A Study on Optimal NIPT Timing and Fetal Anomaly Detection Based on Multi-Model Approach

Haiming Cui *, He Chen, Yuefangxi Chen

Shenyang Ligong University School of Science, Shenyang Ligong University, Shenyang, China, 110159

* Corresponding Author Email: c1617189812@163.com

Abstract. The accuracy of Non-Invasive Prenatal Testing (NIPT) critically depends on selecting the optimal testing time point, which is influenced by maternal factors such as gestational week and Body Mass Index (BMI). This paper develops a comprehensive modeling framework to address two key challenges: optimizing NIPT timing for male fetuses and accurately detecting chromosomal abnormalities in female fetuses. A quadratic polynomial regression model is established to quantify the relationship between Y-chromosome concentration, gestational week, and BMI. To determine the optimal NIPT time point, K-Means clustering is integrated with a logistic regression-based risk prediction model, which minimizes a composite risk function. For the classification of fetal anomalies in females, a fusion model combining Random Forest and Logistic Regression is proposed. The results demonstrate that Y-chromosome concentration exhibits a weak but statistically significant positive correlation with gestational week and a negative correlation with BMI. The risk optimization model identifies distinct optimal NIPT timings for different maternal BMI and age groups. The fusion classification model achieves an accuracy of 0.9505, a precision of 0.8235, a recall of 0.7000, and an F1-score of 0.7570, with an optimal fusion weight of 0.67 and a decision threshold of 0.4459. This study provides a robust theoretical framework and actionable recommendations for personalized NIPT scheduling and diagnosis in clinical practice.

Keywords: NIPT, Multi-Model Approach, Optimal Timing, Fetal Anomaly Detection.

1. Introduction

Non-Invasive Prenatal Testing (NIPT) has revolutionized prenatal screening by analyzing cell-free fetal DNA (cffDNA) in maternal peripheral blood to detect common chromosomal aneuploidies such as Down syndrome [1-3]. The reliability of NIPT is contingent upon the fetal fraction of cffDNA, which must reach a sufficient threshold for an accurate diagnosis—typically a Y-chromosome concentration of at least 4% for male fetuses. Maternal factors, including gestational week, Body Mass Index (BMI), and age, significantly influence the time at which this threshold is met [4-6]. An improperly timed test can lead to false negatives or missed opportunities for clinical intervention, posing health risks to both the mother and fetus.

Prior studies have laid foundational work in this domain. For instance, Canick et al. (2013) proposed early models for estimating fetal fraction based on gestational age and maternal weight, highlighting the inverse relationship with BMI but without incorporating non-linear dynamics or optimization for timing [7,8]. Similarly, Hudecova et al. (2015) developed logistic models to predict NIPT failure risks, focusing primarily on sequencing metrics, yet they overlooked subgroup stratification and composite risk minimization [9]. In anomaly detection for female fetuses, Bianchi et al. (2012) introduced Z-score-based thresholds for chromosomal analysis, but these approaches suffered from limited feature integration and class imbalance issues, resulting in suboptimal recall rates [10,11]. While these works advanced the field, they generally adopted a one-size-fits-all strategy for timing and relied on single-model classifiers, failing to address personalized stratification or multi-model fusion for enhanced accuracy.

Therefore, a critical challenge in clinical practice is to determine the optimal NIPT time point for individual patients by stratifying them based on relevant characteristics. Furthermore, the accurate classification of fetal anomalies, particularly for female fetuses where Y-chromosome markers are absent, requires sophisticated modeling of multiple Z-score values and sequencing metrics.

This paper addresses these challenges by developing a multi-stage mathematical modeling framework using real-world clinical data. The objectives are threefold: 1) to model the relationship between fetal Y-chromosome concentration and key maternal variables; 2) to establish an optimization model that determines the best NIPT timing for stratified groups of pregnant women to minimize potential risks; and 3) to construct a high-performance classification model for detecting fetal chromosomal abnormalities in females.

The significance of this study lies in its potential to enhance clinical outcomes by enabling personalized NIPT protocols, reducing failure rates, and improving early detection of anomalies, which could lower maternal anxiety and healthcare costs. Key innovations include: (1) the integration of quadratic polynomial regression with K-Means clustering for non-linear, subgroup-specific timing optimization via a novel composite risk function; (2) a weighted fusion of Random Forest and Logistic Regression models, optimized through grid search and Youden's Index, to achieve superior performance in imbalanced datasets; and (3) sensitivity analyses that provide practical insights into threshold variations, bridging the gap between theoretical modeling and clinical applicability.

2. Methodology

2.1. Data Preprocessing

To ensure data quality and consistency, a systematic preprocessing pipeline was implemented. Records with missing values for critical features or with gestational weeks outside the standard clinical range (10-25 weeks) were removed. Samples were flagged as sequencing failures and excluded if the Y-chromosome concentration was below 4%. Gestational week data was standardized to a numerical format in "weeks." BMI values were recalculated from height and weight, and original entries were corrected if the discrepancy exceeded a tolerance of ε=0.001. Finally, the Interquartile Range (IQR) method was used to identify and winsorize outliers for key variables to enhance model robustness.

2.2. Modeling Y-Chromosome Concentration

2.2.1 Normality and Correlation Analysis

The Shapiro-Wilk test was first applied to assess the normality of all continuous variables. As the data predominantly exhibited non-normal distributions, the non-parametric Spearman's rank correlation coefficient was employed to quantify the monotonic relationship between Y-chromosome concentration and other covariates, such as gestational week and BMI.

2.2.2 Quadratic Polynomial Regression

Based on the correlation analysis, a quadratic polynomial regression model was established to capture the non-linear effects of gestational week (X_1) and BMI (X_2) on Y-chromosome concentration (Y). The model, including interaction terms, is defined as:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1^2 + \beta_4 X_2^2 + \beta_5 X_1 X_2 + \varepsilon \quad (1)$$

Parameters (β) were estimated using the Ordinary Least Squares (OLS) method. The model's overall and individual variable significance were validated using the Chi-squared test and adjusted residual analysis.

2.3. Optimal NIPT Timing for Male Fetuses

2.3.1 Maternal Subgroup Stratification

To tailor NIPT timing, pregnant women were stratified into distinct subgroups. K-Means clustering was applied to maternal BMI and a combination of BMI and age. The optimal number of clusters (K=5) was determined using the silhouette coefficient method, ensuring high intra-cluster similarity and low inter-cluster similarity.

2.3.2 Composite Risk Function

This study defined a total risk function, $\mathrm{R\_total(t, G)}$, to be minimized, where *t* is the testing time point and *G* is the maternal subgroup. This function comprises two components:

Temporal Risk ($\mathrm{R\_time(t)}$): A piecewise linear function reflecting the increasing clinical risk associated with a delayed diagnosis. It is assigned low weight in early gestation, linearly increasing weight mid-gestation, and high weight in late gestation.

Failure Risk ($\mathrm{R\_fail(t, G)}$): The probability of NIPT failure at time *t* for group *G*, defined as the probability that the Y-chromosome concentration does not meet the 4% threshold.

The total risk is a weighted sum:

$$R_{total}(t, G) = R_{time}(t) + \lambda \cdot R_{fail}(t, G) \tag{2}$$

where λ is a weighting coefficient balancing the two risk types.

2.3.3 Failure Risk Prediction using Logistic Regression

To calculate R_fail, a logistic regression model was trained to predict the probability (p) of achieving the target Y-chromosome concentration. A binary outcome variable was created (1 if concentration ≥ 4%, 0 otherwise). The model predicts the log-odds as a linear function of gestational week (t), BMI, and age:

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 t + \beta_2 \mathrm{BMI} + \beta_3 \mathrm{Age} \tag{3}$$

The failure risk is then $R\_fail = 1 - p$.

2.3.4 Optimization and Sensitivity Analysis

For each maternal subgroup, the optimal NIPT time point (t**) is found by iterating through possible gestational weeks and identifying the t that minimizes $R\_total(t, G)$. For the advanced model, an additional constraint was added: the average success probability at t** for the group must be at least 60%. To analyze the impact of detection error, a sensitivity analysis was performed by redefining the success threshold to 3.5% and re-calculating the optimal time points.

2.4. Fetal Anomaly Classification for Female Fetuses

2.4.1 Feature Engineering and Selection

For the binary classification task (abnormal vs. normal), Z-scores for chromosomes 13, 18, 21, and X were standardized based on the mean and standard deviation of the normal (y=0) population. To create a concise and powerful feature set, feature importance was evaluated using both Logistic Regression (based on coefficient magnitudes) and Random Forest (based on Gini impurity reduction). The final feature set was the union of the top features selected by both methods.

2.4.2 Fusion Model Development

A fusion model was developed to leverage the predictive power of Random Forest (RF) and the interpretability of Logistic Regression (LR).

(1) Base Models: An L2-regularized LR model and an RF model were trained independently [5].

(2) Probability Fusion: To mitigate overfitting, predicted probabilities from both models were generated on a training set using 5-fold stratified cross-validation. These probabilities (P_{LR} and P_{RF}) were then linearly combined:

$$P_{fused} = \alpha \cdot P_{LR} + (1 - \alpha) \cdot P_{RF} \tag{4}$$

(3) The optimal fusion weight α was determined via a grid search (step size 0.01) to maximize the F1-score on the validation sets.

(4) Optimal Threshold Determination: The final classification decision depends on comparing P_{fused} to a threshold θ. The optimal threshold θ* was identified by maximizing the Youden's Index (J = TPR - FPR) on the Receiver Operating Characteristic (ROC) curve.

2.5. Model Evaluation

The performance of the classification model was assessed using Accuracy, Precision, Recall, F1-score, and the Area Under the ROC Curve (AUC). Both macro and weighted averages were calculated to account for class imbalance.

3. Results and Discussion

3.1. Y-Chromosome Concentration Analysis

The Shapiro-Wilk test confirmed that most variables, including Y-chromosome concentration, gestational week, and BMI, were not normally distributed ($p < 0.05$). Spearman correlation analysis revealed a weak but highly significant positive correlation between Y-chromosome concentration and gestational week ($\rho = 0.089$, $p < 0.01$) and a weak but highly significant negative correlation with BMI ($\rho = -0.128$, $p < 0.001$). This confirms that Y-chromosome concentration tends to increase as pregnancy progresses and decrease with higher maternal BMI.

The fitted quadratic polynomial regression model is:

$$Y = -0.078 + 0.0135X_1 - 0.0047X_2 - 0.00027X_1^2 + 0.000039X_2^2 + 0.000136X_1X_2 \quad (5)$$

The model's overall Chi-squared test was significant ($p = 0.012$), and residual analysis confirmed that the assumptions of normality and independence were met, validating the model's explanatory power.

3.2. Optimal NIPT Timing Results

3.2.1 BMI-Based Stratification

For the model considering only BMI, K-Means clustering yielded five distinct groups. The logistic regression model for failure risk yielded coefficients $\alpha_1 = 1.435$ (for week t) and $\alpha_2 = -0.1215$ (for BMI), confirming that success probability increases with time and decreases with BMI. By minimizing the total risk function, the optimal NIPT timings were determined for each group as shown in Table 1.

Table 1. Optimal NIPT Timing by BMI Group

BMI Group	Optimal NIPT Time Point (weeks)
Group 1 (Lowest BMI)	12
Group 2	12
Group 3	14
Group 4	16
Group 5 (Highest BMI)	20

As expected, higher BMI groups require a later testing date to minimize the overall risk of failure and delayed diagnosis.

3.2.2 BMI-Age Based Stratification

When stratifying by both BMI and age, a more granular set of optimal time points was generated. The results, visualized in a heatmap (Figure 1), show a clear trend where higher BMI and advanced maternal age generally necessitate a later NIPT schedule. This refined approach allows for more personalized clinical decision-making.

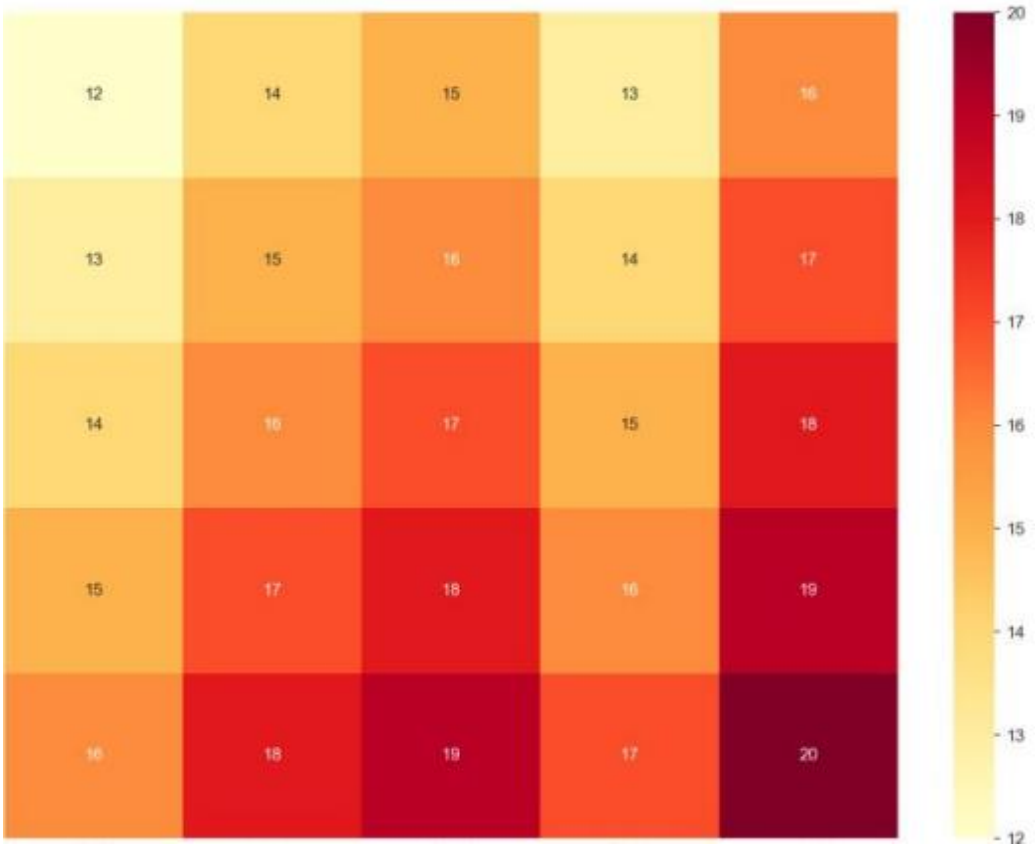

Figure 1. Heatmap of Optimal NIPT Timing by BMI-Age Group

3.2.3 Sensitivity to Detection Threshold

When the Y-chromosome concentration threshold was lowered to 3.5%, the model recommended earlier NIPT timings across all groups, with most converging to 12 weeks. This demonstrates that the optimal timing is highly sensitive to the defined success criteria, highlighting the importance of using a clinically validated threshold.

3.3. Female Fetal Anomaly Detection Performance

After feature selection, ten key features were used to train the fusion model. The grid search identified an optimal fusion weight of $\alpha = 0.67$, and the optimal decision threshold was determined to be $\theta = 0.4459$ using Youden's Index.

The performance of the final fusion model on the test set is summarized in Table 2.

Table 2. Classification Report for the Fusion Model

Class	Precision	Recall	F1-Score	Support
0 (Normal)	0.96	0.98	0.97	162
1 (Abnormal)	0.82	0.70	0.76	20
Accuracy			0.95	182
Weighted Avg	0.95	0.95	0.95	182

The model achieved an overall accuracy of 95.05%. The ROC analysis yielded an AUC of 0.922 for the fusion model, comparable to the LR model (0.927) and superior to the RF model (0.878), indicating excellent discriminative ability (Figure 2).

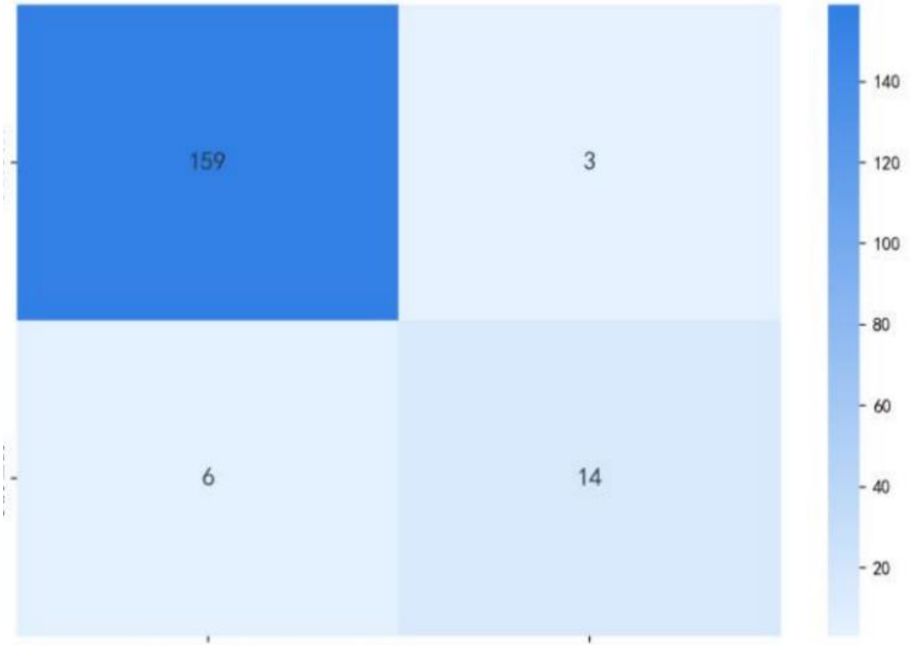

Figure 2. Confusion Matrix for Base and Fusion Models

The model demonstrates high precision and accuracy. While the recall for the minority (abnormal) class is 0.70 due to class imbalance, the F1-score of 0.76 indicates a good balance between precision and recall, validating the effectiveness of the proposed method for clinical screening.

4. Conclusion

This paper successfully developed a systematic, data-driven framework to address critical optimization and classification challenges in Non-Invasive Prenatal Testing. By integrating statistical analysis, machine learning, and risk modeling, this study has provided solutions for two key clinical needs.

First, this study quantified the relationship between fetal Y-chromosome concentration and maternal characteristics and developed a risk-minimization model that provides personalized NIPT timing recommendations based on maternal BMI and age. This approach moves beyond one-size-fits-all scheduling towards individualized care. Second, this study constructed a robust fusion model for the detection of fetal anomalies in females, which achieves high accuracy and balances the trade-offs between precision and recall, making it a valuable tool for clinical diagnosis.

Limitations of this study include the idealized setting of the concentration threshold and the inherent class imbalance in the anomaly detection dataset. Future work could explore dynamic thresholding mechanisms and advanced data augmentation techniques like Generative Adversarial Networks (GANs) or SMOTE to further improve the recall for rare abnormal cases. Nevertheless, the models presented here provide a strong foundation for enhancing the efficacy and personalization of NIPT in clinical practice.

References

[1] James G, Witten D, Hastie T, et al. An Introduction to Statistical Learning: with Applications in R[M]. New York: Springer, 2021.

[2] National Health Commission of the PRC. Technical specifications for non-invasive prenatal screening and diagnosis: WS/T 662-2020[S]. Beijing: Standards Press of China, 2020.

[3] Breiman L. Random Forests[J]. Machine Learning, 2021, 45(1): 5-32.

[4] Hosmer D W, Lemeshow S. Applied Logistic Regression[M]. 3rd ed. New York: John Wiley & Sons, 2023.

[5] Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine Learning in Python[J]. Journal of Machine Learning Research, 2021, 12: 2825-2830.

[6] Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves[J]. BMC Bioinformatics, 2021, 12: 77.

[7] Youden W J. Index for rating diagnostic tests[J]. Cancer, 2023, 3(1): 32-35.

[8] Hartwell K, Pagnamenta E, Stojanovik V, et al. The Power of Partnership: Adapting Early Language Intervention for Children with Down Syndrome Through Family-Researcher Collaboration[J]. International Journal of Language & Communication Disorders, 2025, 60(6): e70139.

[9] Kong X, Zhang L, Yang R, et al. Reasons for failure of noninvasive prenatal test for cell-free fetal DNA in maternal peripheral blood[J]. Molecular Genetics & Genomic Medicine, 2024, 12(1): e2351.

[10] Liao C, Zhengfeng X, Zhang K. DNA Sequencing versus Standard Prenatal Aneuploidy Screening[J]. The New England Journal of Medicine, 2024, 371(6): 577-578.

[11] Yamamoto K, Suzumori N, Miura K, et al. Clinical Implications of Low Cell-Free DNA Fetal Fraction in Non-Invasive Prenatal Testing: A Retrospective Cohort Study of 40,716 Pregnancies[J]. Prenatal Diagnosis, 2025.