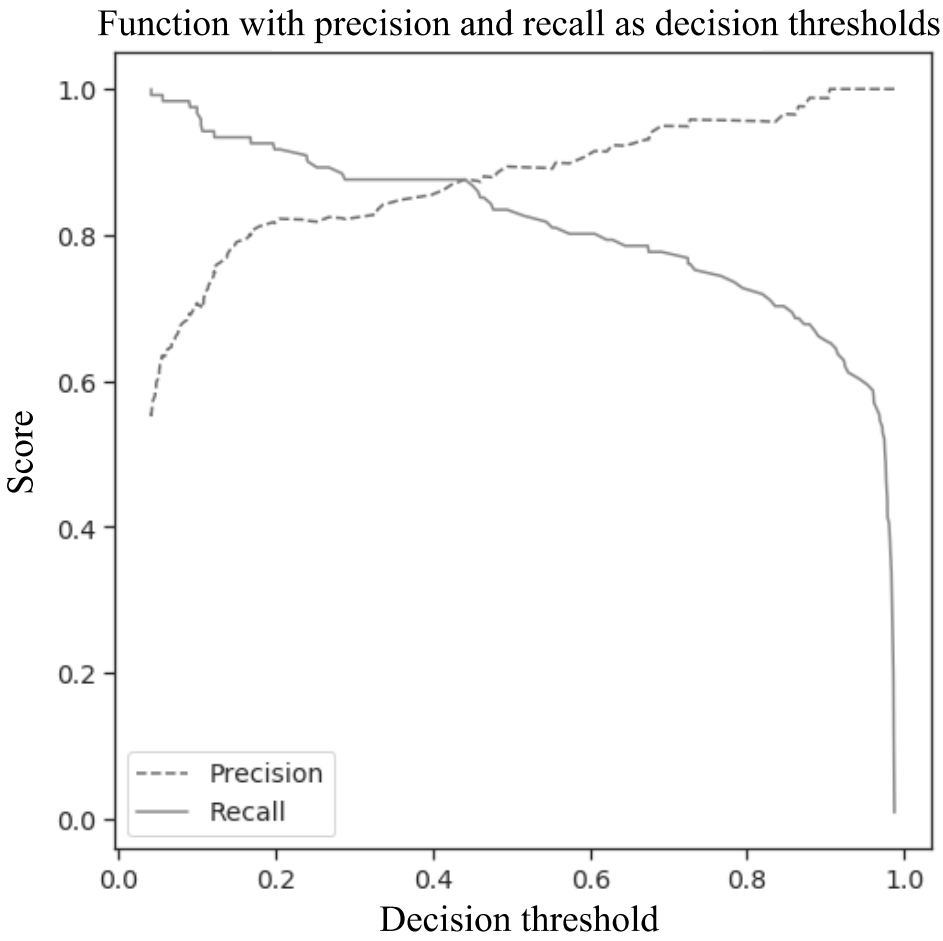
Function with precision and recall as decision thresholds
Score
Decision threshold
Precision
Recall
0.0
0.2
0.4
0.6
0.8
1.0

FIG. 12

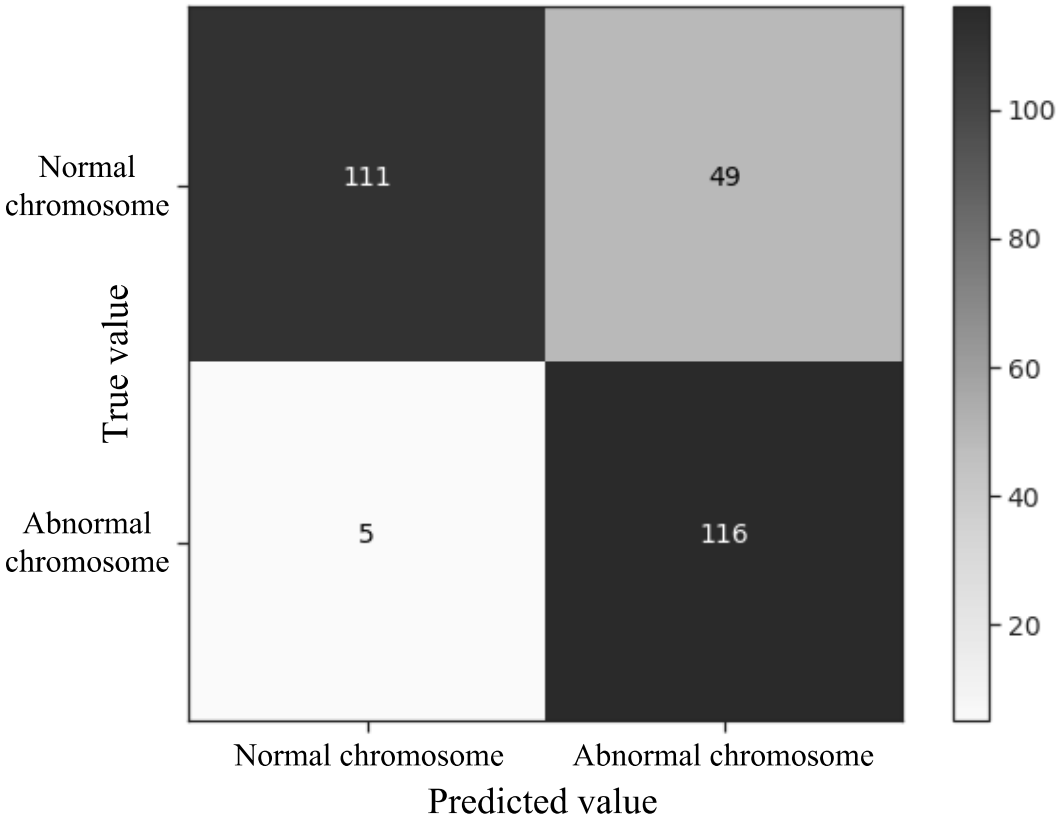
Normal chromosome
Abnormal chromosome
True value
111
49
5
116
Normal chromosome
Abnormal chromosome
Predicted value
100
80
60
40
20

FIG. 13