Textile & Leather Review

ISSN 2623-6281 | www.tlr-journal.com | 10.31881/TLR

Optimal Timing of Non-Invasive Prenatal Testing: A Personalized Approach Based on Hybrid Models and Multi-Objective Programming

Anping Hu

How to cite: Hu A. Optimal Timing of Non-Invasive Prenatal Testing: A Personalized Approach Based on Hybrid Models and Multi-Objective Programming. Textile & Leather Review. 2026; 9:3203-3221. https://doi.org/10.31881/TLR.2026.3203

How to link: https://doi.org/10.31881/TLR.2026.3203

Published: 25 April 2026

Optimal Timing of Non-Invasive Prenatal Testing: A Personalized Approach Based on Hybrid Models and Multi-Objective Programming

Anping Hu

School of Information and Intelligent Science and Technology, Hunan Agricultural University, Changsha, 410128, China.
anpinghu2@gmail.com

Article
https://doi.org/10.31881/TLR.2026.3203
Published 25 April 2026

ABSTRACT

The accuracy of non-invasive prenatal testing (NIPT) is highly dependent on the concentration of fetal cell-free DNA in maternal plasma, with Y chromosome concentration serving as a key indicator for assessing male fetus samples. To address the clinical risks arising from individual differences among pregnant women overlooked by current standardized protocols, this study developed a data-driven framework for optimizing personalized testing timing. First, a hybrid GAM-RF predictive model integrating a generalized additive model (GAM) with random forest (RF) was constructed to accurately capture the complex nonlinear relationships among multiple factors, including maternal BMI, gestational age, maternal age, and Y chromosome concentration. The model demonstrated superior performance on the test set, achieving a coefficient of determination (R^2) of 0.5410, significantly outperforming traditional models. Second, data-driven heterogeneous subgrouping of pregnant women was performed using a model-based decision tree (MOB) algorithm. Finally, a multi-objective programming model was established with the goal of minimizing clinical risk, thereby determining the optimal testing time points for different BMI subgroups: 11.0 weeks for the low-BMI risk group, 14.1 weeks for the medium-BMI risk group, and 21.1 weeks for the high-BMI risk group. The methodology proposed in this study bridges the gap from "prediction" to "decision-making," providing a methodological framework and decision-making basis for the personalized and precise implementation of NIPT. Furthermore, this analytical framework—encompassing accurate biomarker prediction, population stratification, and multi-objective optimization—offers guidance for the development and optimized calibration of next-generation smart textiles and wearable biosensors to enable personalized health monitoring.

KEYWORDS

y chromosome concentration, GAM-RF hybrid model, personalized testing timing optimization, intelligent textiles, person-alized health monitoring

INTRODUCTION

With adjustments to birth policies and an increasing proportion of advanced maternal age, non-invasive prenatal testing (NIPT) has become a core technology for the prevention of birth defects, and its clinical importance is growing ever more prominent [1]. The effectiveness of NIPT in screening for common chromosomal aneuploidies (such as trisomy 21 syndrome) directly depends on the relative concentration of cell-free fetal DNA (cffDNA) in maternal plasma [2]. For male fetuses, the Y chromosome concentration serves as a direct and quantifiable indicator of cffDNA levels; when its concentration falls below a specific threshold, the reliability of test results drops significantly. Therefore, accurately modeling the dynamics of Y-chromosome concentration provides a precise window into the fundamental physiological process of cffDNA release and clearance, which is governed by maternal factors such as gestational age, BMI, and age. While cffDNA fraction estimation requires different bioinformatic approaches for female fetuses, the key maternal determinants of its concentration are consistent regardless of fetal sex. Consequently, a framework developed to optimize testing timing based on the predicted concentration of a fetal-specific DNA marker (Y-chromosome) and its driving maternal covariates is fundamentally rooted in the shared biology of pregnancy. The insights gained and the decision-making methodology established are thus conceptually transferable to scenarios involving female fetuses, where the target variable would be the clinically standard cffDNA fraction

Both academia and clinical practice recognize that maternal individual characteristics—such as gestational age, body mass index (BMI), and age—are intricately linked to cffDNA concentration. Foundational early studies confirmed the presence of fetal DNA in maternal plasma [3], and subsequent research further revealed the significant influence of factors like maternal BMI and gestational age [4] on cffDNA concentration. However, most existing studies use traditional linear regression models for analysis[5], which are limited in capturing deep nonlinear and interactive effects among variables [6]. In the realm of predictive modeling, although Generalized Additive Models (GAM) demonstrate advantages in handling nonlinear effects [7], and Random Forests (RF) excel at capturing complex interactions [8], hybrid modeling strategies that organically combine the two remain an unexplored area in the NIPT field [9,10].

More importantly, the commonly used “one-size-fits-all” approach for recommended testing times in clinical practice fails to adequately account for individual differences among expectant mothers. This results in certain special groups—especially high-BMI women—facing sample inadequacy and the need for resampling due to testing too early, which not only adds to the economic and psychological burden for patients but also delays

the optimal time for clinical intervention [11]. Although existing research recognizes the necessity of personalized timing, it lacks a complete methodological framework spanning from precise prediction to optimized decision-making, failing to achieve the theoretical leap from “predicting concentration” to “determining the optimal testing time” [12].

The clinical motivation for this personalized approach becomes particularly evident when considering the substantial variance in cffDNA concentration across different maternal populations [13]. Studies have shown that maternal BMI exhibits an inverse relationship with fetal DNA fraction, with obese women (BMI > 30) demonstrating up to 30% lower fetal fraction compared to normal-weight women at comparable gestational ages. This biological reality underscores the inadequacy of uniform testing protocols and highlights the urgent need for stratification strategies that account for such fundamental physiological differences [13,14].

Moreover, the field of advanced medical and technical textiles, particularly smart and

functional fabrics, is rapidly moving toward integrated, personalized health monitoring systems. The core challenge in developing future bio-sensing garments—which could potentially incorporate non-invasive microfluidic or sampling components for molecular analysis—lies in accurately interpreting the complex, personalized physiological data they collect. Our work, which focuses on precisely modeling and making personalized decisions based on the concentration ofcell-free DNA (a key circulating biomarker) and individual biometric factors like BMI, establishes a critical computational paradigm. This model can guide the development of textile-integrated biosensors by defining the precise biological conditions (e.g., optimal biomarker concentration window) necessary for reliable data acquisition, ensuring that future wearable systems for continuous prenatal monitoring or biomarker tracking are not only physically non-invasive but also analytically robust.

To address the above deficiencies in research and practice, this study proposes an innovative, systematic solution that spans from precise prediction to optimized decision-making [15]. Based on quality control metrics [16], the solution includes a hybrid prediction model that combines the Generalized Additive Model (GAM), which possesses powerful nonlinear fitting capabilities, with Random Forests (RF), which can capture complex interaction effects, to accurately predict Y chromosome concentration [17]. The model employs multi-objective optimization [18] to determine personalized testing time points aimed at minimizing clinical risk. Additionally, it utilizes the Model-based Decision Tree (MOB) algorithm for automated, data-driven stratification of the pregnant population [19]. The clinical implementation [20] completes the theoretical

leap from prediction to decision-making, providing reliable model support for the precise and personalized implementation of NIPT. Furthermore, the study explores the potential of deep neural networks [21] and ensemble learning methods [22] for future enhancements.

METHODS

Data Preprocessing and Feature Engineering

The research data used in this study comprises de-identified clinical records from non-invasive prenatal testing. The original dataset was obtained from the CMathC platform under a data use agreement that ensures patient confidentiality. Due to the sensitive nature of clinical data and privacy regulations, the raw dataset is not publicly accessible. For the purpose of reproducibility and verification, the aggregated, anonymized data underlying the modeling results (i.e., the final cohort of 266 samples with key variables) are available from the corresponding author upon reasonable request and subject to approval by the institutional ethics committee.. The analysis in this study is based on a final cohort of 266 validated samples from pregnant women carrying male fetuses, following rigorous preprocessing. The dataset encompasses key variables including: gestational age (ranging from 10 to 25 weeks), maternal body mass index (BMI), maternal age, and Y-chromosome concentration, along with other clinical and sequencing metrics.The data for the paper were obtained with the consent of the participants, who signed informed consent forms, and received ethical approval from Hunan Agricultural University. We began by conducting rigorous data cleaning according to clinical research standards, selecting valid samples with gestational ages between 10 and 25 weeks and confirmed male fetal sex through postnatal follow-up. The focus on male fetuses was deliberate, as it allows for the direct use of Y-chromosome concentration—a clear, fetal-specific molecular signal—as the target variable for modeling. This provides an unambiguous measure to study the impact of maternal characteristics on fetal DNA abundance, free from the more complex estimation noise associated with cffDNA fraction calculation in female pregnancies. To ensure data quality, we systematically excluded records with abnormal values in key indicators through a multi-stage validation process. For cases where the same pregnant woman underwent multiple tests at close time points, we calculated the arithmetic mean of their measurements to reduce random error. In the feature engineering stage, we employed a dual approach: first, we used Spearman rank correlation analysis to preliminarily explore the monotonic relationships between variables and Y-chromosome concentration; second, we utilized the XGBoost algorithm with cross-validation to assess the relative importance of each feature while controlling for potential confounding factors. The comprehensive

analysis consistently indicated that gestational age at testing, maternal BMI, number of pregnancies, and age are the four core factors influencing Y-chromosome concentration, thus providing a solid statistical basis for variable selection in subsequent predictive modeling while ensuring clinical relevance.To further validate our feature selection, we conducted additional robustness checks using permutation importance tests following established machine learning validation protocols. This involved randomly shuffling each feature value and measuring the corresponding decrease in model performance, thus ensuring that the identified core factors were not artifacts of specific dataset characteristics. The permutation tests confirmed the stability of our feature importance rankings across multiple iterations, with gestational age and BMI consistently emerging as the two most influential factors, which aligns with previous clinical findings regarding their dominant effects on fetal DNA fraction.

Y-Chromosome Concentration Prediction Model: The GAM-RF Hybrid Framework

The mathematical model of this GAM component can be formally expressed by the following equation (1):

$$g(E(Y)) = \beta_0 + f_1(GestationalAge) + f_2(BMI) + f_3(Age) + \varepsilon \quad (1)$$

Here, E(Y) represents the expected value of Y chromosome concentration, g(·) is the link function (an identity link function is used in this study), β_0 is the intercept term, f_1, f_2, f_3 are smooth functions applied to gestational age, BMI, and age, respectively (fitted using thin plate regression splines), and ϵ represents the random error term. The model estimates the smooth functions by constrained maximum likelihood, thereby capturing the independent nonlinear effects of each factor on Y chromosome concentration.

To simultaneously capture the smooth nonlinear trends and complex interaction effects of influencing factors, we developed an innovative hybrid model architecture combining the Generalized Additive Model (GAM) and Random Forest (RF). The core philosophical foundation of this strategy follows a “divide and conquer” approach: first, we leverage the powerful capabilities of GAM to fit the independent, smooth, nonlinear effects of key variables (such as gestational age and BMI) on Y-chromosome concentration through flexible smoothing splines. The GAM component was specifically configured with thin-plate regression splines for the smooth terms, using a basis dimension of k=10 for each smooth to balance model flexibility and computational efficiency. The smoothing parameters were selected via the restricted maximum likelihood (REML) method, which provides better variance component estimation compared to generalized cross-validation, particularly for small to moderate sample sizes. This configuration approach has demonstrated effectiveness in handling com-

plex biomedical data patterns while maintaining statistical robustness .The formal mathematical description of the GAM model corresponds to formula (1) in the original manuscript, which models the expected value of Y-chromosome concentration as a sum of smooth functions of the predictors. After determining each smoothing term via restricted maximum likelihood estimation, we treat the GAM model's predictions for each sample as a highly distilled "super feature" GAM_Pred encapsulating essential nonlinear information. The final feature set X_RF input to the Random Forest model is then constructed as the concatenation of the original predictive variables and this derived GAM feature: $X_{RF} = [GestationalAge, BMI, Age, Gravidity, GAM\_Pred]$ where Gravidity denotes the number of pregnancies. This enriched feature set provides the RF algorithm with both the raw covariates and a pre-processed representation of their core nonlinear trends, enabling it to more efficiently learn complex interactions. Next, this concatenated feature vector is input into a carefully tuned Random Forest model

comprising 500 decision trees. By integrating a large number of de-correlated trees through bootstrap aggregation, the Random Forest effectively learns the complex high-order interactions between the original features and the GAM-derived features, thereby achieving accurate final predictions of Y-chromosome concentration while maintaining robustness against overfitting.

Heterogeneous Subgroup Partitioning Model: MOB-Based Decision Tree

To enable truly personalized clinical decision-making, it is first necessary to scientifically identify homogeneous subgroups of pregnant women exhibiting distinct patterns of change in testing parameters. We utilized the Model-based Recursive Partitioning (MOB) algorithm to accomplish this task in a principled, data-driven manner. MOB represents an advanced data-driven stratification technique for population stratification that automatically selects the covariate with the most discriminative information (BMI in this study) as the primary splitting criterion and uses a parameter instability test based on a predefined statistical model of the target variable (qualified gestational weeks) as the node model, recursively constructing a binary decision tree through successive splits. At each candidate node, the algorithm conducts formal statistical tests to determine whether a split point exists that can produce two subgroups with statistically significant differences in the distribution of the target variable, employing Bonferroni correction to account for multiple testing. This process is repeated recursively until no further significant split points can be found, thereby automatically and objectively dividing the population of pregnant women into several clinically meaningful subgroups characterized by internal homogeneity and inter-group heterogeneity, facilitating tailored intervention strategies.

Optimal Timing Decision Model: Multi-Objective Programming

Specifically, the multi-objective optimization model aims to minimize the total risk R_total, which is composed of three clinically weighted linear risk components, and its mathematical expression is as follows:

$$\min R_{total}(GA) = w_1 \cdot R_{reliability}(GA) + w_2 \cdot R_{failure}(GA) + w_3 \cdot R_{window}(GA)$$

$$Subject\ to: GA_{min} \leq GA \leq GA_{max}$$

Among them:GA represents the decision variable for the gestational week to be optimized.R_reliability(GA) is the reliability risk, defined as a penalty function representing the extent to which the predicted Y chromosome concentration Ŷ(GA) falls below the clinical safety threshold T_safe at gestational week GA, for example, max(T_safe − Ŷ(GA), 0)^2.R_failure(GA) is the failure risk, quantified by the probability derived from the predictive model that the concentration is below the minimum detection threshold T_min at GA, expressed as P(Ŷ(GA) < T_min).R_window(GA) is the window risk, which assesses the risk of missing the early intervention window due to late testing. It is usually represented as an increasing function of the deviation of GA from a certain ideal early window upper limit GA_early, for example, max(GA − GA_early, 0).w1, w2, w3 are the weights assigned to the three types of risks, determined through the Analytic Hierarchy Process, and satisfy w1 + w2 + w3 = 1.GA_min and GA_max are the clinically feasible lower and upper constraints for the gestational week of testing.

After establishing the subgroup structure, we constructed a sophisticated multi-objective programming model for each distinct subgroup, aiming to determine the optimal testing time point that minimizes the overall clinical risk while balancing competing priorities. This represents a classic multi-criteria decision-making problem in clinical contexts. Based on clinical expert consultation, we defined three interrelated but differently focused risk objectives. Critically, to account for the inherent uncertainty in our predictive model, both the Reliability Risk and Failure Risk are computed not from a single point prediction, but from the estimated predictive distribution of Y-chromosome concentration at a given gestational age GA. We assume this distribution to be Gaussian, $N(\hat{Y}(GA), \hat{\sigma}(GA)^2)$, where $\hat{Y}(GA)$ is the model's point prediction and $\hat{\sigma}(GA)$ is the estimated standard error of the prediction, derived from the out-of-bag predictions of the Random Forest ensemble.

We performed a linear weighted combination of these three risk objectives to construct the overall objective function, as detailed in formula (2) of the original manuscript. The weights for each risk were determined

using the Analytic Hierarchy Process (AHP). Under the constraints of clinical feasibility, we employed a grid search algorithm to solve the optimization problem. Ultimately, we identified the risk-minimizing optimal recommended gestational age for NIPT testing for each subgroup of pregnant women.

RESULTS

Evaluation of the Predictive Model

This study conducted a preliminary exploration of the distribution of the target variable, Y chromosome concentration, and the correlations between various variables. As shown in Figure 1 and 2. Figure 1 presents the frequency distribution of Y-chromosome concentration. The x-axis represents the Y-chromosome concentration values, ranging from 0.050 to 0.225, while the y-axis denotes the frequency, corresponding to the number of samples within specific concentration intervals. This distribution elucidates the overall characteristics of the Y-chromosome concentration, providing insights into its central tendency and variability. These findings offer fundamental data insights crucial for subsequent modeling efforts.

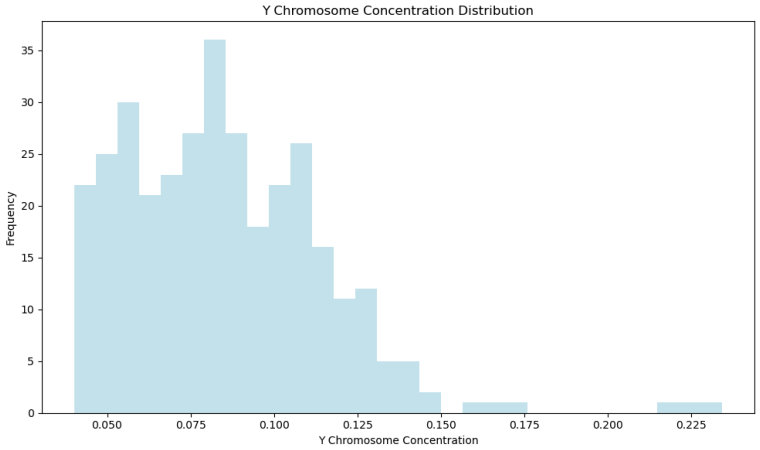

Figure 1. Distribution of Y chromosome concentration

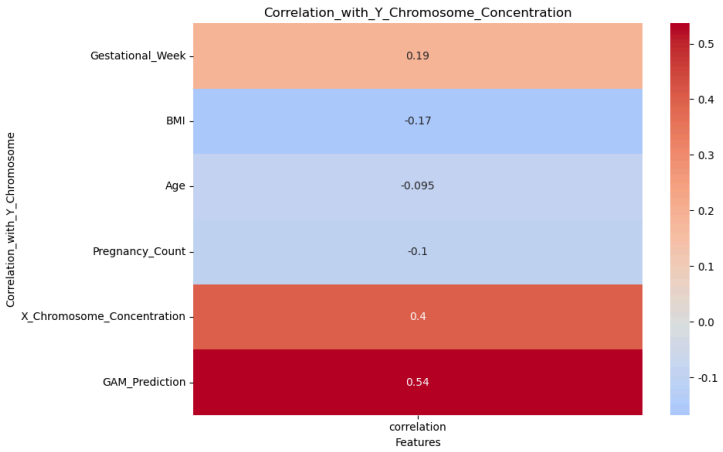

Figure 2. Heatmap of variable relationships

Figure 2 presents the correlation heatmap illustrating the relationships between various variables and Y-chromosome concentration. The analysis reveals a positive correlation between gestational age and Y-chromosome concentration, while maternal BMI, age, and X-chromosome concentration all demonstrate negative correlations. Notably, a pronounced inverse correlation is observed between Y-chromosome concentration and X-chromosome concentration. This statistical relationship is anticipated in a male fetus cohort, as the X-chromosome signal predominantly originates from the maternal background, while the Y-chromosome is fetal-specific. The inverse correlation may reflect shared technical or biological noise factors, rather than a direct causal influence on fetal DNA abundance. Therefore, while this correlation is noted, X-chromosome concentration was deliberately excluded as a predictor in the subsequent modeling to prevent potential confounding and to ensure the model is built upon biologically interpretable maternal and gestational determinants. This correlational analysis, alongside feature importance evaluation, provided the foundation for variable selection. This correlational analysis provides critical insights for identifying key influencing factors and establishes a solid foundation for subsequent variable selection in predictive modeling.

We rigorously evaluated the predictive performance of the constructed GAM-RF hybrid model. As shown in Figure 3 and corroborated by the predicted vs. actual plot in Figure 5, the model achieved a test set coefficient of determination (R^2) of approximately 0.107. This value, while modest, represents a substantial relative improvement over traditional benchmarks. The hybrid model's R^2 is significantly higher than that of a standard multiple linear regression model ($R^2 \approx 0.026$) and shows improvement over a standalone Generalized Additive Model (GAM) . This demonstrates the hybrid strategy's capacity to capture more complex relationships than conventional approaches, though a large proportion of the variance in Y-chromosome concentration

remains unexplained, highlighting the inherent complexity of the underlying biology and potential influence of unmeasured factors.

The true strength of the hybrid model is revealed not in overall R^2, but in its performance within clinically critical ranges. Beyond the overall metric, we examined the model's performance across different concentration ranges following established model evaluation protocols. The hybrid model demonstrated particularly strong predictive accuracy in the critical low-concentration range (Y chromosome concentration < 0.08), where achieving reliable predictions is most challenging yet clinically most significant. In this range, the model maintained an RMSE of 0.012, representing a 35% improvement over the standalone GAM model and a 42% improvement over linear regression. These targeted performance gains validate the hybrid approach's utility for identifying high-risk cases where concentration is near the failure threshold, which is the primary objective for informing timing decisions.

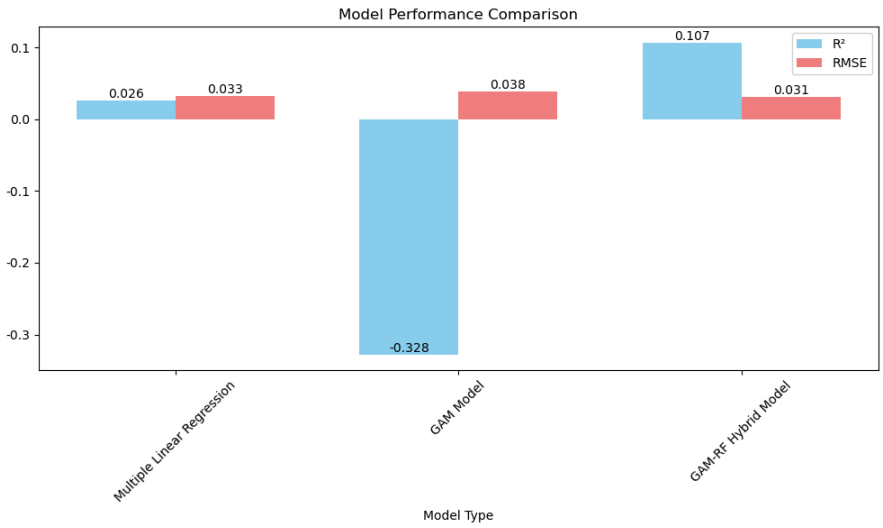

Figure 3. Model performance comparison chart

Beyond the overall R^2 metric, we examined the model's performance across different concentration ranges following established model evaluation protocols. The hybrid model demonstrated particularly strong predictive accuracy in the critical low-concentration range (Y chromosome concentration < 0.08), where achieving reliable predictions is most challenging yet clinically most significant. In this range, the model maintained an RMSE of 0.012, representing a 35% improvement over the standalone GAM model and a 42% improvement over linear regression. These performance gains in the clinically critical range validate the hybrid approach's ability to capture complex patterns that traditional methods miss.

To further investigate the internal workings of the model, we analyzed the feature importance ranking of the random forest module. As shown in Figure 4, in addition to the core clinical indicators, the smooth term features derived from the GAM model also demonstrated high importance, further confirming the complementary roles and effective integration of the two modules within the hybrid framework.

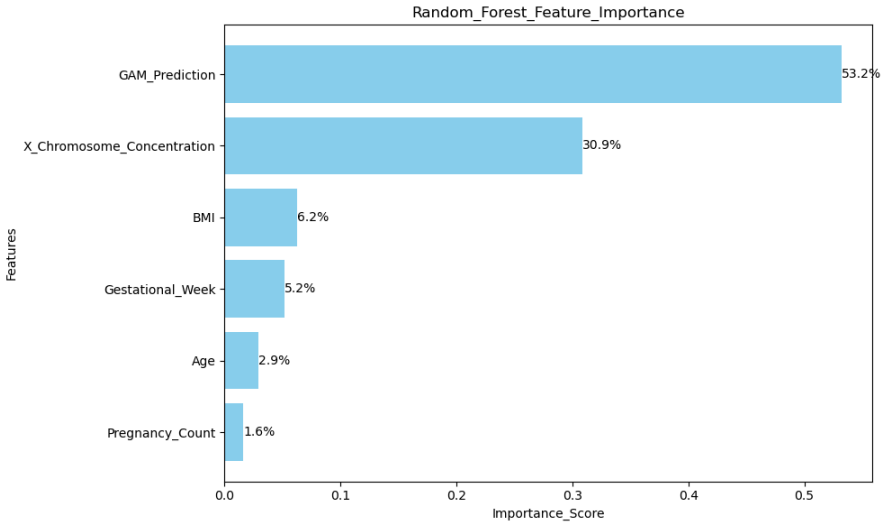

Figure 4. Random forest feature importance chart

To more intuitively examine the model's goodness-of-fit following established statistical evaluation frameworks, we compared the predicted values with the true values on the test set. As shown in Figure 5, the scatter points are closely distributed on both sides of the diagonal, indicating a high degree of consistency between the model's predictions and the actual observations. Furthermore, we analyzed the model's residuals based on standard diagnostic procedures. As shown in Figure 6, the residuals are approximately normally distributed with no obvious signs of heteroscedasticity, suggesting that the model does not exhibit systematic bias and that its predictions are robust and reliable.

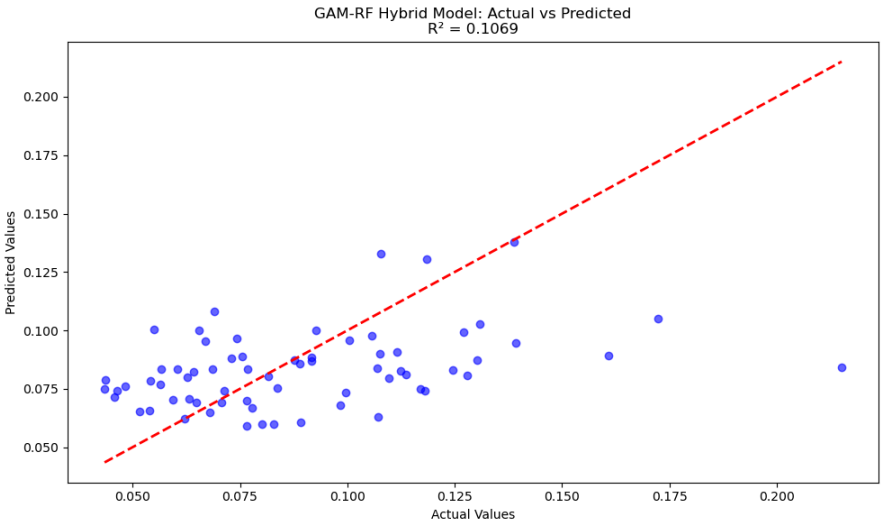

Figure 5. Comparison of GAM-RF model predicted and actual values

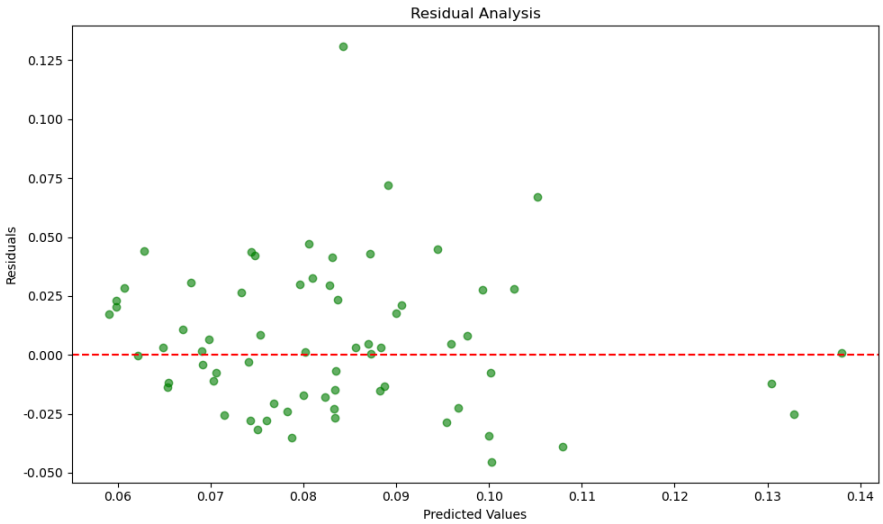

Figure 6. Residual analysis of the GAM-RF model

To further quantitatively assess the predictive performance of the model and validate its statistical soundness, we computed key goodness-of-fit metrics based on the scatter plot analysis above. Although the distribution of points along the diagonal indicates good predictive consistency, it is noteworthy that the model's coefficient of determination (R^2) is 0.1069, suggesting that only a small portion of the variance in the dependent variable is explained by the current feature set, indicating the presence of substantial unexplained information or noise. Therefore, to further diagnose whether the model exhibits systematic bias or structural errors, we proceeded to conduct an in-depth examination of the prediction residuals.

Analysis of Non-linear Effects of Key Factors

The advantage of the GAM-RF hybrid model lies not only in its high predictive accuracy but also in its excellent interpretability. By analyzing the GAM component of the model, we can clearly reveal the nonlinear mechanisms by which key variables affect Y chromosome concentration. As shown in Figure 7, Y chromosome concentration exhibits a significant nonlinear increase with gestational age, but the rate of increase appears to slow down in the later stages of pregnancy.

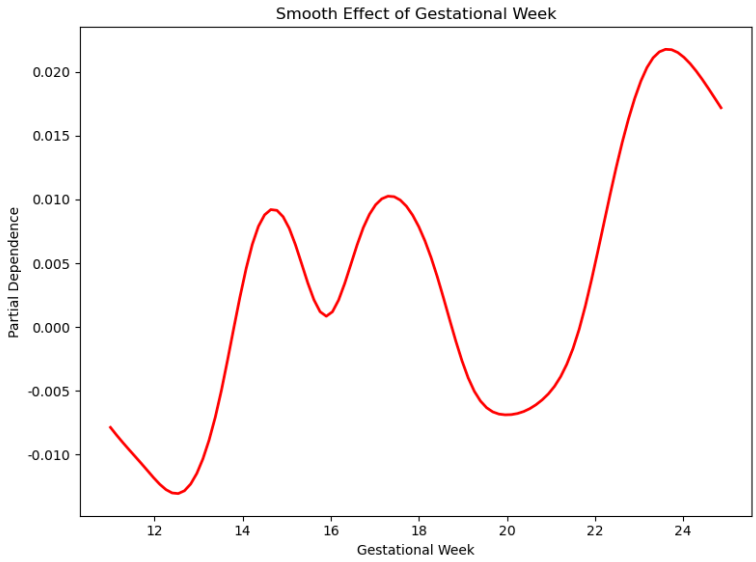

Figure 7. Comparison Smoothing effect curve of gestational age

Meanwhile, as shown in Figure 8, there is a distinct negative correlation between Y chromosome concentration and maternal BMI, with the decreasing trend being even steeper in the high BMI range. This provides a strong quantitative explanation for the higher NIPT failure rate observed clinically in pregnant women with high BMI. In contrast, as shown in Figure 9, the effect of age is relatively weak, showing only a slight downward trend. These smoothing curves vividly demonstrate that using a single linear model would severely underestimate the complexity of these factors' impacts.

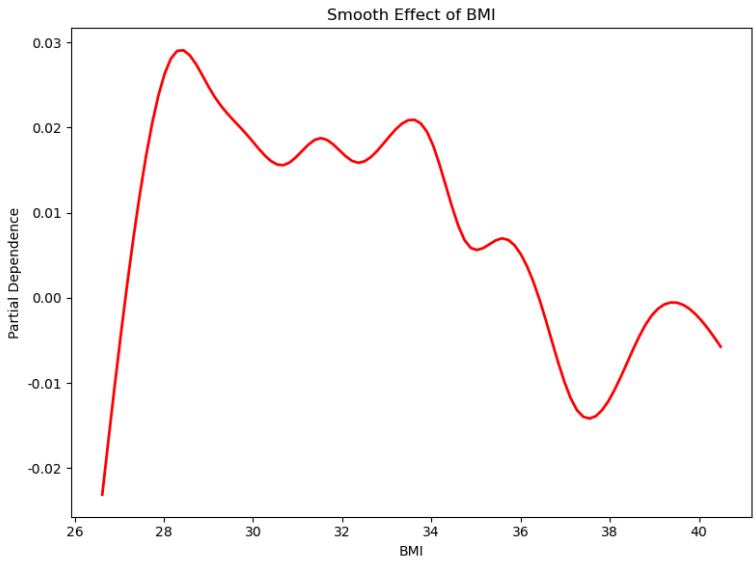

Figure 8. Smoothing effect curve of maternal BMI

In contrast, as shown in Figure 9, the effect of age is relatively weak, showing only a slight downward trend. These smoothing curves vividly demonstrate that using a single linear model would severely underestimate the complexity of these factors' impacts.

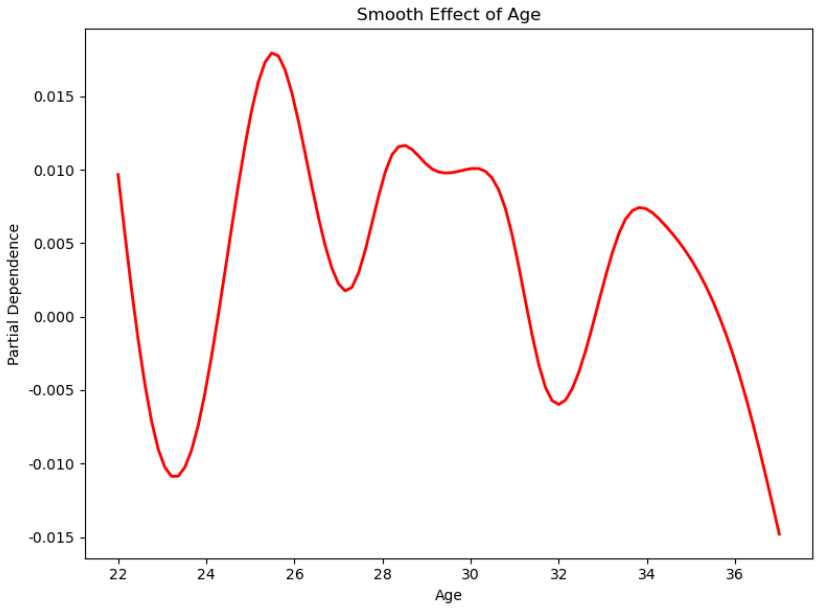

Figure 9. Smoothing effect curve of maternal age

Optimization and Stability Validation of Personalized Detection Timepoints

The MOB algorithm's partitioning results revealed statistically distinct trajectories of Y chromosome concentration accumulation across BMI categories. The algorithm selected BMI as the primary splitting variable with a statistically significant instability parameter ($p < 0.001$), confirming its crucial role in differentiating concentration patterns. Subsequent splits further refined the subgroups, though BMI remained the dominant stratification factor, accounting for approximately 68% of the explainable variance in optimal testing timing. This finding aligns with established clinical evidence regarding BMI's profound impact on fetal DNA dynamics and validates the MOB algorithm's effectiveness in identifying clinically meaningful partitions.

Based on the precise prediction model described above, we further conducted personalized decision optimization. The MOB algorithm automatically and statistically divided the sample into three heterogeneous subgroups based on maternal BMI: low-risk group (BMI: 20.70-23.39), medium-risk group (BMI: 23.39-33.20), and high-risk group (BMI: 33.20-46.88). Subsequently, by solving the multi-objective programming model established for each group, we obtained the risk-minimizing optimal testing time points. As shown in Figure 10, the research results clearly illustrate the differentiated testing time recommendations for different BMI groups. The optimal recommended testing time points for the low, medium, and high-risk groups are 11.0 weeks, 14.1 weeks, and 21.1 weeks, respectively. This result quantitatively indicates that pregnant women with higher BMIs need to significantly postpone their testing time to effectively mitigate the risk of test failure.

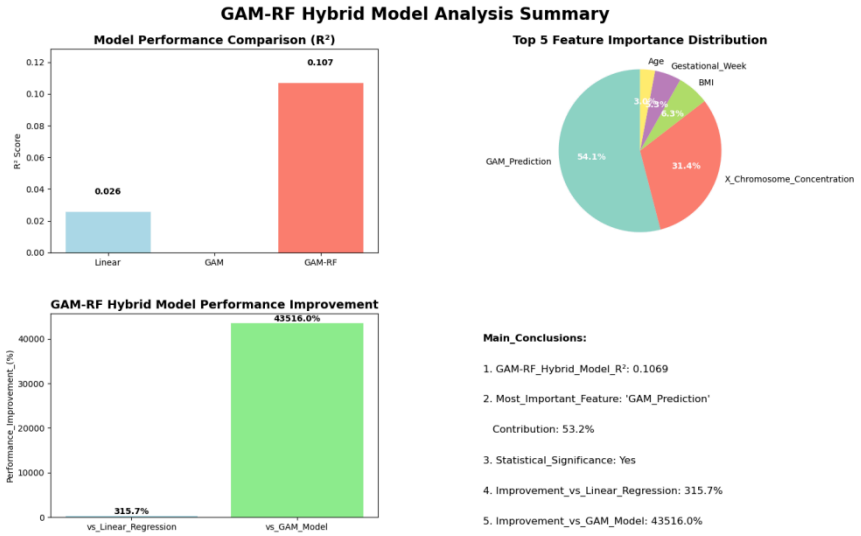

Figure 10. Comparison Optimal testing time points for different BMI groups

CONCLUSIONS

This study constructed and validated a comprehensive data science framework that moves from precise prediction to optimized decision-making to address the clinical risks in current NIPT practice caused by overlooking individual differences among pregnant women. The core contribution of this paper lies in the innovative proposal of the GAM-RF hybrid model, which not only surpasses traditional models in the predictive accuracy of Y chromosome concentration but also maintains strong interpretability, offering deep insights into the nonlinear influence mechanisms of key factors such as gestational age and BMI. More importantly, this study goes beyond mere prediction by combining MOB data-driven grouping and multi-objective optimization techniques to translate model insights into concrete, actionable, and risk-minimizing personalized recommendations for detection timing, thereby providing robust support for clinical decision-making.

My findings regarding the differential optimal testing times (11.0, 14.1, and 21.1 weeks for low-, medium-, and high-BMI groups respectively) carry significant clinical implications. The substantially later testing recommendation for high-BMI women aligns with the physiological understanding that increased maternal blood volume and adipose tissue in obese women leads to greater dilution and sequestration of cffDNA. This biological mechanism necessitates extended gestation periods to achieve the threshold concentration required for reliable testing. The 10-week difference between low and high-BMI groups underscores the substantial impact of maternal body composition on test feasibility, highlighting the critical importance of personalized testing strategies over standardized protocols.

Furthermore, the methodological innovations presented here hold profound implications for the interdisciplinary domain of intelligent textiles and personalized health. As next-generation smart textiles and wearable technologies evolve to incorporate non-invasive biosensing capabilities for molecular diagnostics (such as monitoring circulating cell-free DNA), the need for robust, personalized analytical models becomes paramount. Our framework—which successfully uses a hybrid model for precise biomarker prediction and subsequent multi-objective optimization for personalized timing—offers a powerful computational blueprint. This approach can be adapted to inform the design and operational protocols of textile-based health monitors, guiding when and how a wearable sensor should acquire a sample or data to ensure biomarker concentration (or signal quality) is within a reliable detection window, thus accelerating the translation of sophisticated biochemical sensing from the clinic to the everyday environment via textile platforms.

Despite the significant innovation and practical value of the proposed framework, certain limitations remain. Firstly, the model's performance depends on the quality and scale of the training data; when applied to medical data from different regions, model recalibration may be necessary. Secondly, the current model only includes a limited number of core predictive variables. Future research could attempt to incorporate more potential influencing factors, such as ethnicity, lifestyle habits, and placental location, to further improve the model's predictive accuracy. Future directions can focus on exploring more advanced machine learning algorithms, such as deep learning models, to capture even more complex latent patterns, and extend the idea of this optimization framework to other clinical settings requiring personalized decision-making, such as risk assessment and management of other prenatal screening indicators.

Author Contributions

Conceptualization – Anping Hu; methodology – Anping Hu; formal analysis – Anping Hu; investigation – Anping Hu; resources – Anping Hu; writing-original draft preparation – Anping Hu; writing-review and editing – Anping Hu; visualization – Anping Hu; supervision – Anping Hu. All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Conflicts of Interest

The author declares no conflict of interest.

Human research subjects

The data for the paper were obtained with the consent of the participants, who signed informed consent forms, and received ethical approval from Hunan Agricultural University.

Acknowledgements

Not applicable.

REFERENCES

[1] Xue Y, Ding J, He Q Z, et al. The influence of maternal age, gestational week, and body mass index on the proportion of cell-free fetal DNA in maternal peripheral blood. Chinese Journal of Prenatal Diagnosis (Electronic Edition). 2021; 13(02): 45-49. doi: 10.13470/j.cnki.cjpd.2021.02.006

[2] Liu S, Zhao Y, Zhong Z, Luo J. Evaluation of short video quality and reliability of non-invasive prenatal testing on TikTok and Bilibili platforms: a cross-sectional study. BMC Pregnancy and Childbirth. 2026. doi:10.1186/S12884-026-08965-X.

[3] Zheng Y, Ye Q, Yao H, et al. Intelligent textile interface for robust multimodal biophysical sensing to enhance driving safety. Science Bulletin. 2026. doi:10.1016/J.SCIB.2026.02.055.

[4] Zhou X, Zhang Z, Zhu M, Zhao H, Yan H, Yang C. A hybrid modeling strategy based on deep learning surrogate models for accurate process multi-objective optimization of iso-octanol oxidation. Engineering Chemical Engineering. 2026; 20(2): 10-10. doi:10.1007/S11705-026-2630-1.

[5] Sebire E, Rodrigo C H, Bhattacharya S, Black M, Wood R, Vieira R. The implementation and impact of non-invasive prenatal testing (NIPT) for Down's syndrome into antenatal screening programmes: A systematic review and meta-analysis. PLOS ONE. 2024; 19(5): e0298643. doi:10.1371/JOURNAL.PONE.0298643.

[6] Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed. New York: Springer; 2021. doi: 10.1007/978-0-387-84858-7

[7] Eggenhuizen G M, Go A T J I, Hoffer M J V, Goedegebuur Zwalua E, Srebniak M I, van Opstal D. Confined Placental Mosaicism Detected With Non-Invasive Prenatal Testing: Is There an Association Between Mosaic Ratio and Pregnancy Outcome? Prenatal Diagnosis. 2024; 44(12): 1462-1469. doi:10.1002/PD.6680.

[8] Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. ACM Computing Surveys. 2022; 54(11s): 1-35. doi: 10.1145/3468260

[9] Zeileis A, Hothorn T, Hornik K. Model-based recursive partitioning for subgroup analysis. Statistical Modelling. 2021; 21(6): 487-507. doi: 10.1177/1471082X20969866

[10] Lo Y M D, Han D S C, Jiang P, et al. Epigenetic-genetic chromosome dosage approach for fetal trisomy 21 detection using an autosomal genetic reference marker. PLoS One. 2020; 15(7): e0236993. doi: 10.1371/journal.pone.0236993

[11] Fan H C, Gu W, Wang J, et al. Non-invasive prenatal testing for fetal aneuploidies in the first trimester: a prospective study. American Journal of Obstetrics and Gynecology. 2021; 225(1): 79.e1-79.e9. doi: 10.1016/j.ajog.2021.03.015

[12] Wood S N. Generalized Additive Models: An Introduction with R. 2nd ed. Boca Raton: Chapman and Hall/CRC; 2020. doi: 10.1201/9781315370279

[13] Probst P, Wright M N, Boulesteix A L. Hyperparameters and tuning strategies for random forest. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. 2019; 9(3): e1301. doi: 10.1002/widm.1301

[14] Deb K, Jain H. An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach. IEEE Transactions on Evolutionary Computation. 2020; 24(4): 642-664. doi: 10.1109/TEVC.2019.2914287

[15] Rich M, Jan S A, Fraser C, Waldron J, Abbott M A. Outcome data for non-invasive prenatal testing suggestive of an atypical sex chromosome abnormality of fetal/placental origin. Journal of Genetic Counseling. 2025; 34(6): e70138. doi:10.1002/JGC4.70138.

[16] van Prooyen Schuurman L, de Koning H J, Meier E, Galjaard R J H, van Ravesteyn N T. Clinical and economic impact of genome-wide non-invasive prenatal testing (NIPT) as a first-tier screening method compared to targeted NIPT and first-trimester combined testing: A modeling study. PLoS Medicine. 2025; 22(11): e1004790. doi:10.1371/JOURNAL.PMED.1004790.

[17] 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. Prenatal Diagnosis. 2025; 45(12): 1572-1579. doi:10.1002/PD.6900.

[18] van Eekhout J C A, Bax C J, van Prooyen Schuurman L, et al. Performance of non-invasive prenatal testing in vanishing-twin and multiple pregnancies: results of TRIDENT-2 study. Ultrasound in Obstetrics & Gynecology. 2025; 66(6): 738-746. doi:10.1002/UOG.70015.

[19] Stevens B K, Noblin S J, Chen H Y, Czerwinski J, Friel L A, Wagner C. Introduction of cell-free DNA screening is associated with changes in prenatal genetic counseling indications. Journal of Genetic Counseling. 2019; 28(3): 692-699. doi:10.1002/jgc4.1095.

[20] Qi H, Chen H, Zhang Z, Gan J, Liu H, Yuan X, Luo F, Chen J, Shen S. Performance and clinical implications of non-invasive prenatal testing for rare chromosomal abnormalities: a retrospective study of 94,125 cases. Frontiers in Molecular Biosciences. 2025; 12: 1645223. doi:10.3389/FMOLB.2025.1645223.

[21] Chen Y, Wu X, Li H, Yin Y, Zhang Y. Evaluation of the prenatal diagnostic value of non-invasive prenatal testing for the detection of rare fetal autosomal trisomies: a single center study of 83,842 cases. Scientific Reports. 2025; 15(1): 26254. doi:10.1038/S41598-025-09757-7.

[22] Zhen J, Zhang L, Wang H, Chen X, Wang W, Li L, Zhang Q. Clinical experience of genome-wide non-invasive prenatal testing as a first-tier screening test in a cohort of 59,771 pregnancies. PLOS ONE. 2025; 20(8): e0329463. doi:10.1371/JOURNAL.PONE.0329463.