Am J Transl Res 2025;17(4):2541-2551
www.ajtr.org /ISSN:1943-8141/AJTR0161626

Original Article

Association between elevated maternal serum MSX1 and IRF6 levels and fetal orofacial clefts: implications for clinical prediction

Juan Cao, Bailei Zhang, An'er Chen

Department of Obstetrics, The Affiliated Women and Children's Hospital of Ningbo University, Ningbo 315000, Zhejiang, China

Received October 31, 2024; Accepted March 14, 2025; Epub April 15, 2025; Published April 30, 2025

Abstract: Objective: To investigate the association between maternal serum expression levels of Msh homeobox 1 (MSX1) and interferon regulatory factor 6 (IRF6) and fetal cleft lip and palate (CLP), as well as their potential role in clinical prediction. Methods: A prospective case-control study was conducted. A total of 100 pregnant women carrying fetuses diagnosed with CLP via prenatal screening and diagnosis at The Affiliated Women and Children's Hospital of Ningbo University from July 2021 to June 2024 (CLP group) and 105 pregnant women with healthy fetuses (normal group) were selected as the research subjects. Clinical data and relative expression levels of MSX1 and IRF6 were collected. Binary logistic regression was used to analyze the influencing factors of fetal CLP. A Gradient Boosting Machine (GBM) model was developed using R language 4.4.1 software to predict fetal CLP and evaluate its predictive performance. Results: The serum levels of MSX1 and IRF6 in the CLP group were higher than those in the normal group ($t = 6.536, 9.907$; both $P < 0.001$). Independent influencing factors for fetal CLP included maternal age, family history of CLP, malnutrition during pregnancy, and expression of MSX1 and IRF6. A GBM model was constructed based on these factors, with their relative importance ranked as follows: IRF6 > MSX1 > family history of CLP > maternal malnutrition > age. In the training set, the GBM model achieved an area under the curve (AUC) of 0.898 (95% CI: 0.849, 0.948), with a sensitivity of 79.2% and specificity of 88.7%. In the validation set, the AUC was 0.895 (95% CI: 0.818, 0.973), with sensitivity of 85.7% and the specificity of 82.4%. The calibration curve demonstrated good agreement between predicted and actual probabilities. Decision curve analysis showed a threshold probability range of 0.30-0.72 in the training set and 0.30-0.73 in the validation set. Conclusion: Elevated maternal serum MSX1 and IRF6 expression levels are closely associated with fetal CLP and may serve as potential biomarkers for its clinical prediction.

Keywords: Maternal, MSX1, IRF6, fetus, cleft lip and palate, association, prediction

Introduction

Fetal cleft lip and palate (CLP) is a common congenital birth defect, imposing a heavy burden on families and society. Beyond its impact on facial appearance, CLP can lead to various consequences such as feeding difficulties, speech disorders, hearing impairment, and psychological problems [1]. Globally, the incidence of fetal CLP is about 1 in every 500 newborns [2].

Despite extensive research, the etiology of fetal CLP remains incompletely understood. Genetic factors are believed to play a crucial role, with mutations or abnormal expression of multiple genes implicated in its pathogenesis [3]. In addition, environmental factors such as maternal infections, drug exposure, malnutrition, and contact with harmful substances during pregnancy may also increase the risk of fetal CLP [4]. However, the specific pathogenic mechanisms are not yet fully understood. Currently, prenatal ultrasound examination is a commonly used diagnostic tool for fetal CLP. However, it has limitations, including potential missed diagnoses in cases of mild CLP and variability in accuracy due to factors such as operator's experience and fetal position [5]. Furthermore, there is a lack of reliable biomark-

https://doi.org/10.62347/ZDZC2364

ers for early risk assessment and prediction of fetal CLP.

Msh homeobox 1 (MSX1) and interferon regulatory factor 6 (IRF6) are key regulatory genes involved in craniofacial development, influencing multiple downstream targets essential for facial tissue formation [6]. MSX1 plays a critical role in palatogenesis, and its mutation or abnormal expression has been linked to CLP [7]. Similarly, IRF6 mutation have been associated with an increased risk of fetal CLP [8]. Given their roles in craniofacial development, the detection of MSX1 and IRF6 expression levels in maternal serum may provide a novel approach for early diagnosis and risk assessment of fetal CLP.

This study aims to investigate the association between maternal serum MSX1 and IRF6 expression levels and fetal CLP, as well as their potential for clinical prediction. By improving early risk assessment, this research seeks to aid clinicians in timely intervention and optimizing fetal outcomes.

Materials and methods

Patient recruitment

A prospective case-control study was conducted. A total of 100 pregnant women carrying fetuses diagnosed with CLP through prenatal screening and diagnosis at The Affiliated Women and Children's Hospital of Ningbo University between July 2021 and June 2024 (CLP group) and 105 pregnant women with healthy fetuses (normal group) were recruited as the research subjects.

(1) Inclusion criteria: 1) Regular attendance at all scheduled prenatal check-ups; Timely submission of required biological samples (including blood samples for MSX1 and IRF6 expression analysis); Accurate completion of all study-related questionnaires and surveys. 2) Confirmed intrauterine singleton pregnancy verified by ultrasound examination. 3) Availability of complete serum MSX1 and IRF6 expression data; Samples processed using standardized laboratory protocols with strict quality control measures.

(2) Exclusion criteria: 1) A history of teratogenic medication use during current pregnancy. Teratogenic drugs are known to increase the risk of birth defects, which could potentially confound the relationship between MSX1 and IRF6 expression levels and fetal CLP. Common teratogenic drugs include thalidomide, isotretinoin, and some antiepileptic medications. 2) Pregnant women with pre-existing liver diseases (such as hepatitis B, cirrhosis), kidney diseases (such as chronic kidney failure, nephrotic syndrome), or heart diseases (such as congenital heart disease, coronary artery disease). These medical conditions can affect the physiological state of the pregnant woman and potentially interfere with the expression levels of MSX1 and IRF6, as well as the development of the fetus. 3) Pregnant women with cognitive impairment, such as dementia or severe intellectual disability, or a history of mental illness (such as schizophrenia, bipolar disorder). These conditions may affect the pregnant woman's ability to provide accurate information, comply with the study procedures, or understand the implications of the informed consent. 4) Pregnant women carrying fetuses with other congenital birth defects in addition to CLP. The presence of other birth defects could complicate the analysis and make it difficult to isolate the specific association between MSX1 and IRF6 expression levels and fetal CLP.

This study was approved by the Ethics Committee of The Affiliated Women and Children's Hospital of Ningbo University.

Establishment of gradient boosting machine (GBM) model

The 205 cases were randomly divided into a training set of 143 cases and a validation set of 62 cases in a ratio of 7:3. A GBM prediction model for fetal CLP was developed using R 4.4.1 software, and its predictive efficacy was evaluated.

The sequential backward search method was used for feature selection. Variables with the least impact were iteratively removed from the full-feature model until the remaining features met a predefined threshold for model performance. Feature selection and model optimization were completed based on the training set. To enhance model robustness, 10-fold cross-validation was performed within the training set, and internal validation was conducted. Grid search was used to determine the

optimal hyperparameters. The established model was verified in an independent test set. Feature selection, model construction, and verification were completed using the R packages "mlr" and "gbm".

Detection of relative expression levels of MSX1 and IRF6 genes

A 2 mL venous blood sample was collected from each pregnant women, and serum was separated. The relative expression levels of MSX1 and IRF6 were determined using fluorescence quantitative polymerase chain reaction (PCR). Total serum ribose nucleic acid (RNA) was extracted following manufacturer's instructions, ensuring strict precautions against RNase contamination. The purity and concentration of RNA were assessed using an ultraviolet spectrophotometer. RNA was then reverse-transcribed into complementary DNA (cDNA) using a reverse transcription kit, with β-actin serving as the internal reference gene. The primer sequence of forward primer (5'→3') is MSX1: 5'-AGCAGCTGCTGCTGCTGC-TG-3', IRF6: ATGTGCCCCATCACTTGTTG, β-action: TGGCACCCAGCACAATGAA. The primer sequence of reverse primer (5'→3') is MSX1: CTGCTGCTGCTGCTGCTGCTG, IRF6: GGTGGCT-GCTTCTCTATCTG, β-action: CTAAGTCATAGTCC-GCCTAGAAGCA. The relative expression levels of MSX1 and IRF6 genes was calculated using 2-ΔΔCt method.

Data collection

Clinical data of pregnant women were collected from electronic medical records, including age, parity, gestational age, family history of CLP, smoking history, drinking history, abortion history, educational level, malnutrition during pregnancy, anemia during pregnancy, trauma during pregnancy, early pregnancy infection, medication use during early pregnancy, hyperemesis gravidarum, and concurrent gestational complications.

Smoking history is defined as cigarette smoking before (at least one pack of cigarettes daily for over one year) and during pregnancy (continued cigarette use after pregnancy confirmation until delivery).

Drinking history is defined as alcohol consumption before (consuming liquor or other spirits ≥ 10 times per month) and during (continued intake of any alcoholic beverage) pregnancy.

Trauma during pregnancy is defined as physical injuries sustained during pregnancy, such as abrasions, contusions, fractures and other external injuries resulting from accidental falls, collisions, or traffic accidents. It also encompassed internal organ trauma, such as injuries to abdominal organs [9].

Early pregnancy infection is defined as an infection occurring between 1 and 12 weeks of pregnancy, caused by viruses, bacteria, chlamydia, mycoplasma, or other pathogens [10].

Diagnostic criteria for hyperemesis gravidarum [11]: positive urine pregnancy test; frequent nausea and vomiting, leading to an inability to eat or drink, lethargy, and excessive sleepiness; electrolyte imbalance and metabolic disturbances; positive urine ketone body test; exclusion of nausea and vomiting caused by gastrointestinal infections, cholecystitis, pancreatitis, or neurological diseases. Concurrent gestational complications included gestational hypertension, diabetes, and kidney disease.

Data analysis methods

Statistical analyses were conducted using SPSS 23.0. Quantitative data were expressed as mean ± standard deviation ($\bar{x}$ ± sd), and group comparisons were performed using the t-test. Categorical data were presented as n (%), and comparisons between groups were analyzed using the chi-square test. The goodness of fit of the GBM model was assessed using a calibration curve, generated through 1000 bootstrap resampling iterations. The predictive performance of the GBM model for fetal cleft lip and palate was evaluated using the receiver operating characteristic (ROC) curve. A *P*-value < 0.05 was considered statistically significant.

Results

Differences in basic characteristics between the CLP and normal groups

The proportions of pregnant women aged > 34 years, with a family history of CLP, malnutrition during pregnancy, and anemia during pregnancy were significantly higher in the CLP group

Table 1. Comparison of basic characteristics between pregnant women with and without CLP

Basic characteristics	CLP group (n = 100)	Normal group (n = 105)	χ^2/t	*P*
Age (years)			20.117	< 0.001
< 24 years old	15 (15.00%)	21 (20.00%)		
24-34 years old	28 (28.00%)	56 (53.33%)		
> 34 years old	57 (57.00%)	28 (26.67%)		
Parity			0.195	0.659
Primipara	44 (44.00%)	43 (40.95%)		
Multipara	56 (56.00%)	62 (59.05%)		
Gestational weeks	37.52 ± 1.03	37.55 ± 1.10	0.204	0.839
CLP family history			15.608	< 0.001
Yes	24 (24.00%)	5 (4.76%)		
No	76 (76.00%)	100 (95.24%)		
Smoking history			0.277	0.599
Yes	20 (20.00%)	18 (17.14%)		
No	80 (80.00%)	87 (82.86%)		
Drinking history			0.273	0.201
Yes	24 (24.00%)	22 (20.95%)		
No	76 (76.00%)	83 (79.05%)		
Abortion history			0.095	0.758
Yes	22 (22.00%)	25 (23.81%)		
No	78 (78.00%)	80 (76.19%)		
Educational level				
Below junior high school	24 (24.00%)	21 (20.00%)		
Junior high school/high school	61 (61.00%)	65 (61.90%)		
Above high school	15 (15.00%)	19 (18.10%)		
Prenatal malnutrition				
Yes	31 (31.00%)	11 (10.48%)	13.244	< 0.001
No	69 (69.00%)	94 (89.52%)		
Anemia during pregnancy				
Yes	21 (21.00%)	8 (7.62%)	7.551	0.006
No	79 (79.00%)	97 (92.38%)		
Trauma during pregnancy			0.488	0.485
Yes	18 (18.00%)	23 (21.90%)		
No	82 (82.00%)	82 (78.10%)		
Early pregnancy infection			0.785	0.376
Yes	21 (21.00%)	17 (16.19%)		
No	79 (79.00%)	88 (83.81%)		
Taking medicine in early pregnancy			0.298	0.585
Yes	15 (15.00%)	13 (12.38%)		
No	85 (85.00%)	92 (87.62%)		
Hyperemesis gravidarum			0.420	0.517
Yes	45 (45.00%)	52 (49.52%)		
No	55 (55.00%)	53 (50.48%)		
Concurrent gestational complications			0.095	0.758
Yes	22 (22.00%)	25 (23.81%)		
No	78 (78.00%)	80 (76.19%)		

Note: CLP, cleft lip and palate.

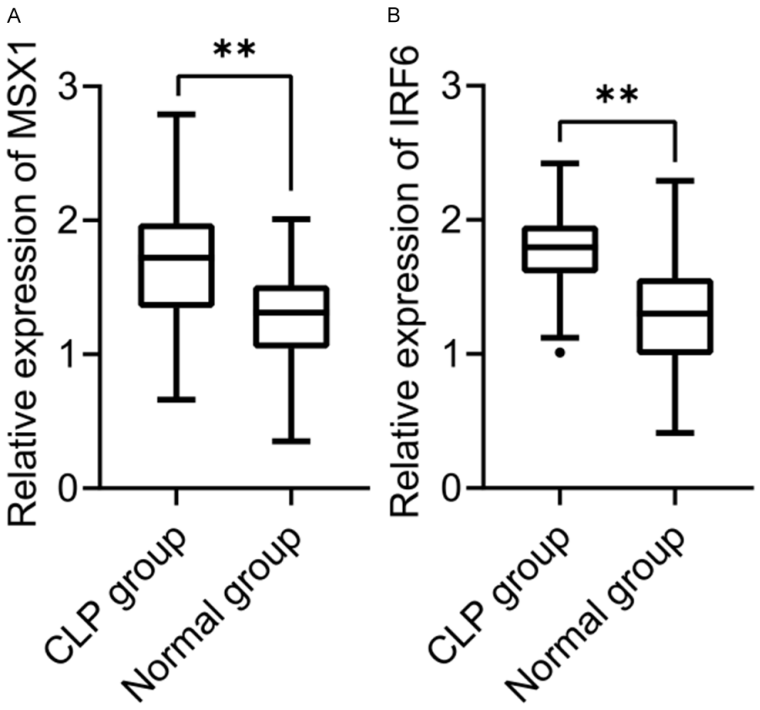

Figure 1. Differences in the maternal serum levels of MSX1 (A) and IRF6 (B) between the two groups. Note: CLP, cleft lip and palate; MSX1, Msh homeobox 1; ***P* < 0.001.

than those in the normal group (all P < 0.05). However, there were no statistically significant differences between the two groups in parity, gestational weeks, smoking history, drinking history, abortion history, educational level, trauma during pregnancy, early pregnancy infection, medication use in early pregnancy, hyperemesis gravidarum, or concurrent gestational complications (all P > 0.05) (**Table 1**).

Differences in serum MSX1 and IRF6 expression levels between the CLP and normal groups

The relative expression levels of MSX1 and IRF6 in the serum of pregnant women in the CLP group were 1.69 ± 0.46 and 1.74 ± 0.28, respectively, significantly higher than 1.32 ± 0.33 and 1.26 ± 0.41 in the normal group (t = 6.536, 9.907, both P < 0.001) (**Figure 1**).

Association between MSX1 and IRF6 expression levels and fetal CLP

A binary logistic regression model was used to analyze the relationship between fetal CLP and the variables that showed significant differences between the two groups (maternal age, family history of CLP, malnutrition during pregnancy, anemia, MSX1, and IRF6). The results showed that maternal age, family history of CLP, malnutrition during pregnancy, MSX1, and IRF6 were independent influencing factors for fetal CLP (all P < 0.05). Moreover, regardless of whether other variables were included in the analysis, MSX1 and IRF6 remained independent influencing factors for fetal CLP (P < 0.05) (**Table 2**).

Predictive ability of MSX1 and IRF6 for fetal CLP

The AUC, sensitivity, and specificity for MSX1 in predicting fetal CLP were 0.755 (95% confidence interval (CI): 0.688-0.822), 69.0%, and 77.1%, respectively. For IRF6, they were 0.833 (95% CI: 0.777-0.889), 78.0%, and 78.1%, respectively. ROC curve analysis demonstrated that MSX1 and IRF6 exhibited superior predictive performance for fetal CLP compared to maternal age (AUC: 0.640), family history of CLP (AUC: 0.596), and malnutrition during pregnancy (AUC: 0.603) (**Figure 2**; **Table 3**).

Predictive mechanism of MSX1 and IRF6 for fetal CLP in the GBM model

A GBM model was developed using independent influencing factors for fetal CLP as predictor variables. The relative importance of each variable was ranked as follows: IRF6 > MSX1 > family history of CLP > malnutrition during pregnancy > age (**Figure 3**).

Predictive performance of the GBM model

The ROC curve was drawn to evaluate the model's predictive performance. The results showed that the AUC of this model in the training set was 0.898 (95% CI: 0.849, 0.948), with sensitivity of 79.2% and specificity of 88.7%, indicating good discrimination (**Figure 4A**); while in the validation set, the model

Table 2. Relationship between MSX1, IRF6 expression and fetal CLP

Variable	β	SE	Wals	*P*	OR (95% CI)
Step[a]					
Age			6.736	0.034	
< 24 years old	-0.944	0.583	2.628	0.105	0.389 (0.124-1.218)
24-34 years old	-1.162	0.460	6.384	0.012	0.313 (0.127-0.771)
CLP family history	2.118	0.723	8.587	0.003	8.316 (2.017-34.288)
Prenatal malnutrition	1.385	0.557	6.178	0.013	3.995 (1.340-11.907)
Anemia during pregnancy	0.649	0.726	0.797	0.372	1.913 (0.461-7.941)
MSX1	2.286	0.577	15.718	< 0.001	9.833 (3.176-30.438)
IRF6	3.436	0.624	30.283	< 0.001	31.058 (9.135-105.592)
Constant	-8.691	1.425	37.218	< 0.001	< 0.001
Step[b]					
Age			6.768	0.034	
< 24 years old	-0.898	0.574	2.443	0.118	0.408 (0.132-1.256)
24-34 years old	-1.168	0.459	6.491	0.011	0.311 (0.127-0.764)
CLP family history	2.179	0.711	9.406	0.002	8.839 (2.196-35.583)
Prenatal malnutrition	1.372	0.552	6.174	0.013	3.945 (1.336-11.647)
MSX1	2.361	0.576	16.812	< 0.001	10.597 (3.429-32.749)
IRF6	3.480	0.626	30.903	< 0.001	32.469 (9.518-110.756)
Constant	-8.821	1.438	37.653	< 0.001	< 0.001
Step[c]					
MSX1	2.345	0.516	20.617	< 0.001	10.434 (3.792-28.712)
IRF6	3.712	0.605	37.601	< 0.001	40.919 (12.494-134.012)
Constant	-9.285	1.301	50.904	< 0.001	0

[a]Input variables: age, family history of CLP, prenatal malnutrition, anemia during pregnancy, MSX1, IRF6; [b]Input variables: age, CLP family history, malnutrition during pregnancy, MSX1, IRF6; [c]Input variables: MSX1, IRF6. Note: CLP, cleft lip and palate; MSX1, msh homeobox 1; IRF6, interferon regulatory factor 6; SE, standard error; OR, odds ratio.

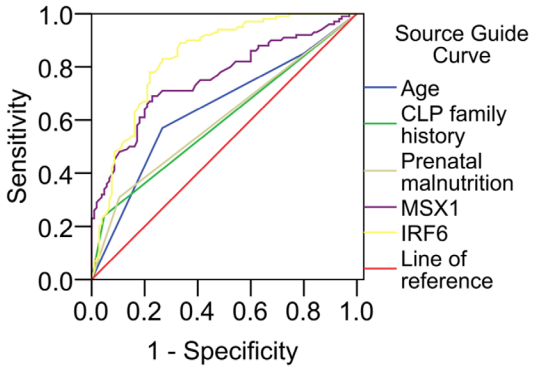

Figure 2. ROC curves for each independent factor in predicting CLP. Note: ROC, receiver operator characteristic; CLP, cleft lip and palate; MSX1, msh homeobox 1; IRF6, interferon regulatory factor 6.

achieved an AUC of 0.895 (95% CI: 0.818, 0.973), with sensitivity of 85.7% and specificity of 82.4%, also indicating good predictive accuracy (**Figure 4B**).

Calibration curve analysis

To evaluate the predictive accuracy of the GBM model, a calibration curve was constructed. The results demonstrated that the model's predicted probabilities closely aligned with the ideal calibration curve, indicating that the predicted probability was highly consistent with the actual probability (**Figure 5**).

Decision curve analysis (DCA)

DCA showed that the model's curve remained consistently higher than the "treat all" and "treat none" strategies across the entire threshold probability range. This suggests that the GBM model provides a positive net benefit in most cases. The optimal threshold probability range was 0.30-0.72 in the training set (**Figure 6A**), and that in the validation was 0.30-0.73 (**Figure 6B**), indicating that patients

Table 3. Predictive performance of each factor for CLP

Test result variable	AUC (95% CI)	SE	*P*	Sensitivity (%)	Specificity (%)	Optimum cutoff value
Age	0.640 (0.563-0.716)	0.039	0.001	57.0	73.3	-
CLP family history	0.596 (0.518-0.674)	0.040	0.017	24.0	95.2	-
Prenatal malnutrition	0.603 (0.525-0.680)	0.040	0.011	31.0	89.5	-
MSX1	0.755 (0.688-0.822)	0.034	< 0.001	69.0	77.1	1.53
IRF6	0.833 (0.777-0.889)	0.028	< 0.001	78.0	78.1	1.6

Note: CLP, cleft lip and palate; MSX1, msh homeobox 1; IRF6, interferon regulatory factor 6; AUC, area under the curve; SE, standard error.

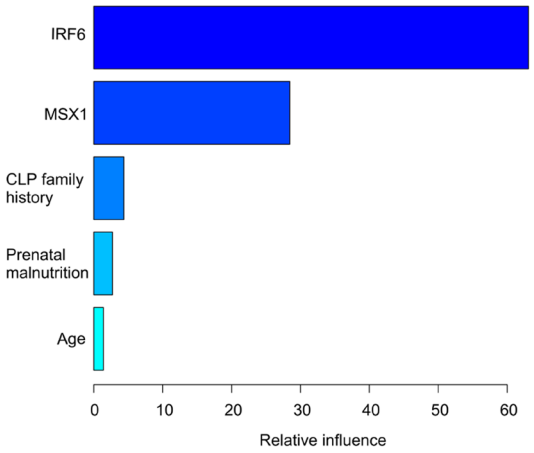

Figure 3. Importance of each factor in the predictive model. Note: CLP, cleft lip and palate; MSX1, msh homeobox 1; IRF6, interferon regulatory factor 6.

can obtain the greatest net benefit within this range.

Discussion

As a common congenital birth defect, CLP presents both physical and psychological challenges for affected children while imposing a heavy burden on families and healthcare systems. Early and accurate prediction of fetal CLP is crucial for clinical intervention, informed family decision-making, and efficient allocation of medical resources. This study investigated the association between maternal serum expression levels of MSX1 and IRF6 and fetal CLP, and constructed a GBM prediction model for providing a novel approach in clinical prediction.

This study found that the relative expression levels of MSX1 and IRF6 in maternal serum were significantly higher in the CLP group than those in the normal group, suggesting their potential role in fetal craniofacial development. MSX1 plays an important role in craniofacial development, particularly in facial bone and palatal development. Multiple studies have linked abnormal MSX1 expression to various craniofacial malformations. For example, Yang et al. [12] reported that MSX1 mutations could lead to abnormal facial bone development, increasing the risk of CLP. Its mechanism may involve regulating processes such as cell proliferation, differentiation, and apoptosis, influencing facial tissue formation. IRF6 also plays an indispensable role in craniofacial development. A study by Parmar et al. [13] demonstrated that IRF6 gene polymorphisms are closely related to CLP susceptibility. IRF6 is involved in epithelial cell differentiation and adhesion, and its dysfunction can compromise facial epithelial integrity, contributing to CLP formation. The elevated expression levels of MSX1 and IRF6 in maternal serum may reflect abnormalities in fetal development. Under normal conditions, these genes maintain expression levels within a specific range to support proper facial structure formation. However, in cases of fetal malformations such as CLP, their expression profiles may shift, potentially serving as biomarkers for abnormal craniofacial development. In summary, the elevated levels of MSX1 and IRF6 in maternal serum are likely associated with abnormal fetal development in CLP cases.

This study identified maternal age as an independent factor for fetal CLP. Advanced maternal age is associated with declining oocyte quality and an increased risk of chromosomal abnormalities, which may contribute to abnormal fetal development. Debbarh et al. [14] reported a significantly higher risk of CLP in older pregnant women compared to younger mothers. This may be related to germ cell aging,

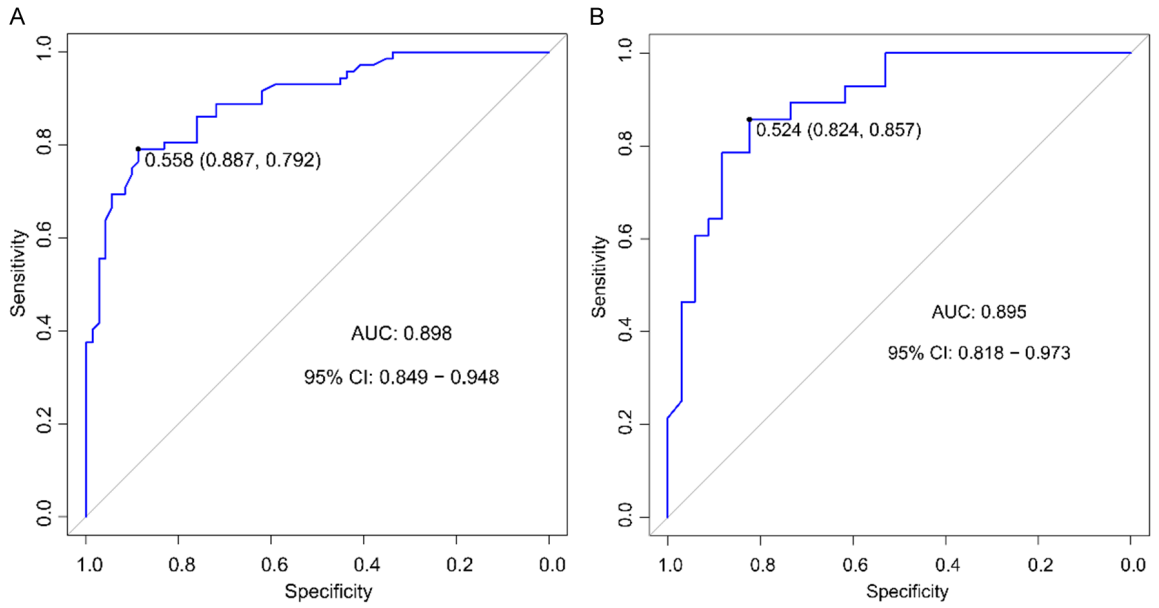

Figure 4. ROC curves for GBM model in predicting CLP in both the training (A) and validation (B) sets. Note: ROC, receiver operator characteristics; GBM, gradient boosting machine; CLP, cleft lip and palate.

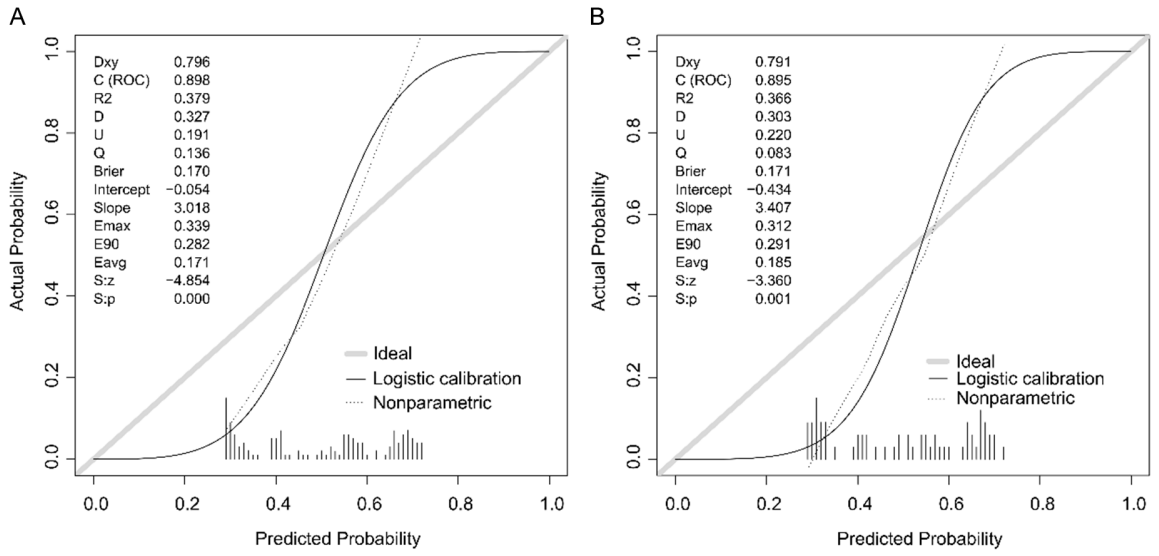

Figure 5. Predictive accuracy of GBM model analyzed by calibration curve diagram in both the training (A) and validation (B) sets. Note: GBM, gradient boosting machine.

accumulation of gene mutations, and changes in endocrine function. Pregnant women with a family history of CLP also face a significantly higher risk of having a fetus with CLP, highlighting the role of genetic predisposition. Familial CLP aggregation may be driven by mutations in key genes, which can disrupt facial structure development and increase CLP risk [15, 16]. Maternal malnutrition during pregnancy is another critical influencing factor. The fetal growth and development process is highly dependent on maternal nutritional status. Deficiencies in essential nutrients, such as folic acid and vitamin B12, have been linked to an increased risk of congenital anomalies, including CLP. Jain et al. [17] and Barrientos et al. [18] emphasized the importance of adequate prenatal nutrition in reducing birth de-

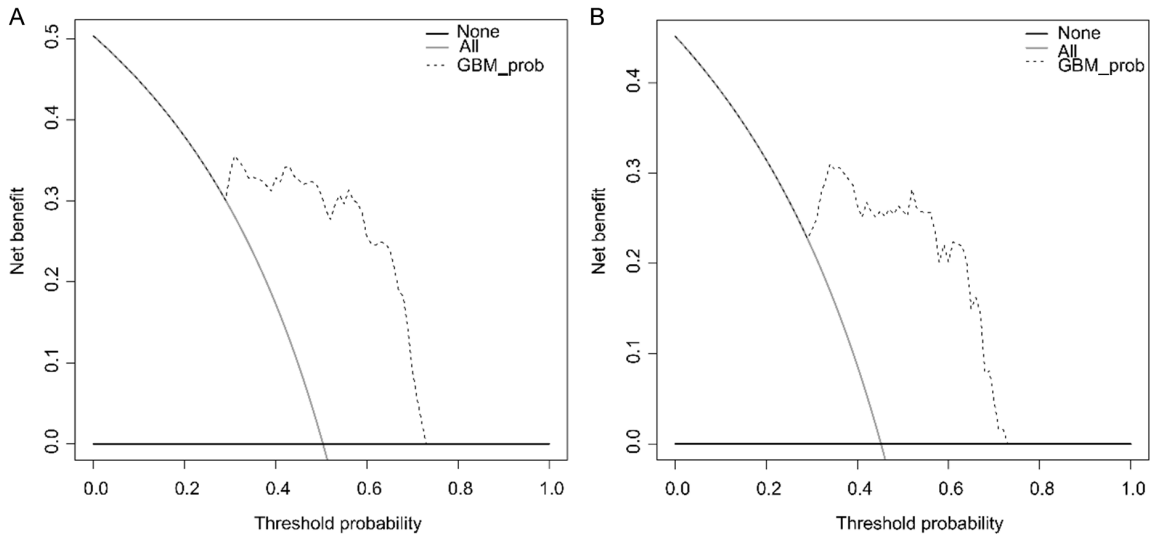

Figure 6. Net benefit of GBM model analyzed by decision curve diagrams in both the training (A) and validation (B) sets. Note: GBM, gradient boosting machine.

fect risks, suggesting that a balanced diet and proper supplementation can help prevent CLP. As mentioned earlier, abnormal MSX1 and IRF6 expression is closely related to CLP. These genes may contribute to fetal CLP susceptibility by disrupting facial tissue development, cellular signaling pathways, and gene regulatory networks. Overall, multiple factors including maternal age, family history, malnutrition, and gene expression contribute to the occurrence of fetal CLP.

The GBM prediction model developed in this study showed high predictive efficacy in both the training and the validation sets. The model's AUC, sensitivity, and specificity indicate strong accuracy and reliability, effectively predicting fetal CLP occurrence. The AUC approached to 0.9, indicating high discriminatory ability, allowing the model to accurately distinguish between CLP and normal fetuses. The high sensitivity and specificity indicate that the model effectively detects positive cases while minimizing false positives. The calibration curve demonstrated that the model's predicted probabilities closely aligned with actual probabilities, further verifying its accuracy. DCA established the threshold probability ranges for the training and validation sets, providing a reference for clinical decision-making. In practical applications, doctors can make reasonable clinical decisions by comprehensively considering various factors based on the model's prediction results and threshold probability range. The GBM model offers significant advantages in handling high-dimensional data and complex nonlinear relationships [19, 20]. In this study, the model integrated multiple influencing factors, including maternal age, family history of CLP, malnutrition during pregnancy, and MSX1 and IRF6 expression levels. Given the potential interactions among these factors, the GBM model effectively captured complex relationships, enhancing predictive accuracy. In addition, the GBM model also demonstrated strong stability and generalizability. Its performance in both the training and validation sets confirmed its reliability and applicability, supporting its potential for clinical implementation. To conclude, the GBM model shows great potential in effectively predicting the occurrence of fetal CLP in clinical settings.

This study also has limitations. As a single-center study, it may be subject to geographical limitations. Differences in genetic background, living environment, and dietary habits across populations in different regions may affect the generalizability of the findings. In addition, this study focused solely on maternal serum expression levels of MSX1 and IRF6, without investigating other potential biomarkers. However, fetal CLP is likely influenced by multiple genes, biomarkers, and environmental factors. Other maternal lifestyle factors, such as exposure to harmful substances, and environmental

influences, including pollution, were not analyzed, despite their potential role in CLP risk. Therefore, in the future, multi-center studies with a larger and more diverse sample size should be conducted to improve generalizability. Besides, future study can focus on exploring additional biomarkers associated with fetal CLP to develop a more comprehensive and precise prediction model. Additionally, incorporating broader maternal and environmental factors into risk evaluation models represents a promising strategy to refine risk assessment and increase predictive accuracy.

Conclusion

Elevated maternal serum MSX1 and IRF6 levels are closely related to fetal CLP, supporting their potential as biomarkers for clinical prediction. However, multi-center studies, biomarker exploration, and mechanistic research are needed to enhance prediction accuracy, improve prevention strategies, and facilitate early diagnosis and intervention. These advancements could ultimately reduce the burden on affected families and healthcare systems.

Acknowledgements

This work was supported by Ningbo Top Medical and Health Research Program (2022-020405) and Ningbo science and technology project (2023Z178) and Obstetrics Key Discipline (2022-B16).

Disclosure of conflict of interest

None.

Address correspondence to: An'er Chen, Department of Obstetrics, The Affiliated Women and Children's Hospital of Ningbo University, Ningbo 315000, Zhejiang, China. Tel: +86-0574-870834-41; E-mail: chenaner723@163.com

References

[1] Chimedtseren I, Niimi T, Inoue M, Furukawa H, Imura H, Minami K, Garidkhuu A, Gantugs AE and Natsume N. Prevention of cleft lip and/or palate in A/J mice by licorice solution. Congenit Anom (Kyoto) 2023; 63: 141-146.

[2] Lauer G, Pradel W and Birdir C. Cleft lip and palate : one of the most frequent congenital malformations. HNO 2023; 71: 276-284.

[3] Huang C, Song X, Zhang Q, Liu M, Mao J, Xiang J, Liu Y, Li H and Wang T. Prenatal diagnosis of a fetus with cleft lip and palate by using chromosomal microarray analysis. Zhonghua Yi Xue Yi Chuan Xue Za Zhi 2020; 37: 471-474.

[4] Lili Y, Jian M, Junpeng G, Kun Z, Jinfang Z and Yongqing H. Association between non-syndromic cleft lip with or without cleft palate and environmental factors in Ningxia. Hua Xi Kou Qiang Yi Xue Za Zhi 2017; 35: 291-295.

[5] Wan Y, Zhou Y, Xu X, Lu X, Wang Y and Zhang C. An effective ultrasound fetal palate screening software based on the "sequential sector scan through the oral fissure" and three-dimensional ultrasound. BMC Pregnancy Childbirth 2024; 24: 526.

[6] Zawiślak A, Woźniak K, Kawala B, Gupta S, Znamirowska-Bajowska A, Janiszewska-Olszowska J, Lubiński J, Calvo-Guirado JL, Grocholewicz K and Jakubowska A. IRF6 and FGF1 polymorphisms in non-syndromic cleft lip with or without cleft palate in the Polish population. Open Med (Wars) 2023; 18: 20230677.

[7] Memon A, Khidri FF, Waryah YM, Nigar R, Bhinder MA, Shaikh AM, Shaikh H and Waryah AM. Association of MSX1 gene variants with non-syndromic cleft lip and/or palate in the Pakistani population. Cleft Palate Craniofac J 2024; 61: 1845-1852.

[8] Nasroen SL, Tammama T, Darwis RS, Adil A, Rahmutia S, Maskoen AM and Gani BA. The IRF6 rs2013162 and MTHFR A1298C rs-1801131 gene polymorphisms related to non-syndromic cleft lip and palate among deutero-malay in indonesia. Cleft Palate Craniofac J 2024; 61: 2009-2016.

[9] Fabricant SP, Greiner KS and Caughey AB. Trauma in pregnancy and severe adverse perinatal outcomes. J Matern Fetal Neonatal Med 2021; 34: 3070-3074.

[10] Pesch MH, Saunders NA and Abdelnabi S. Cytomegalovirus infection in pregnancy: prevention, presentation, management and neonatal outcomes. J Midwifery Womens Health 2021; 66: 397-402.

[11] Okeke IG, Camarda AR, Okeke R and Chaughtai S. Semaglutide-induced hyperemesis gravidarum. JCEM Case Rep 2024; 2: luad167.

[12] Yang K, Dong XY, Wu J, Zhu JJ, Tan Y, Yan YS, Lin L and Zhang DL. A clinical and multi-omics study of Van der Woude syndrome in three generations of a Chinese family. Mol Med Rep 2020; 22: 2925-2931.

[13] Parmar B, Verma U, Khaire K, Danes D and Balakrishnan S. Inhibition of cyclooxygenase-2 alters craniofacial patterning during early embryonic development of chick. J Dev Biol 2021; 9: 16.

[14] Debbarh H, Jamil M, Jelloul H, Zakaria A, Louanjli N and Cadi R. Assessment of cell-free DNA and apoptosis in an oocyte microenvironment: promising biomarkers to predict intracytoplasmic sperm injection outcomes. Zygote 2023; 31: 296-302.
[15] Loenarz C, Ge W, Coleman ML, Rose NR, Cooper CD, Klose RJ, Ratcliffe PJ and Schofield CJ. PHF8, a gene associated with cleft lip/palate and mental retardation, encodes for an Nepsilon-dimethyl lysine demethylase. Hum Mol Genet 2010; 19: 217-22.
[16] Cui X, Zhu G, Han M, Li X, Lou S, Xing C, Xu S, Pan Y and Wang L. Genetic variants in BCL-2 family genes influence the risk of non-syndromic cleft lip with or without cleft palate. Birth Defects Res 2024; 116: e2288.
[17] Jain S, Maheshwari A and Jain SK. Maternal nutrition and fetal/infant development. Clin Perinatol 2022; 49: 313-330.
[18] Barrientos G, Ronchi F and Conrad ML. Nutrition during pregnancy: influence on the gut microbiome and fetal development. Am J Reprod Immunol 2024; 91: e13802.
[19] Tong C, Du X, Chen Y, Zhang K, Shan M, Shen Z, Zhang H and Zheng J. Machine learning prediction model of major adverse outcomes after pediatric congenital heart surgery: a retrospective cohort study. Int J Surg 2024; 110: 2207-2216.
[20] Ren Z, Zhu J, Gao Y, Yin Q, Hu M, Dai L, Deng C, Yi L, Deng K, Wang Y, Li X and Wang J. Maternal exposure to ambient PM10 during pregnancy increases the risk of congenital heart defects: evidence from machine learning models. Sci Total Environ 2018; 630: 1-10.