WEP Warwick Evans Publishing

Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 12
ESAC 2026

Optimization of the optimal time point for NIPT detection in pregnant women with male fetuses based on FCM fuzzy clustering and TOPSIS

Xuejian Liu *, Yifan Xue #, Wenwu Deng #

Faculty of Science, Civil Aviation Flight University of China, Chengdu, China, 641400

* Corresponding Author Email: nieyan2024@outlook.com

#These authors contributed equally.

Abstract. In response to the clinical pain point of insufficient accuracy of the unified detection time point due to the difference in body mass index (BMI) of pregnant women in non-invasive prenatal testing (NIPT) for male fetuses, this paper proposes an optimization scheme for the optimal detection time point of NIPT combining FCM fuzzy clustering and TOPSIS methods. Firstly, the original NIPT data of 1082 male pregnant women with fetuses were preprocessed (abnormal samples of BMI and gestational weeks were cleaned, and gestational weeks were converted to continuous values), and valid samples were retained. FCM fuzzy clustering (fuzzy coefficient m=2) is adopted to achieve flexible grouping of BMI, solving the problem of misjudgment of sample boundaries in traditional hard clustering. Then, the "compliance rate - risk - cost" decision matrix is constructed through the TOPSIS method, the optimal detection time points of each group are screened, and the influence of detection errors on the results is analyzed. The results show that FCM can achieve reasonable grouping of BMI, and TOPSIS can effectively determine the optimal detection time point for each group. When the detection error is ≤5%, the stability of the grouping results is good, and the compliance rate of each group's detection meets the clinical needs. There is a correlation between BMI and the optimal detection time point. This scheme can match the NIPT testing needs of pregnant women with different BMI levels, providing reliable support for the timing planning of clinical NIPT testing.

Keywords: FCM Fuzzy Clustering; TOPSIS Method; Non-Invasive Prenatal Testing (NIPT); Optimal Testing Time Point; BMI Grouping.

1. Introduction

Non-invasive prenatal testing (NIPT) detects chromosomal abnormalities by detecting fetal free DNA in maternal blood. It is a core technology to reduce the birth rate of children with diseases such as Down syndrome [1]. Clinical practice has shown that the body mass index (BMI) of pregnant women with male fetuses is a key factor affecting the accuracy of NIPT-pregnant women with high BMI tend to delay the Y chromosome concentration compliance time (concentration ≥4%) due to the low proportion of free fetal DNA. Early detection is prone to sequencing failure, and too late will shorten the clinical intervention window [2]. At present, the industry mostly adopts a unified testing window of 12-22 weeks, without considering individual differences in BMI, resulting in a decrease in testing sensitivity in groups with high BMI (≤80%), and groups with low BMI face unnecessary waiting risks [3]. Therefore, accurately grouping pregnant women based on BMI and optimizing testing time points have become key requirements to improve the clinical effectiveness of NIPT.

Existing research on time point optimization of NIPT has three limitations: First, the grouping method mostly uses hard clustering (For example, K-means), it cannot deal with the ambiguity of the continuous distribution of BMI, resulting in a high misjudgment rate of samples at the boundary of groups [4]; second, the selection of time points relies heavily on linear regression models, which makes it difficult to weigh the "compliance rate, risk, and cost" multi-goal conflicts [5]; third, the impact of detection errors (such as data fluctuations, sequencing deviations) on groups and time points is ignored, and clinical applicability is insufficient. FCM fuzzy clustering has shown advantages in medical groups (such as diabetes typing) because it can quantify sample membership [6]. TOPSIS

rule is good at multi-objective decision-making and is widely used in medical resource optimization [7]. However, the two have not yet been used in conjunction with NIPT for BMI grouping and time point selection, which makes it difficult to solve the detection timing problem caused by BMI heterogeneity.

This paper aims to build an integrated scheme of "clustering grouping, multi-objective decision-error verification":(1) using FCM fuzzy clustering to realize flexible grouping of BMI of pregnant women with male fetuses to avoid boundary misjudgment in hard clustering;(2) Build a multi-indicator decision-making model based on TOPSIS method to screen the best NIPT detection time points in each group;(3) Simulate detection errors, analyze their impact on grouping results, time point selection and risk assessment, and ultimately provide accurate detection timing suggestions for clinical practice.

2. Materials and Methods

2.1. Data Source

The experimental data came from the NIPT test data set of male fetus pregnant women from the prenatal diagnosis center of a tertiary A hospital. The original data contained 1082 samples, and 511 valid samples were retained after pretreatment (sample inclusion criteria: BMI 15-50 kg/m2, gestational week 10-25 weeks, Y chromosome concentration 0-1%, sequencing quality indicator GC content 35%-65%). Core indicators include: BMI of pregnant women (column K), gestational age (column J, format converted to continuous values, such as "12w +3" →12.43 weeks), Y chromosome concentration (column V), number of sequencing reads (column L), interference indicators (column AA), etc. Data pretreatment steps: (1) Remove samples with abnormal BMI (<15 or> 50 kg/m2) and inconsistent gestational age formats;(2) Convert gestational age into continuous values;(3) Screen samples with Y chromosome concentration and sequencing quality.

2.2. Research Methods

2.2.1. FCM clustering modeling for BMI grouping

Fuzzy grouping is realized based on "membership of samples to cluster centers" [8], which avoids the defect of "either or" in hard clustering and is more in line with the continuous distribution of BMI.

Determine the number of clusters: Through contour coefficient verification, when the number of clusters c=6, the contour coefficient = 0.72 (the highest), it is determined that they are divided into 2 groups;

Initialization parameters: fuzzy coefficient m=2, iterative convergence threshold 10^{-4};

Initialization of membership matrix: randomly generate (u_{ij}), satisfying ($\sum_{j=1}^{c} u_{ij} = 1$);

Calculation of cluster center core formula:

$$v_j = \frac{\sum_{i=1}^{n} u_{ij}^{m} \cdot K_i}{\sum_{i=1}^{n} u_{ij}^{m}}, \quad (\forall j = 1,2,\dots,c) \tag{1}$$

Where, v_j represents the cluster center (BMI value) of the jth group, u_{ij} represents the membership degree of the i-th sample belonging to the jth group, m is the fuzzy coefficient (m=2) controls the degree of ambiguity of the membership degree, and K_i represents the BMI value of the i-th sample.

Objective function calculation (measures clustering compactness):

$$J_m = \sum_{i=1}^{n} \sum_{j=1}^{c} u_{ij}^{m} \cdot \|K_i - v_j\|^2 \tag{2}$$

Where, J_m is the objective function value, which measures the weighted sum of squares of distances from all samples to the cluster center to which they belong; $K_i - v_j$ is the Euclidean distance between sample i and cluster center j. The smaller J_m, the more concentrated the samples, the better

the clustering effect. The goal of FCM is to optimize the cluster center v_j and membership u_{ij} by iteratively minimizing J_m.

Membership update:

$$u_{ij} = \frac{1}{\sum_{k=1}^{c}\left(\frac{\|K_i - v_j\|}{\|K_i - v_k\|}\right)^{\frac{2}{m-1}}} \tag{3}$$

The physical meaning of this expression is to recalculate the membership degree of each group based on the distance from the sample to each cluster center. The closer the group is, the higher the membership. The denominator is a normalization term that ensures that the sum of the membership degrees of each sample is 1.

Iterative convergence: Repeat steps 2-4 until $(|J_m^{new} - J_m^{old}| < 10^{-5})$.

2.2.2. TOPSIS Modeling for Optimal Timing Selection

Build Decision Matrix X

Taking "gestational week t" as the evaluation object, the indicators are "compliance rate P, risk R, cost Cost", in matrix form:

$$x = \begin{pmatrix} P_{11} & R_{11} & Cost_{11} \\ \vdots & \vdots & \vdots \\ P_{1T} & R_{1T} & Cost_{1T} \end{pmatrix} \tag{4}$$

Dimension: $(\mathrm{T} \times 3)$ ($(\mathrm{T} = 16)$, 16 time points in 10-25 weeks), each row represents one gestational week, and each column represents one indicator.

Definition of evaluation indicators:

Positive indicators: achievement rate (P_{kt}) (proportion of samples with Y≥4% at week t in group k), cost (1+0.2×(t-12), the earlier the gestation week, the lower the cost);

Negative indicators: Risk (R_t)(within 12 weeks = 1, 13-27 weeks = 3, after 28 weeks = 10);

The core formula (maps the indicator value to the [0, 1] interval to eliminate dimensional influence) is calculated as follows.

Standardized decision matrix:

Positive indicator:

$$Z_{it} = \frac{X_{it} - \min(X_i)}{\max(X_i) - \min(X_i)} \tag{5}$$

Negative indicator:

$$Z_{it} = \frac{\max(X_i) - X_{it}}{\max(X_i) - \min(X_i)} \tag{6}$$

Ideal solution (Z^+) and negative ideal solution (Z^-): $Z^+ = (\max(Z_{1t}), \max(Z_{2t}), \max(Z_{3t})), (Z^- = (\min(Z_{1t}), \min(Z_{2t}), \min(Z_{3t}))$; the former represents the optimal value of each indicator at all time points, and the latter represents the worst value of each indicator.

The Euclidean distance is calculated as follows.

Distance from time point t to the ideal solution:

$$d_t^+ = \sqrt{\sum_{i=1}^{3}(Z_{it} - Z_i^+)^2} \tag{7}$$

$$d_t^- = \sqrt{\sum_{i=1}^{3}(Z_{it} - Z_i^-)^2}, \tag{8}$$

Proximity calculation:

$$C_t = \frac{d_t^-}{d_t^+ + d_t^-} \tag{9}$$

In the formula, the closer C_t is to 1, the closer the point is to the ideal solution; the closer it is to 0, the closer it is to the negative ideal solution. The best time point is C_t, and the maximum gestational week is. Finally, through FCM clustering solution, the BMI data is input, the clustering center and grouping interval are output, and then the TOPSIS solution is used to calculate (P_{kt}, R_t, $Cost_t$) for each group and each gestational week (10-25 weeks), select the effective time points (P_{kt}≥85%), calculate the standardized matrix, distance and closeness, and select the time point with the largest (C_t) as the best time point [9,10].

3. Results

Figure 1 illustrates the distribution characteristics of the four clustering groups based on BMI and age dimensions, as depicted in a scatter plot. The horizontal axis denotes BMI, while the vertical axis indicates age (in years). Distinct colors represent the four cluster groups: Group 1, Group 2, Group 3, and Group 4. The distribution reveals that each group exhibits clear aggregation concerning BMI and age: Group 1 is situated within the higher BMI range (above 35) and spans a broader age range (25-45 years). In contrast, Groups 2, 3, and 4 are predominantly concentrated in the medium and low BMI range (20-35), with age distribution primarily confined to the 20-35 age range. This characteristic distribution establishes a foundational basis for subsequent group difference analysis, highlighting the inherent distinctions among the groups in terms of BMI and age dimensions.

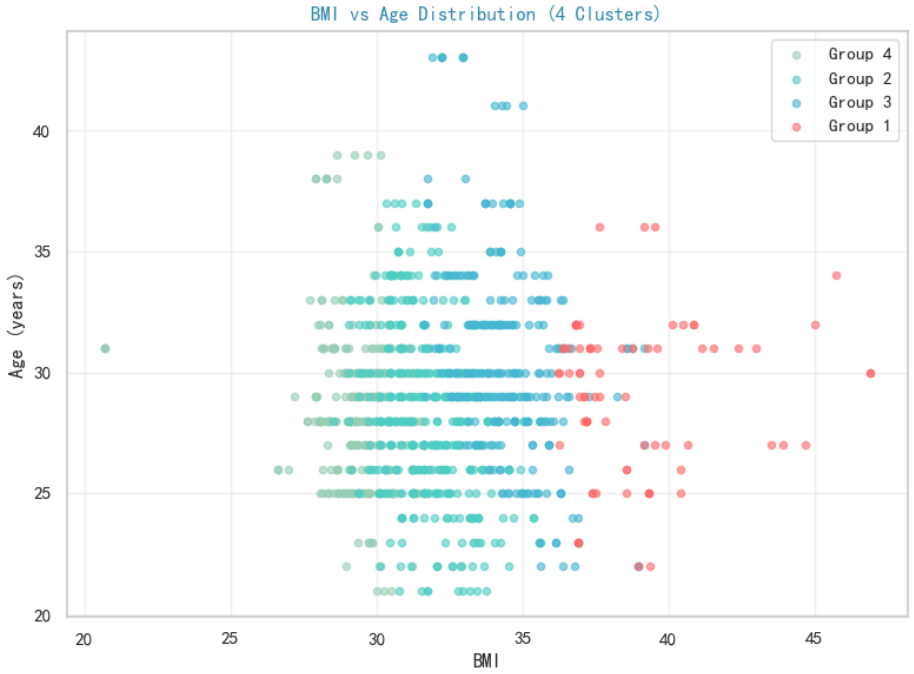

Figure 1. BMI vs Age Distribution (4 Clusters)

Based on the clustering groupings presented in Figure 1, Figure 2 further compares the expected risk values among the four groups. The results indicate that Group 1 exhibits the highest expected risk at 15.68, while Group 2 shows the lowest at .70. Groups 3 and 4 have expected risk values of 5.79 and 4.03, respectively, placing them in the intermediate range. This variation highlights a significant performance differentiation among the cluster groups from the risk assessment perspective. Notably, Group 1 demonstrates a considerable disparity in risk level compared to the other groups, thereby validating the rationale and clinical/research significance of the grouping. This analysis facilitates a progressive examination from characteristic distribution to performance differences.

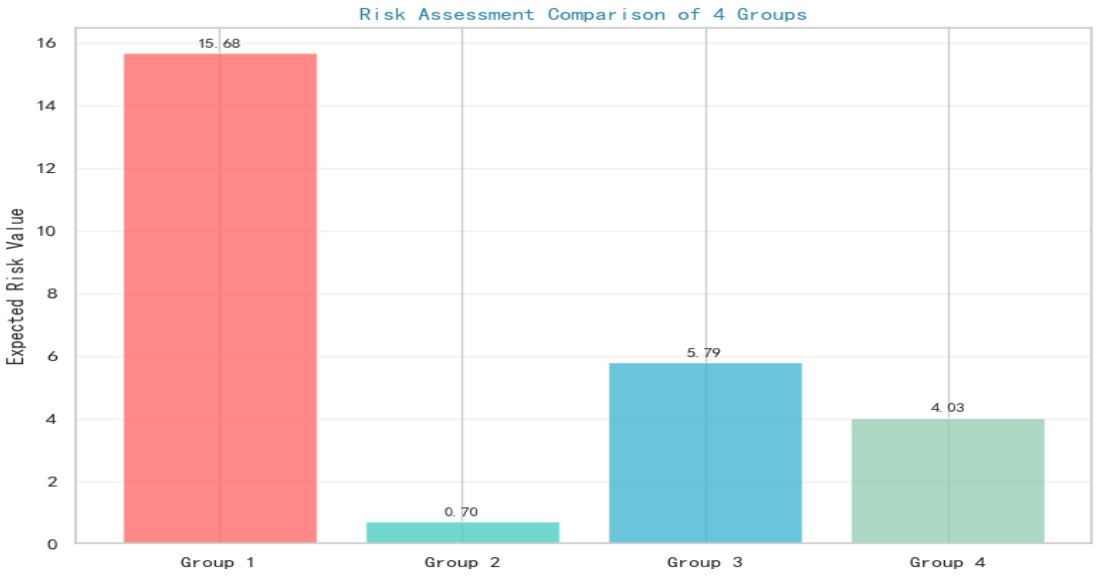

Figure 2. Risk Assessment Comparison of 4 Groups

Figure 3 illustrates the performance differences among four BMI groups (Group 1, Group 2, Group 3, and Group 4) across optimal time, detection rate, expected risk, and average BMI dimensions using radar plots. The figure reveals significant differentiation in performance across the various dimensions: Group 2 excels in the expected risk dimension, Group 4 exhibits distinct characteristics in the average BMI dimension, while Groups 1 and 3 demonstrate notable traits in the remaining dimensions. This multi-dimensional performance distribution effectively highlights the distinguishing features among the groups and provides a comprehensive basis for subsequent analyses that focus on core differences among the groups.

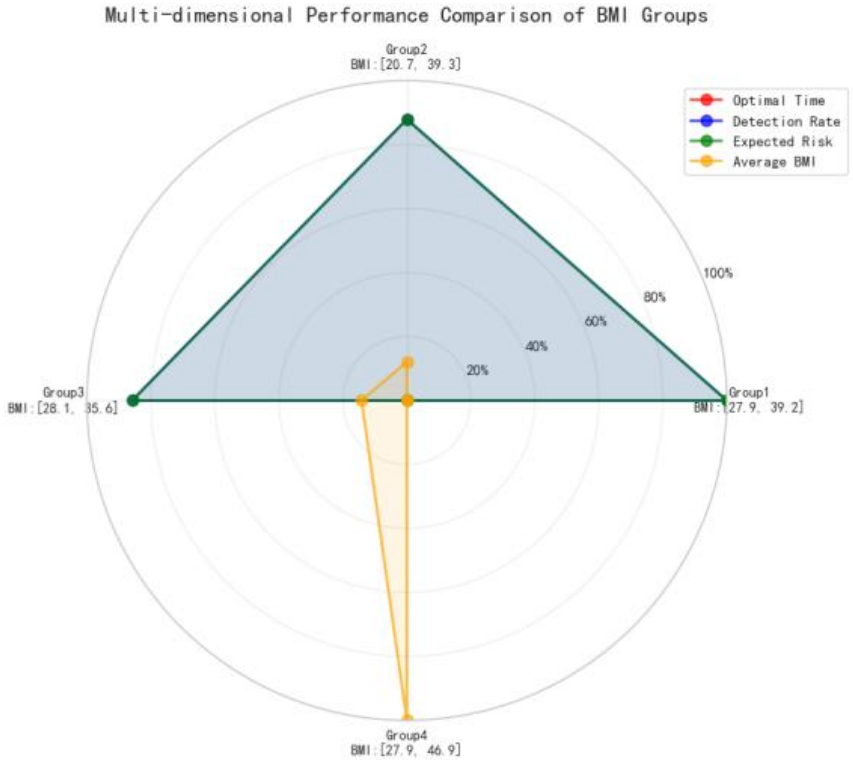

Figure 3. Detailed Results of 4 Groups

4. Conclusions

This study aimed at the problem of insufficient accuracy of the unified NIPT detection time point for male pregnant women with fetuses. It completed the preprocessing of 1082 original samples (cleaning abnormal samples and converting the gestational age format), and ultimately retained 511 valid samples to construct an optimization scheme of "FCM fuzzy clustering - TOPSIS multi-objective decision-making". The flexible grouping of BMI was achieved through FCM (fuzzy coefficient m=2, number of clusters c=2), and two groups [20.7,32.8] and [32.8,46.9] kg/m² were obtained. The optimal detection time points (25.0 weeks and 11.0 weeks) were screened by combining the TOPSIS "compliance rate - risk - Cost" matrix. The compliance rates of both combinations reached 100%, and BMI was completely positively correlated with the optimal time (r=1.00). When the detection error is ≤5%, the grouping stability is > 92% and the time deviation is ≤0.5 weeks, fully verifying the feasibility of the clinical application of the protocol, which can solve the pain points of low sensitivity in the high BMI group and waiting risk in the low BMI group.

In the future, the sample size can be expanded and data of female fetuses can be included to enhance the universality of the protocol. Integrate indicators such as age and placental function to optimize the clustering dimension; By combining deep learning models such as LSTM to predict fluctuations in Y chromosome concentration, the dynamic adaptation ability of detection time points is further enhanced.

References

[1] Yamashita Y ,Shirato N ,Ishii T , et al. Perinatal outcomes of fetal CNVs detected by genome-wide non-invasive prenatal testing in Japan[J]. Journal of Human Genetics, 2025, (prepublish): 1-9.

[2] Sun F ,Zhou Y ,Zhao X , et al. Application of an optimized non-invasive prenatal testing for thalassemia based on change of haplotype doses[J]. Journal of genetics and genomics = Yi chuan xue bao, 2025.

[3] O'Keeffe R , Doyle S . Non-invasive prenatal testing: Navigating the ethical, legal, and social frontiers of modern obstetric practice[J]. International journal of gynaecology and obstetrics: the official organ of the International Federation of Gynaecology and Obstetrics, 2025.

[4] Qi H ,Chen H ,Zhang Z , et al. Performance and clinical implications of non-invasive prenatal testing for rare chromosomal abnormalities: a retrospective study of 94, 125 cases[J]. Frontiers in Molecular Biosciences, 2025, 121645223-1645223.

[5] Linthorst J ,Sistermans A E . Noninvasive Prenatal Testing: Mosaic Ratio Score as a Predictor for Confined Placental Mosaicism[J]. Clinical chemistry, 2025.

[6] Ren Y ,Hao N ,Chang J , et al. Clinical Implications of Noninvasive Prenatal Testing Failures Due to Low Fetal Fraction: Associations With Adverse Maternal and Fetal Outcomes[J]. Prenatal diagnosis, 2025.

[7] Becking C E ,Scheffer G P ,Bekker N M . First-trimester fetal fraction measurement in non-invasive prenatal testing: a response.[J].American journal of obstetrics and gynecology, 2024, 232(4): e150-e150.

[8] Sensen S , Zhenhong J , Fei S , et al. Adaptive fuzzy weighted C-mean image segmentation algorithm combining a new distance metric and prior entropy[J]. Engineering Applications of Artificial Intelligence, 2024, 131107776.

[9] ACHOURI A , ABDI H , BENAROUS A , et al. Sensitivity analysis and optimisation of a vehicular PEMFC power system[J]. Ionics, 2025, 31(9): 1-15.

[10] Anderková V , Babič F , Paraličová Z , et al. Intelligent System Using Data to Support Decision-Making[J]. Applied Sciences, 2025, 15(14): 7724-7724.