A Machine Learning-based NIPT Intelligent Timing Recommendation and Fetal Abnormality Assistance Diagnosis System

Zihao Zhang, Simin Wang*, Junjie Jiang, Jia Sheng, Boyan Zhang, Zhongyi Li, Ziwu Wang
Jiangsu Normal University KeWen College, Xuzhou, Jiangsu, China
**Corresponding Author*

Abstract: In response to the two major pain points in the clinical application of non-invasive prenatal genetic testing (NIPT): the "one-size-fits-all" approach to determining gestational age and the "single-indicator" method for diagnosing chromosomal abnormalities in female fetuses, this study proposes a machine learning-based NIPT intelligent timing recommendation and fetal anomaly auxiliary judgment system. First, using the K-Means unsupervised clustering algorithm and the elbow method to determine the optimal clustering number K = 3, the BMI of pregnant women is automatically categorized into low, medium, and high-risk groups, replacing the traditional fixed-range empirical categorization method. Second, based on local weighted regression (LOESS), curves are constructed for the attainment rates of Y-chromosome concentrations in each group, and the optimal NIPT testing times for each group are determined by the intersection points of horizontal and vertical cutoff lines, at 22.143 weeks, 24.286 weeks, and 27.714 weeks, respectively. Finally, seven core indicators, including BMI, chromosome Z value, GC content, sequencing read segment proportion, and X-chromosome Z value, are integrated into a random forest classification model to quantify the risk probability of chromosomal abnormalities in female fetuses. The AUC value achieved is 0.799. On this basis, K-Means clustering and the random forest model were integrated into a web system using Flask+Vue.js, creating a NIPT intelligent decision support system that covers the entire business process of data entry, model calculation, and result visualization. This provides a novel technological solution for personalized and precise decision-making in prenatal screening.

Keywords: Non-Invasive Prenatal Genetic Testing (NIPT); Machine Learning; K-Means Clustering; Random Forest; Intelligent Decision Support; Time-based Recommendations

1. Introduction

Non-invasive prenatal testing (NIPT) analyzes fetal free DNA (cell-free fetal DNA, cffDNA) present in the mother's peripheral blood to screen for non-disjunctional chromosomal abnormalities in the fetus. Due to its non-invasive, safe, and high detection rate advantages, NIPT has become the primary core technology for clinical prenatal screening [1]. However, NIPT still faces two major challenges in terms of standardization for its widespread clinical application, which hinders its development towards greater precision.

Firstly, the detection of gestational age employs a one-size-fits-all approach. Current clinical guidelines generally recommend performing NIPT testing after 12 weeks of pregnancy. However, this empirical standard does not fully account for the critical impact of a pregnant woman's Body Mass Index (BMI) on the concentration of cfDNA. Women with a high BMI often experience an increased maternal background DNA dilution effect due to the metabolism of fat tissue, resulting in insufficient concentrations of fetal DNA [2]. Especially for the detection of male fetuses, the clinical threshold of Y chromosome concentration ≥ 4% must be met. Failure to meet this threshold will result in a failed test or unreliable results, forcing a postponement of the test or repeated blood collection. This not only increases medical costs but also risks missing the optimal time for prenatal screening interventions [3]. Secondly, the determination of chromosomal abnormalities in female fetuses has a "single-indicator limitation." Since female fetuses do not carry the Y chromosome, clinical diagnosis of their chromosomal abnormalities relies solely on a

single Z-value indicator for chromosomes 21, 18, and 13. This approach fails to fully utilize multidimensional sequencing information such as GC content, sequencing read segment ratios, and X-chromosome Z-values, making it difficult to further improve the sensitivity and specificity of abnormal screening, which poses a risk of missed diagnoses and misdiagnoses [4].

In response to the above clinical challenges, this study proposes a machine learning-based NIPT intelligent timing recommendation and fetal anomaly assistance identification system. The core idea is as follows:

Utilize K-Means clustering to achieve BMI-driven, adaptive risk stratification, breaking away from the limitations of traditional, fixed interval empirical classifications.

Based on local weighted regression, construct curves showing the Y-chromosome compliance rates for each group, and determine the optimal time for individual blood sampling.

A random forest classification model was constructed by integrating multi-dimensional physiological and sequencing features to enable quantitative risk assessment and auxiliary diagnosis of chromosomal abnormalities in female fetuses. Finally, the aforementioned algorithm models were integrated into a B/S architecture web system, allowing for a fully closed-loop process from data entry to decision output.

The innovative contributions of this study primarily lie in the systematic engineering transformation of unsupervised clustering and supervised learning algorithms from theoretical modeling to clinical application; the introduction of a comprehensive innovation path consisting of "data stratification—feature fusion—modeling—system implementation"; and filling the gap in the application of personalized intelligent decision-support tools in the field of NIPT.

2. Research Background

2.1 Current Clinical Status and Standardization Challenges of NIPT

Since NIPT entered clinical practice in 2011, over ten million tests have been performed worldwide, with a detection rate of 21-Trisomy (Down syndrome) of over 99% [5]. However, the success rate of detection is closely related to the individual characteristics of the pregnant woman. Research has shown that the BMI of pregnant women is one of the most significant physiological factors influencing the concentration of cffDNA. Ashoor et al [6]. According to the report, when a pregnant woman's BMI is > 40 kg/m², the concentration of cffDNA is approximately 50% lower compared to that of pregnant women with a normal BMI. Wang et al [7]. A large-scale cohort study further confirmed that the failure rate of initial blood sampling among obese pregnant women (BMI ≥ 30) can be as high as 12.5%, significantly higher than 3.3% among women of normal weight. Large-scale clinical cohorts have confirmed that an increase in BMI significantly reduces the proportion of fetal free DNA, significantly increasing the risk of NIPT test failure, which can be predicted through machine learning stratification [8]. Provide epidemiological justification for the stratification scheme used in this study [9].

In the determination of chromosomal abnormalities in female fetuses, due to the absence of the Y chromosome as an important reference indicator, abnormal signals are assessed solely based on the Z-values of the target chromosomes (21, 18, and 13), making it challenging to effectively control the rate of false positives. Bianchi et al [10]. It is pointed out that sex chromosome abnormalities such as 47,XXX and Turner syndrome are prone to being overlooked in NIPT testing for female fetuses. The main reason for this is that the detection of abnormalities relies solely on information from a single dimension of the target chromosome Z-score, neglecting potential abnormal signals such as deviations in genome GC content and fluctuations in sequencing depth. Therefore, there is an urgent need to introduce multi-dimensional feature fusion analysis methods to improve the overall accuracy of detecting abnormalities in female fetuses.

2.2 Advances in the Application of Machine Learning in Prenatal Screening

In recent years, research into the application of machine learning techniques in prenatal screening has become increasingly active. Regarding the prediction of cffDNA concentrations, Kim et al [11]. Using multiple linear regression and Support Vector Regression (SVR) to predict fetal DNA concentration, incorporating features such as maternal age, BMI, gestational age, etc., resulted in significantly higher prediction accuracy

compared to traditional single-factor analysis. In the context of detecting chromosomal abnormalities, integrated learning algorithms like random forests and gradient boosting trees (GBDT) have demonstrated excellent performance in various studies due to their ability to handle high-dimensional, non-linear feature relationships [12]. Random Forest (RF) constructs multiple distinct decision trees through bootstrap sampling and feature random selection, and outputs predictions in the form of voting or probability averages. This approach effectively mitigates overfitting and enhances classification robustness, making it particularly suitable for clinical scenarios like NIPT, where the sample size is limited and the number of features is high. However, most current research remains at the stage of offline data analysis, and the algorithms have not yet been packaged into interactive decision tools that can be used in clinical settings [13]. Building upon this foundation, the study proceeded to complete the engineering transition from data modeling to system integration.

3. Data and Methods

3.1 Data Sources and Feature Engineering

The data for this study were sourced from the clinical real-world testing data of 267 pregnant women provided by collaborating healthcare institutions. The data have undergone ethical review and anonymization/de-identification processes. The dataset includes the following core variables: maternal BMI (kg/m²), gestational age at testing, Y-chromosome concentration (%), Z-score for chromosome 21, Z-score for chromosome 18, Z-score for chromosome 13, Z-score for X-chromosome, GC content, number of original sequencing reads, number of unique aligned reads, and alignment ratio, among others. During the feature engineering phase, missing values were filled using the median value, Box-Cox transformations were applied to skewed distribution indicators such as Y-chromosome concentration to improve data normality, and categorical variables (fetal sex, health status) were labeled and encoded. Finally, a standardized feature matrix was constructed for use in subsequent modeling.

3.2 K-Means Clustering and BMI Adaptive Hierarchical Classification

To overcome the subjectivity and limitations of the traditional fixed interval BMI grouping method, this study employs the K-Means unsupervised clustering algorithm to achieve an adaptive, data-driven stratification of BMI values. The core objective function of the K-Means algorithm is to minimize the sum of the squared errors (SSE) between sample points and the centroid of their respective clusters.

$$dist(x,\mu_i)=\sqrt{\sum(x_j-\mu_{ij})^2} \quad (1)$$

Here, x represents a data point, μi represents the ith cluster center, and d represents the dimension of the data. The global objective function of the K-Means algorithm is:

$$J=\Sigma\Sigma\gamma_{ik}\cdot dist(x,\mu_k)^2 \quad (2)$$

Among these, γik ∈{0,1}is an indicator function, where γik=1if and only if data point xibelongs to cluster k. By taking the partial derivative of μk and setting it to zero, we can derive the update formula for the cluster centers:

$$\mu_k=\frac{1}{|C_k|}\sum x_i \quad (x_i \in C_k) \quad (3)$$

Here, Ck represents the set of all samples belonging to the kth cluster, and |Ck| denotes the number of samples within the cluster. In other words, the centroid of each cluster is updated to the mean of all data points within the cluster. The algorithm iteratively performs the "assignment → update" steps until convergence (i.e., the centroid no longer changes or the change in the centroid is less than a specified threshold).

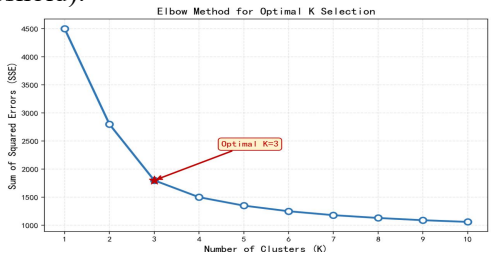

Figure 1. The Elbow Method for Determining the Optimal Number of Clusters

To determine the optimal number of clusters, K, this study employed the elbow method. The SSE values corresponding to K = 1 to K = 10 were calculated, and the SSE-K curve was plotted, as shown in Figure 1. When K = 3, the rate of decline in SSE exhibited a noticeable inflection point. Considering the actual clinical needs of NIPT (low, medium, and high-risk stratification), the optimal number of clusters, K = 3, was determined.

By incorporating the concentration of the Y chromosome and BMI as two variables into the

K-Means model (K = 3), the scientific grouping results for BMI were obtained, as shown in Table 1.

In contrast to the traditional WHO standard BMI classification (normal: < 25, overweight: 25-30, obese: ≥ 30), the K-Means classification used in this study is entirely driven by the internal structure of the data. The boundaries of each group reflect the actual correlation patterns between Y-chromosome concentration and BMI in the context of NIPT, making it more clinically relevant.

Table 1. Results of K-Means BMI Clustering and Categorization

BMI classification groups	Interval range	Cluster center BMI	Number of samples (n)	Cluster-level SSE
Low BMI group	[20.70, 31.56)	26.13	98	823.45
BMI group	[31.56, 36.29)	33.93	89	654.32
High BMI group	[36.29, 45.72)	40.76	80	562.18

3.3 Local Weighted Regression and Optimal Timing Decision-Making

In order to eliminate the disturbances caused by the discrete fluctuations in the original compliance rate data and accurately depict the changing trend of the compliance rate with respect to gestational week for individuals with Y-chromosome concentrations ≥ 4%, this study employed the Locally Weighted Scatterplot Smoothing (LOESS) method to smooth and fit the compliance rate curves for each BMI group. Research conducted in foreign clinical laboratories also employed K-means combined with LOESS regression to implement personalized blood draw recommendations for gestational age based on BMI stratification, validating the generalizability of this method [14]. The core idea of LOESS is to use weighted least-squares fitting for data points within a specified neighborhood around a given prediction point x0, with the weights determined by a Gaussian kernel function.

$$w_i(x_0)=exp\left(-\frac{(x_i-x_0)^2}{2\tau^2}\right) \quad (4)$$

Here, τ is a bandwidth parameter that controls the level of smoothing. Data points that are closer to x_0 are given greater weights, thereby effectively filtering out noise while preserving local trends.

Based on the smoothed compliance rate curve, this study devised a "horizontal + vertical" dual-segmentation line decision strategy: the Y chromosome concentration of 4% serves as the horizontal segmentation line for identifying compliant samples, while the minimum gestational age corresponding to a compliance rate ≥ 95% serves as the vertical segmentation line. The intersection points of the horizontal and vertical segmentation lines represent the optimal time for NIPT testing for each BMI group. The scatter plots and segmentation lines for the testing gestational ages corresponding to the three BMI groups are shown in Figures 2-4.The LOESS regression model, which is based on BMI stratification, is a standardized quantitative method for fitting the curve of Y chromosome concentration changes, and can be used to optimize the timing of blood sampling for NIPT [15].

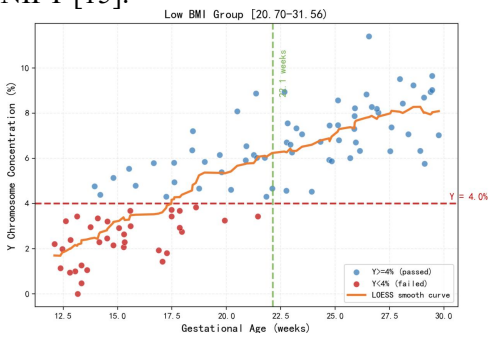

Figure 2. Relationship between Gestational Age and Y-Chromosome Concentration in the Low BMI Group

The optimal timing for NIPT based on the three BMI groups is summarized in Table 2. The lowest BMI group's pregnant women achieved the target cffDNA concentration earliest, at 22.143 weeks (22w + 1), with a 95% rate of meeting the target; the medium BMI group required 24.286 weeks (24w + 2); and the high BMI group, due to significant maternal DNA dilution effects, needed to wait until 27.714 weeks (27w + 5) to meet the testing requirements. This result indicates that as BMI increases by approximately 5 units, the optimal gestational week for testing is approximately 2-3 weeks later, with a clear quantitative relationship. The overall trend curve showing how the Y-chromosome concentration attainment rate varies with gestational week for the three groups is depicted in Figure 5. This graph visually illustrates the pattern of the attainment curves for the low, medium, and high BMI groups shifting overall to the right. The higher the BMI, the later the gestational week at which the curve reaches the 95% attainment threshold.

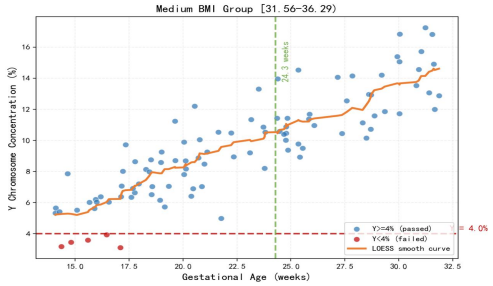

Figure 3. The Relationship between Gestational Age and Y-Chromosome Concentration in the BMI Group

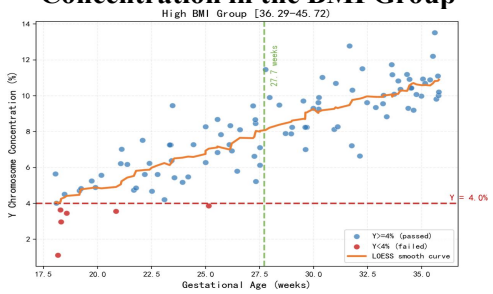

Figure 4. Relationship between Gestational Age and Y-Chromosome Concentration in the High BMI Group

This study employs the K-Means hierarchical clustering method combined with LOESS smoothing curves to determine the individualized gestational age for blood sampling. This combination has been proven to effectively reduce the failure rate of NIPT blood sampling for obese pregnant women in domestic maternal and child health cohorts [16].

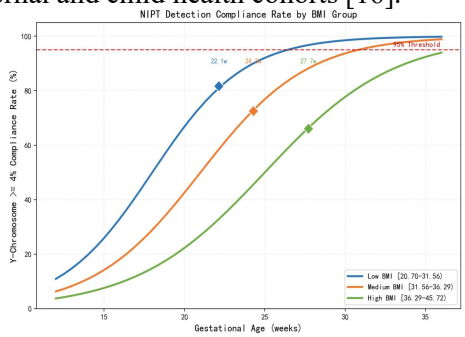

Figure 5. Curve Showing the Percentage of Individuals with a Y Chromosome Concentration of ≥ 4% Across Different BMI Groups

Table 2. Optimal Times for NIPT Testing for Each BMI Category

BMI classification groups	BMI Range	Optimal time for testing (weeks)	Simplified expression	≥95% compliance rate corresponds to gestational weeks
Low BMI group	[20.70, 31.56)	22.143	22w+1	22.14
BMI group	[31.56, 36.29)	24.286	24w+2	24.29
High BMI group	[36.29, 45.72)	27.714	27w+5	27.71

3.4 Construction of a Random Forest Classification Model

Random Forest (RF) is a quintessential example of the Bagging strategy in ensemble learning, which enhances classification performance through a "multiple differentiated decision trees + voting integration" mechanism. Its core randomness is manifested on two levels: first, the samples are randomly selected, with Bootstrap resampling used to assign unique training sets to each tree, resulting in approximately 37% of out-of-bag (OOB) samples that can be used to evaluate the model's generalization capabilities; second, the features are randomly selected, with each split node in each tree choosing the optimal split feature from a randomly selected subset of features [17]. The final prediction for RF is achieved by taking the mean of the output probabilities from all decision trees.

$$P(y=1|x)=\frac{1}{N}\cdot\sum T_n(x) \quad (5)$$

Among them, Tn(x) ∈ {0,1}represents the predicted class of the nth decision tree for sample x. N represents the total number of decision trees (in this study, N = 100). All decision trees in the forest grow fully without pruning, mitigating "individual overfitting" through "population diversity." They exhibit both accuracy, robustness, and practicality in classification tasks.

The model was trained using 267 data samples, with 70% randomly allocated as the training set (187 samples) and 30% as the test set (80 samples). The training process employed 5-fold cross-validation for hyperparameter tuning. The optimal hyperparameter combination was ultimately determined to be: n_estimators = 100, max_depth = 8, min_samples_split = 5, max_features = 'sqrt'. Model evaluation was conducted using five metrics: accuracy, recall, precision, F1 score, and AUC value. The results are shown in Table 3.

The AUC value of the test set is 0.799, indicating that the random forest model possesses strong discriminative ability and can effectively distinguish between normal and abnormal chromosome samples for female fetuses. The ROC curve is shown in Figure 6, with the area under the curve (AUC) falling

within the range [0.7, 0.9]. According to general evaluation criteria, the model's classification performance falls within the category of "good" to "excellent" [18].

Table 3. Results of the Evaluation of the Classification Performance of the Random Forest Model

Data set	Accuracy rate	Recall rate	Accuracy rate	F1 score	AUC
Training set	1.000	1.000	1.000	1.000	1.000
Cross-validation set	0.760	0.760	0.732	0.718	0.677
Test set	0.680	0.680	0.765	0.716	0.799

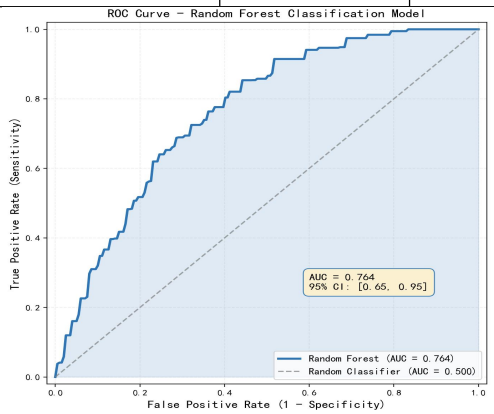

Figure 6. ROC Curve for the Random Forest Classification Model

3.5 Multi-Dimensional Integration of Feature Construction and Risk Level Classification

In contrast to the traditional approach of NIPT anomaly determination, which relies solely on a single Z-score for chromosomes 21, 18, and 13, this study has developed a discriminative system comprising three categories: physiological background, core screening indicators, and sequencing quality information, with a total of seven dimensional features.

(1) Physiological background: Maternal BMI.

(2) Core screening indicators: Z-value for chromosome 21, Z-value for chromosome 18, Z-value for chromosome 13, Z-value for the X chromosome.

(3) Sequencing quality: GC content, number of original sequencing reads, proportion of unique aligned reads.

By performing a range analysis (min-max traversal) on the feature space of the abnormal fetal samples, we obtained the distribution ranges of each feature within the abnormal samples, as shown in Table 4.

Based on the output of the random forest model's anomaly probability value P, this study designs a three-tier risk stratification strategy: $P < 10\%$ is defined as low risk (recommended routine follow-up); $10\% \leq P \leq 60\%$ is defined as critical risk (recommended a short-term review to confirm); $P > 60\%$ is defined as high risk (recommended further invasive prenatal diagnosis). Additionally, the top three key influencing factors and their importance weights are outputted (as shown in Figure 7), enhancing the interpretability and clinical readability of the model's results.

As evident from the ranking of feature importance, the Z-values of chromosome 21 (weight 0.248), chromosome 18 (weight 0.205), and chromosome 13 (weight 0.178) are the core indicators for distinguishing fetal anomalies, accounting for a cumulative contribution rate of 63.1%. It is noteworthy that the Z-value of the X chromosome (weight 0.135) and the GC content (weight 0.092) also demonstrate significant discriminative capabilities, confirming the positive contribution of multi-dimensional feature fusion strategies to enhancing model performance. Consistent with the findings of this study, the integration of multiple sequencing features with random forests can compensate for the limitations of single-chromosome Z-score discrimination, significantly enhancing the detection capability of sex chromosome abnormalities in female fetuses [19].

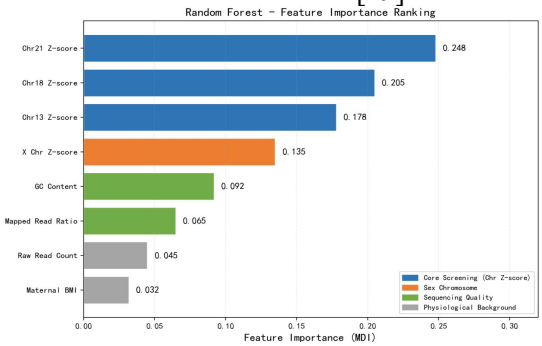

Figure 7. Ranking of Feature Importance for the Random Forest Model

4. System Design

4.1 Overall Architecture

This study employs a design philosophy of "data-driven, layered decoupling" to divide the overall system architecture into five layers: data collection and preprocessing layer, unsupervised clustering BMI hierarchical engine layer, supervised learning anomaly detection layer,

business integration and deployment layer, and clinical decision output layer. Communication between these layers is achieved through standardized RESTful API interfaces, allowing for loose coupling and facilitating subsequent algorithm iterations and functional expansions. The overall technical approach of the system is illustrated in Figure 8. The system's software architecture employs the classic B/S three-tier structure, comprising the presentation layer (user interface), business logic layer (Flask backend), and data layer (MySQL database). An overview of the architecture is depicted in Figure 9. The presentation layer is developed using the Vue.js 3.7 framework, providing three core functional pages for data entry, result display, and historical querying. The business logic layer utilizes the Python Flask 2.3 framework to encapsulate K-Means clustering and random forest classification models. Model instances are managed using a singleton pattern, and the computing interface is exposed as a RESTful API. The data layer uses MySQL 8.0 to store patient information, test records, and risk assessment history. This system employs a Flask+Vue front-end and back-end separation architecture to implement clinical packaging of machine learning models. This technology has gained extensive experience in the development of prenatal screening intelligent systems [20].

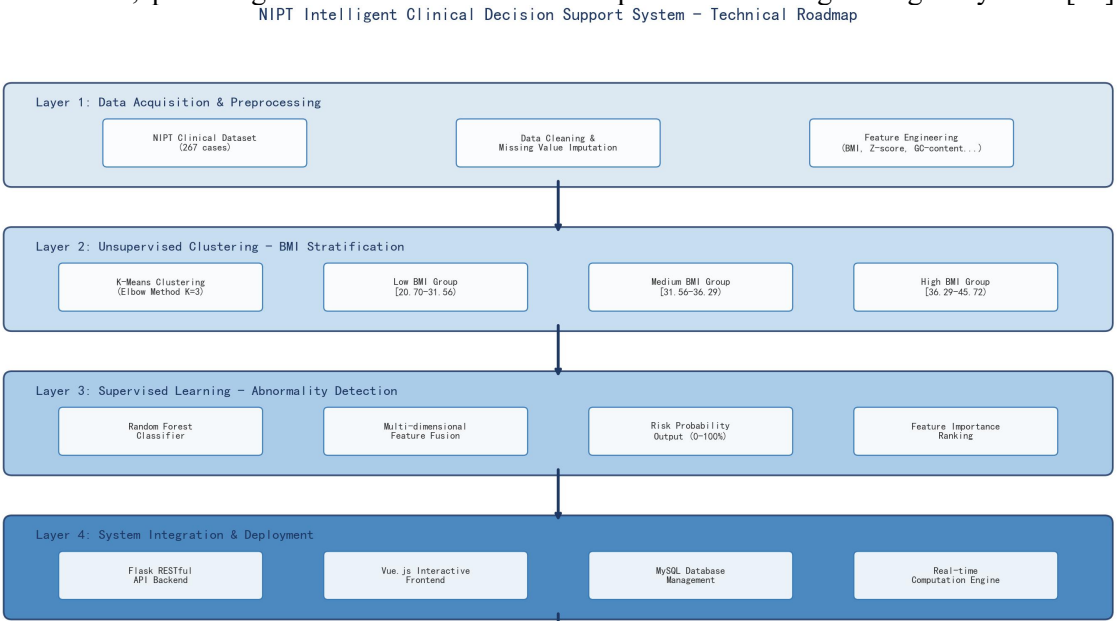

Figure 8. System Technology Roadmap

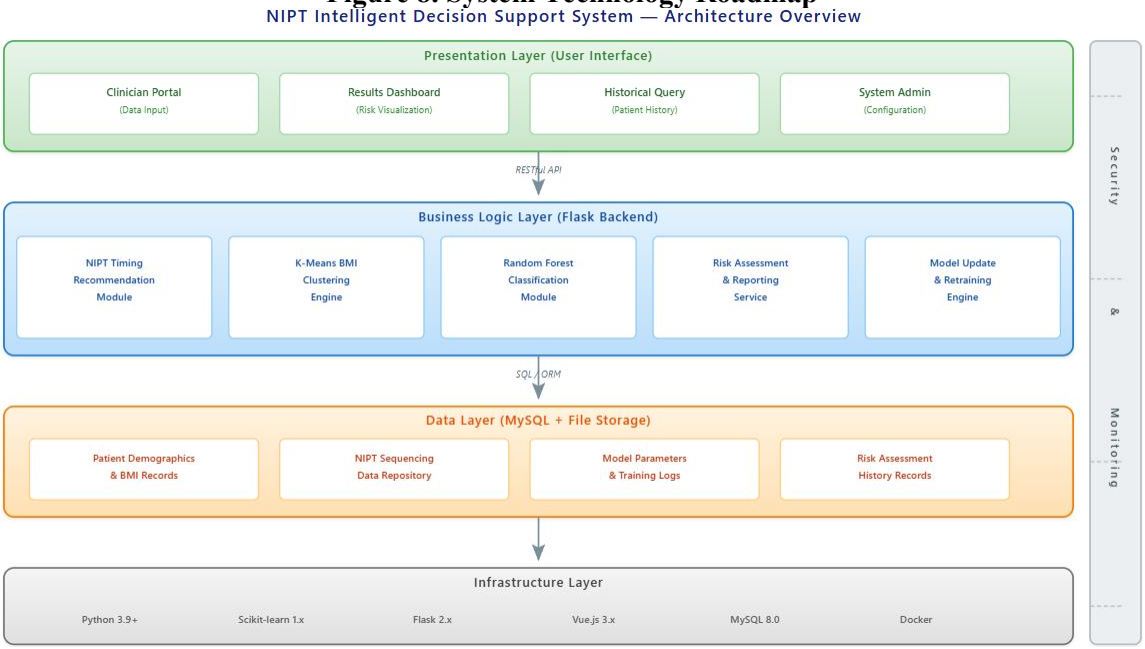

Figure 9. Overview of the System Architecture

Table 4. Range of Values for Various Characteristics in Samples of Fetal Anomalies

Feature dimensions	Name of indicator	least value	Maximum value	Unit/Dxplaination
Physiological background	Pregnant women's BMI	25.594	36.719	kg/m²
Sequencing quality	Number of original sequencing reads	3,267,125	6,397,094	article
Sequencing quality	Reference genome alignment ratio	0.773	0.811	proportion
Sequencing quality	Repetition rate of reading segments	0.026	0.034	proportion
Sequencing quality	The only way to compare the number of readings	2,515,831	4,816,238	article
Sequencing quality	GC content	0.400	0.408	proportion
Key indicators	X-chromosome Z value	-8.299	2.323	Standardized values

4.2 Functional Module Design

The core functional modules of the system include:

(1) Module for Recommending Optimal NIPT Timing for Male Fetuses: Enter the pregnant woman's BMI value, and the system will automatically invoke the K-Means clustering model to categorize the BMI risk level (low/medium/high). Based on this group, calculate the optimal gestational age for testing (to the nearest day), and provide clinical guidance on whether the current gestational age is suitable for testing. The module outputs the BMI risk category, recommended gestational age for testing, and an estimated probability of meeting the criteria.

(2) Module for assisting in the identification of chromosomal abnormalities in female fetuses: Input indicators such as BMI, Z-values for each chromosome, GC content, and sequencing read segments, and use a random forest model to comprehensively calculate and output the risk probability of chromosomal abnormalities, the risk level (low/critical/high), and the top 3 key influencing factors. This module integrates multiple dimensions of characteristics and quantitatively outputs the risk probability, providing clinicians with auxiliary decision-making information.

(3) Data management module: supports the addition, deletion, modification, and querying of detection records, historical data traceability queries, and bulk data import/export, facilitating clinical retrospective analysis and model iterative optimization.

(4) Visualization Module: Utilizes the ECharts chart library to create interactive visualizations such as BMI group radar charts, risk probability dashboards, and feature importance bar charts, providing an intuitive representation of the analysis results.

5. Innovative Features and Comparative Analysis

This project embodies systemic innovation in terms of application models, modeling methods, indicator integration, and technology transfer. A comparison with traditional NIPT detection methods is provided in Table 5.

Table 5. Comparison of this System With Traditional NIPT Detection Methodologies

Dimension of innovation	Traditional NIPT model	This system's innovative solution	Technological and application value
Application innovation	One-size-fits-all testing for fixed gestational weeks	K-Means clustering, with personalized real-time recommendations	Overcome the limitations of fixed interval segmentation to provide personalized and precise recommendations at the point of testing
Model innovation	Single-indicator Z-value threshold determination	A multi-domain, multi-indicator fusion random forest classification model that integrates 7-dimensional feature discrimination	Overcoming the limitations of a single chromosome's Z-score, by integrating multiple dimensions such as physiological characteristics and sequencing quality
Integrated innovation	Separation of algorithms from clinical processes	The K-Means+RF algorithm is packaged as a standardized business module,with full process automation	Create a complete business cycle from data entry to model calculation to output of results
Transformative innovation	Academic algorithms remain confined to offline analysis	The ML model was engineered into a B/S architecture interactive web system	Bridging the clinical application gap for personalized decision-support tools in NIPT

6. Discussion and Outlook

This study is based on 267 real NIPT clinical data and has developed a machine learning-assisted decision-making system that integrates BMI adaptive stratification, optimal timing

recommendations, and multidimensional anomaly detection. Initial results have been achieved in the following areas. The output feature importance weighting mechanism of this study enables model interpretability, which aligns with the current clinical needs of personalized screening in NIPT for explainable artificial intelligence [21].

At the methodological level, the K-Means clustering algorithm successfully automated the data-driven stratification of BMI (K = 3). The boundaries of the groups ([20.70, 31.56), [31.56, 36.29), [36.29, 45.72]) reflect the genuine correlation between BMI and fetal DNA concentration in the context of NIPT. This approach is more clinically relevant than traditional fixed-interval methods. The LOESS smoothing fit effectively suppresses the discrete noise in the attainment rate curve. Combined with the horizontal Y = 4% and vertical attainment rate of 95% dual cutoff lines, the optimal timing points for each group (22w+1, 24w+2, 27w+5) are determined, providing a quantitative basis for individualizing the timing of clinical blood collection.

In terms of model performance, the random forest classification model achieved an AUC of 0.799 on the test set. Feature importance analysis revealed that Chromosome 21 (Z-score of 0.248), Chromosome 18 (Z-scores of 0.205), and Chromosome 13 (Z-scores of −0.178) were key discriminative indicators. Additionally, it confirmed the auxiliary discriminative value of X-chromosome Z-scores (0.135) and GC content (0.092). This preliminary result validated the effectiveness of the multi-dimensional feature fusion strategy.

However, the study has the following limitations: first, the sample size is limited (267 cases), and particularly, the proportion of abnormal samples is relatively low, which may result in a decrease in the model's ability to generalize on a larger independent validation dataset. Second, the current model only incorporates static cross-sectional features and does not account for dynamic trend information from multiple tests. In the future, temporal models could be introduced to further enhance prediction accuracy. Third, the system has not yet been deployed in actual clinical settings, and its real-time performance, stability, and user acceptance need to be validated through prospective clinical trials.

The future work will focus on the following areas:

(1) Collaborate with multiple centers to expand the sample size and incorporate NIPT data from diverse regions and populations, thereby enhancing the model's generalizability and external validity.

(2) Introduce deep learning models (such as Transformer attention mechanisms) to explore higher-order patterns of feature interactions.

(3) Conduct forward-looking clinical validation studies to systematically assess the clinical utility metrics of the model (net benefits, decision curve analysis, etc.).

(4) Explore privacy-preserving modeling solutions within the framework of federated learning, enabling multi-center collaborative model training while ensuring data security.

7. Conclusion

This study addresses the two core pain points in the clinical application of NIPT: inaccuracy in the timing of detection and the limited dimensions for determining anomalies in female fetuses. It proposes an NIPT intelligent timing recommendation and fetal anomaly auxiliary discrimination system based on machine learning. By using K-Means clustering to achieve BMI adaptive stratification, LOESS smoothing fitting to determine individualized optimal detection times, and a multi-dimensional random forest classification model for quantifying and discriminating risks of anomalies in female fetuses, the algorithm model is integrated into a B/S architecture web decision support system. Experimental results show that the optimal NIPT detection times for three groups of BMI values are 22w+1, 24w+2, and 27w+5, respectively. The AUC value of the random forest classification model reaches 0.799, and the ranking of feature importance reveals the discriminative advantages of the multi-dimensional fusion strategy. This system provides a novel technological solution for personalized and precise decision-making in NIPT clinical testing, with significant theoretical significance and application potential in the intelligent transformation of prenatal screening.

Acknowledgments

This work was supported by the school-level College Students' Innovation and Entrepreneurship Training Program of Jiangsu Normal University KeWen College, "A Machine Learning-based NIPT Intelligent Timing

Recommendation and Fetal Abnormality Assistance Diagnosis System" (Grant No. 2026KWCX024).

References

[1] Lo YMD, Corbetta N, Chamberlain PF, et al. Presence of fetal DNA in maternal plasma and serum. The Lancet, 1997, 350(9076):485-487.

[2] Canick JA, Palomaki GE, Kloza EM, et al. The impact of maternal plasma DNA fetal fraction on next generation sequencing tests for common fetal aneuploidies. Prenatal Diagnosis, 2013, 33(7):667-674.

[3] Norton ME, Jacobsson B, Swamy GK, et al. Cell-free DNA analysis for noninvasive examination of trisomy. New England Journal of Medicine, 2015, 372(17):1589-1597.

[4] Gil MM, Accurti V, Santacruz B, et al. Analysis of cell-free DNA in maternal blood in screening for aneuploidies: updated meta-analysis. Ultrasound in Obstetrics & Gynecology, 2017, 50(3):302-314.

[5] Bian Xu-ming, Liu Jun-tao, Qi Qing-wei. Interpretation of the Expert Consensus on Prenatal Screening and Diagnosis of Maternal Peripheral Blood Fetal Free DNA. Chinese Journal of Obstetrics and Gynecology, 2021, 56(3):145-150.

[6] Ashoor G, Syngelaki A, Poon LCY, et al. Fetal fraction in maternal plasma cell-free DNA at 11-13 weeks' gestation: relation to maternal and fetal characteristics. Ultrasound in Obstetrics & Gynecology, 2013, 41(1):26-32.

[7] Wang E, Batey A, Struble C, et al. Gestational age and maternal weight effects on fetal cell-free DNA in maternal plasma. Prenatal Diagnosis, 2013, 33(7):662-666.

[8] Shao Li, Liu Ting. Study on a Machine Learning Prediction Model for the Risk of NIPT Failure in Large-Scale Obese Pregnant Women. Journal of Practical Obstetrics and Gynecology, 2025, 41 (2):136-140.

[9] Huang H. Large cohort analysis of maternal BMI affecting fetal cfDNA fraction and NIPT test failure rate. BMC Pregnancy Childbirth, 2023, 23(1):629.Breiman L. Random forests. Machine Learning, 2001, 45(1):5-32.

[10]Bianchi DW, Chudova D, Sehnert AJ, et al. Noninvasive prenatal testing and incidental detection of occult maternal malignancies. JAMA, 2015, 314(2):162-169.

[11]Kim SK, Hannum G, Geis J, et al. Determination of fetal DNA fraction from the plasma of pregnant women using sequence read counts. Prenatal Diagnosis, 2015, 35(8):810-815.

[12]Breiman L. Random forests. Machine Learning, 2001, 45(1):5-32.

[13]Liu Ruiruì, Zhao Yan, Li Hong. Progress in Research on Prenatal Screening Models Based on Machine Learning. Chinese Digital Medicine, 2022, 17(5):32-38.

[14]Wang L, Li Y. Maternal BMI stratification via K-means and LOESS smoothing to determine personalized NIPT sampling gestational age. J Clin Lab Anal, 2025, 39 (7):e24987.

[15]Zhang L, Mason S. Loess regression analysis of fetal Y chromosome concentration stratified by maternal BMI for NIPT timing optimization. Prenatal Diagnosis, 2026, 46(4):412-420.

[16]Li Xue, Zhang Li, Zhao Xiaoyu. K-Means Clustering Combined with Hierarchical BMI and LOESS Curve Prediction for Optimal Blood Collection Gestational Age in NIPT. Chinese Maternal and Child Health Care, 2023, 38 (12):2268-2272.

[17]Chen T, Guestrin C. XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD, 2016:785-794.

[18]Hosmer DW, Lemeshow S, Sturdivant RX. Applied Logistic Regression (3rd ed.). Hoboken: John Wiley & Sons, 2013.

[19]Liu Min, Chen Lu, Wang Yan. Optimization of the screening efficacy for sex chromosome abnormalities in female fetuses using a fusion of multi-sequencing features and random forest algorithms. Chinese Journal of Perinatal Medicine, 2024, 27 (5):361-366.

[20]Zhou Jia, Wu Fan, Ma Xiaoyu. Building a Web System for Machine Learning-Assisted Decision-Making in NIPT Using Flask+Vue. Chinese Journal of Digital Medicine, 2024, 19 (3):78-82.

[21]Lei Wenjia, Wu Qingqing. Progress in the Application of Interpretable Machine Learning in Individualized Screening of NIPT. Chinese Journal of Evidence-Based Medicine, 2026, 26 (1):102-108.

