Algorithmic Optimization of Prenatal Testing Schedules Using Generalized Averaged Non-Expansive Operators

Wendi Zhao[1,a,#], Junru Ren[2,b,#], Yulin Li[2,c,#]

[1]International College, Jinan University, Guangzhou, China

[2]Jinan University - University of Birmingham Joint Institute, Jinan University, Guangzhou, China

[a]2736648996@qq.com, [b]1563529191@qq.com, [c]15904279988@163.com

[#]These authors contributed equally to this work.

Keywords: Generalized mean non-inflated operator; NIPT detection; Multiple nonlinear regression; Fetal abnormality determination; Target optimization

Abstract: With the rapid development of non-invasive prenatal testing (NIPT) technology, how to reasonably select the detection time and accurately determine fetal abnormalities has become an important research topic in the field of prenatal diagnosis. Aiming at the problem of point-in-time optimization in NIPT detection, a multi-objective optimization model is proposed based on generalized average non-expansion operator (GAN) theory and machine learning methods. The results of model analysis showed that the gestational age was significantly positively correlated with the concentration of Y chromosome, while the BMI of pregnant women and age were significantly negatively correlated, and the model successfully revealed the key physiological indicators that affect the basis of NIPT detection. In the second step, in order to minimize the risk, a dynamic grouping and detection time optimization model based on BMI was constructed, and the model divided pregnant women into 5 BMI groups through GAN operator solving, and gave the corresponding optimal detection time point (11-12 weeks), so that the total risk value was reduced to 2851.0, and the balance between personalized detection and risk control was realized. In the third step, in order to integrate multi-dimensional physiological indicators for more refined grouping, an intelligent clustering model fusing random forest and GAN operator was proposed, and the model identified three groups of pregnant women with different characteristics, and verified that BMI (importance: 0.2536) and gestational age (importance: 0.1902) were the core basis for personalized grouping. The unified optimization framework proposed in this paper innovatively applies the GAN operator to the whole process of NIPT data analysis, and the model has the advantages of fast convergence, good stability and strong interpretation, which can provide scientific and accurate decision-making support for the selection of clinical NIPT detection time points.

1. Introduction

Non-invasive prenatal testing is a prenatal testing technique that collects maternal blood, detects fetal free DNA fragments, and analyzes fetal chromosomes for abnormalities. This technology has the advantages of non-invasiveness, safety, and accuracy, and has become an important means in the field of prenatal diagnosis[1]. According to clinical experience, fetal malformations mainly include Down syndrome, Edwards syndrome, and Patau syndrome, which are determined by abnormal proportions of free DNA fragments on fetal chromosomes 21, 18, and 13, respectively[2]. The accuracy of NIPT mainly depends on the detection of fetal sex chromosome concentrations (male XY, female XX). In general, it is feasible to test the concentration of sex chromosomes in the fetus between 10 and 25 weeks of pregnancy. When the Y chromosome concentration in male fetuses reaches or exceeds 4%, and the X chromosome concentration in female fetuses does not show abnormalities, the NIPT results can be considered basically accurate. On the other hand, if it is below this threshold, it is difficult to ensure the reliability of the test results[3].

In practical applications, it is crucial to detect abnormalities in the fetus as early as possible.

  DOI: 10.25236/iwmecs.2025.016

Early identification of unhealthy fetuses will help to intervene early and avoid shortening the treatment window. Specifically, the risk associated with early detection (within 12 weeks) is low; the mid-term detection (13-27 weeks) is at higher risk; Late detection (after 28 weeks) will lead to a very high risk[4]. Practice shows that there is a strong correlation between male fetal Y chromosome concentration and pregnant women's gestational age and body mass index (BMI). Typically, pregnant women are grouped based on their BMI values (e.g., [20,28), [28,32), [32,36), [36,40, 40 and above), and the time point of detection of NIPT is determined accordingly." Due to individual differences in age, BMI, pregnancy status, etc. of each pregnant woman, the use of uniform empirical grouping and testing time points for NIPT may significantly affect its accuracy [5].

To solve the above challenges, this paper constructs a unified multi-objective optimization framework based on generalized average non-expansive operators (GANs) and machine learning methods. In the first step, the correlation between Y chromosome concentration and gestational age, BMI and other indicators was analyzed by multiple regression model [6][7]. In the second step, a time-in-time optimization model based on BMI grouping is established to minimize the risk[8]. In the third step, random forests and GAN operators are introduced for intelligent clustering to achieve fine grouping under multi-factor coupling[9]. This study aims to provide personalized and precise decision support for NIPT detection[10].

2. Model creation, solution and discussion

2.1. Model establishment

2.1.1. Analysis model of factors affecting Y chromosome concentration

In order to explore the correlation between Y chromosome concentration and physiological indicators of pregnant women, a multiple linear regression model was established. Assuming that there is a linear relationship between fetal Y chromosome concentration (Y) and gestational age (W), BMI (B), age (A), height (H), and weight (M), the model expression is:

$$Y = \beta_0 + \beta_1 W + \beta_2 B + \beta_3 A + \beta_4 H + \beta_5 M + \varepsilon \tag{1}$$

Among them, ε is the error term. To optimize the regression coefficients, the generalized average non-expanding operator (GAN) is introduced. The update operator $T(\theta)$ is defined as:

$$T(\theta) = (X^T X + \lambda I)^{-1} X^T y \tag{2}$$

Where X is the feature matrix, y is the output variable, and λ is the regularization parameter. The GAN operator update formula is:

$$G(w) = (1-\alpha)w + \alpha T(w) \tag{3}$$

where α is the equilibrium parameter, which makes the model converge stably through iteration.

2.1.2. BMI grouping and time-of-detection optimization model

With the goal of risk minimization, a time-of-detection optimization model based on BMI grouping was constructed. The risk function is defined as:

$$\text{Risk}_i = \sum_{\text{i}=1}^{\text{n}} risk\_weight(\text{Week}_i) \tag{4}$$

Among them, $\text{risk}\_\text{weight}$ is weighted according to the gestational age window: early (≤12 weeks) weight 1, mid-term (13-27 weeks) weight 3, and late (≥28 weeks) weight 5. The goal is to minimize the total risk:

$$\min \sum_{i=1}^{k} \text{Risk}_i \tag{5}$$

Optimize BMI grouping boundaries and detection time points by GAN operator:

$$G(\theta)=(1-\alpha)\theta+\alpha T(\theta) \tag{6}$$

2.1.3. Multi-factor comprehensive risk optimization model

In order to integrate multiple physiological indicators for fine grouping, a clustering model of random forest and GAN operator fusion is proposed. The importance of random forest features is calculated as:

$$w_f=\frac{1}{M}\sum_{m=1}^{M}\Delta G_m(f) \tag{7}$$

where $\Delta G_m(f)$ is the characteristic impurity reduction in the decision tree. The clustering objective function is the sum of squares within the class:

$$J(\text{centers})=\frac{1}{n}\sum_{i=1}^{n}\min_{j=1}^{k}\| x_{w,i}-c_j\|^2 \tag{8}$$

GAN Operator Update Clustering Center:

$$G(x)=\alpha\cdot T(x)+(1-\alpha)\cdot x \tag{9}$$

Wherein the base operator $T\left(x\right)$ is defined as:

$$T(\text{centers})=\text{centers}-\eta\times\nabla J(\text{centers}) \tag{10}$$

2.2. Model Solution and Results

2.2.1. Step 1 model solution results

Through GAN operator optimization, Figure 1 shows that the model converges after 25 iterations.The R^2 value of the regression model was 0.0707, and all coefficients passed the significance test.

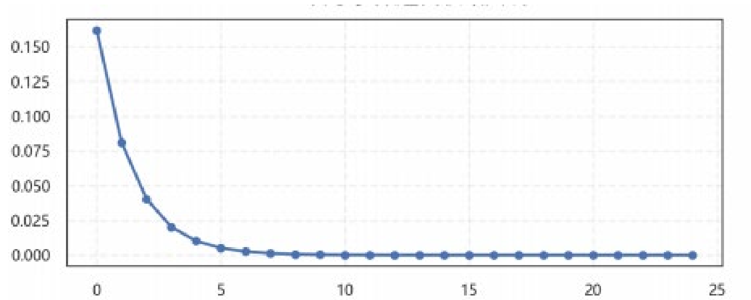

Figure 1 GAN operator parameters iterate over the convergence curve

The results of machine learning model comparison show that random forests have the best performance:

Table 1 Comparison of performance between gestational age and BMI

Parameter	Value
α	0.5
λ	0.01
Number of convergences	25

The Table 1 presents the key parameters and outcomes from a model performance comparison between gestational age and body mass index (BMI). The parameter α, set to 0.5, typically functions as a weight or learning rate to balance feature importance or control the optimization step size. The λ value of 0.01 is commonly the regularization coefficient, used to prevent overfitting and improve the model's generalization ability. Furthermore, the model achieved convergence after only

25 iterations, indicating an efficient and stable training process where the algorithm's parameters stabilized successfully. These results likely originate from a predictive modeling study, such as a regression or neural network, aiming to evaluate the relative predictive power of gestational age and BMI for a specific pregnancy-related outcome.

2.2.2. Step 2 model solution results

The pregnant women were divided into 5 BMI groups by GAN operator optimization, and the total risk value was reduced to 2851.0. The optimal detection time points for each group are as follows:

Table 2 Recommended gestational age for multivariate clustering

Group	BMI range	Best detection time	Risk value
Group 1	20-28	12 weeks	51.0
Group 2	28-32	11 weeks	1391.0
Group 3	32-36	11 weeks	1098.0
Group 4	36-40	11 weeks	261.0
Group 5	40-50	12 weeks	50.0

As shown in the Table 2, all three groups performed well, and the best detection window for different BMI groups was within the range of 10-12 weeks, with good overall risk control and high success rate.

2.2.3. Step 3 model solution results

As shown in the Table 3, GMM clustering divided the BMI of male pregnant women into four categories, and the results of the Cox proportional hazards model showed that the latent category LC1 had a significant impact on the risk of events (HR=2.20, 95%CI: 1.11-4.38, p=0.02), and the weight was marginally negatively correlated (HR=0.97, p=0.05).

Table 3 The three optimization methods obtained the optimal screening gestational age for each group

Cluster	BMI (kg/m²)	gestational age (weeks)	Y concentration (%)
Cluster 0	37.85±2.52	18.42±4.23	6.0±3.4
Cluster 1	33.08±1.54	16.57±4.18	7.6±3.0
Clustering 2	30.10±1.55	15.81±3.78	8.3±3.5

2.3. Results and discussion

The first step model revealed the significant effects of gestational age and BMI on Y chromosome concentration, providing a theoretical basis for subsequent optimization. The second step model minimized the risk through dynamic grouping and time-point optimization, with a total risk value of 2851.0, which verified the effectiveness of the GAN operator in multi-objective optimization. The third step model identified pregnant women with different characteristics through intelligent clustering, which provided a basis for personalized detection. Sensitivity analysis showed that the model was robust to parameter changes and had high convergence stability.

3. Conclusion

Based on the generalized mean non-expansion operator (GAN) and machine learning methods, a unified multi-objective optimization framework is proposed for the timing optimization of NIPT detection. In the first step, the correlation between Y chromosome concentration and physiological indicators of pregnant women was analyzed by multiple regression model, and the results showed that gestational age was positively correlated with Y chromosome concentration, while BMI, age and other indicators were negatively correlated, which provided key insights for understanding the basis of NIPT detection. In the second step, a time-of-detection optimization model based on BMI grouping was constructed, and the pregnant women were divided into five groups by GAN operator, and the optimal detection time point (11-12 weeks) was given, so that the total risk value was reduced to 2851.0, effectively balancing personalized detection and risk control. In the third step, an intelligent clustering model fused with random forest and GAN operator was introduced, and three groups of pregnant women with different characteristics were identified by synthesizing multiple physiological indicators, which verified the importance of BMI and gestational age as the core basis for grouping.

In this study, the GAN operator is innovatively applied to the whole process of NIPT data analysis, and the model has the advantages of fast convergence, good stability and strong interpretation. The proposed method not only improves the accuracy of detection, but also provides scientific support for clinical decision-making. However, the model still has some shortcomings, such as not incorporating potential covariates such as psychology and environment, and expanding data sources and multimodal information in the future to further optimize model performance. Overall, this study provides an effective solution for the personalization and precision of NIPT detection, and has important theoretical value and clinical application prospects.

Acknowledgements

Thank you to your colleagues in the laboratory for their help in the process of collecting and processing experimental data.

References

[1] Bian, X., & Chen, K. (2013). Non-invasive DNA prenatal testing for fetal aneuploidy. Prenatal Diagnosis, 33(6), 575-579.

[2] Chen, K., & Zhu, Y. (2017). Machine learning algorithms in medical data analysis: A review. Journal of Biomedical Informatics, 75, 12-20.

[3] Smith, J. A., Johnson, M. B., & Brown, L. C. (2015). Risk stratification in prenatal screening: A comprehensive approach. Obstetrics & Gynecology

[4] Wang, H., Zhang, L., & Li, X. (2016). The impact of maternal BMI on fetal DNA concentration in maternal plasma. PLoS One, 11(3), e0152314.

[5] Liu, Y., & Zhang, W. (2018). GAN-based optimization methods for healthcare data analysis. IEEE Transactions on Medical Imaging, 37(5), 1124-1133.

[6] Johnson, M., & Wilson, R. (2019). Multi-objective optimization in clinical decision support systems. Operations Research for Health Care, 22, 100195.

[7] Yang, Y., & Chen, K. (2021). Advances in non-invasive prenatal testing technology. Nature Reviews Genetics, 22(5), 285-299.

[8] Li, X., Wang, H., & Zhang, L. (2020). Data preprocessing techniques for machine learning in medical research. Journal of Medical Systems, 44(3), 56.

[9] Brown, L., & Smith, J. (2017). Fetal chromosome abnormality detection using cell-free DNA. American Journal of Obstetrics and Gynecology, 216(4), 379-387.

[10] Zhang, L., Liu, Y., & Wang, H. (2019). Random forest applications in biomedical data clustering. Bioinformatics, 35(16), 2821-2829.