ISSN: 2664-9640

Study of Mother-infant Behavioural Relationships based on Structural Equation Modelling and LightGBM Regression Models

Zhiyuan Lu*

College of Finance and Statistics, Hunan University, Changsha, China

*Corresponding author: 2331178006@qq.com

Abstract

In this paper, the relationship between mothers' physical and mental health and infants' behavior was investigated. Structural equation modeling (SEM) was used to analyze the standardized data, analyze the relationship between the influence of the paths of each factor, understand the composition of the paths of each factor node from a global perspective, and filter the quantitative variables within the factor through the factor loading coefficients, record the standardized coefficients and test the significance of the data to eliminate invalid paths, which in turn led to the correlation law. At the same time, the relationship model of the infant's behavioral characteristics with the mother's physical and psychological indicators was established using the LightGBM regression model for classification, so as to establish the relationship model, and get the model with good performance.

Keywords

SEM; Factor Node; LightGBM.

1. Introduction

Changes in the physical and psychological status of mothers can directly or indirectly affect all aspects of their infants, of which the evaluation of the physical status of mothers can be reflected by five aspects: age of the mother, marital status, education, duration of pregnancy, and mode of delivery, and the evaluation of the psychological status of mothers can be reflected by three indicators: CBTS, EPDS, and HADS. Since the physical and psychological status of mothers changes over time, and in the process of giving birth to and raising children, the physical and psychological status of mothers may be affected by different aspects of the living environment. Therefore, it is of great practical significance to consider how to maximize the protection of infants' healthy growth in today's society.

Some scholars have already studied the relationship between mothers' and infants' behavior. Mary [1] et al. explored the relationship between maternal alcohol consumption, mother-infant interaction, and infants' cognitive development. Structural equation modeling was used to test a model in which infants' cognitive development (as measured by the Bayley Psychological Inventory) was hypothesized to be influenced by the infant's emotions and the mother's ability to interact with the infant. Crockenberg [2] et al. explored the relationship between maternal and infant characteristics and the development of infant temperament and mother-infant interactions at 3 months after birth. Data were analyzed using hierarchical multiple regression techniques. The results showed that neonatal irritability showed some consistency as measured by calm time; observed fussiness and crying were not associated with neonatal irritability but with mothers' unresponsive attitudes and behaviors; and initial irritability was associated with greater responsiveness to female infants than with male infants. Tokuda's [3] study investigated the factors that influence the development of psychological distress in women who did not experience psychological distress in their mothers during pregnancy, and

ISSN: 2664-9640

who had no psychological distress in the 12 post-delivery months after delivery to determine the factors influencing the occurrence of psychological distress. A 3-month pre-pregnancy questionnaire was used to investigate feelings of pregnancy, and the 6-item Kessler Psychological Distress Scale (K6) was used to assess maternal mental health during pre-pregnancy, mid- and late pregnancy, and 12 months postpartum.

Additionally, predictions of behavior have been made. Burg [4] et al. examined the relationship between scores of minor physical abnormalities (MPAs) and behavior in 1- and 2-year-old neonates. These results suggest that neonatal abnormality scores by themselves are unlikely to prove clinically useful in predicting preschool behavioral problems in an unselected population. Aliabadi [5] et al. selected 50 low birth weight infants under the age of 2 months who met the inclusion criteria by using a convenience sampling method in the Neonatal Intensive Care Unit (NICU) and the pediatric outpatient clinic of Shahid Akbarabadi Hospital in Tehran. Motor and behavioral abilities were assessed using the Test of Infant Motor Performance (TIMP) and Neonatal Behavioral Assessment Scale (NBAS), respectively. Chen [6] et al. expressed the relationship between the human body and joint angles by constructing a multivariate regression model, which predicted the optimal postures by the joint angles. Shilo [7] et al. investigated the relationship between the human body and joint angles by constructing a multivariate regression model, which predicted the optimal postures by the joint angles. 5- and 8-year-olds whether they relied on group preferences to predict the behavior of others, and whether their reliance on group preferences differed between in-group and out-group.

However, no scholars have constructed a systematic mathematical model for the mother-infant relationship, and at the same time, no scholars have predicted infants' behavior based on mothers' characteristics. Therefore, this paper first explores the relevant laws of the mother-infant relationship based on the SEM model, and then predicts the behavior of infants based on LightGBM. In this paper, with https://publicqn.saikr.com/2023/08/03/contest/698fed3121efd2f1dc22689b2aca39221691056478428.rar?attname=2023%E5%B9%B4%E7%AC%AC%E5%9B%9B%E5%B1%8A%E2%80%9C%E5%8D%8E%E6%95%B0%E6%9D%AF%E2%80%9D%E5%85%A8%E5%9B%BD%E5%A4%A7%E5%AD%A6%E7%94%9F%E6%95%B0%E5%AD%A6%E5%AD%A6%E5%BB%A6%E5%AD%A6 E5%BB%BA%E6%A8%A1%E7%AB%9E%E8%B5%9B%E8%B5%9B%E9%A2%98.rar as an example of relevant research.

2. Exploration of Correlation Laws based on SEM Modeling

First, in order to eliminate the scale you need to standardize the data first. Data standardization means scaling the data required for the schedule so that it falls into a small specific interval. Data standardization gives the mean and standard deviation of the original data to standardize the data. The processed data conforms to the standard normal distribution, i.e., the mean is 0 and the standard deviation is 1, which facilitates the ability to compare and weigh indicators of different units or magnitudes. Its transformation function is:

$$x *= \frac{x-\mu}{\sigma} \tag{1}$$

where μ is the mean of the sample data and σ is the standard deviation of the sample data. Meanwhile, data standardization can improve the convergence speed of the model, enhance the accuracy of the model, and eliminate the influence of magnitude and scale.

Structural equation modeling (SEM) is a method for building, estimating, and testing causal models that clearly analyze the effects of individual indicators on the aggregate and the interrelationships among individual indicators. The model allows the independent and dependent variables to contain errors, while accurately estimating the relationship between observed and latent variables, estimating the factor structure and factor relationships, and estimating the degree of fit of the overall model and data.

The steps of the analysis are as follows:

1) According to the weighted path diagram of the structural model, analyze the factor path influences relationship situation, and at the same time understand the path composition and relationship of each factor node from the global.

2) Screening the intra-factor quantitative variables through the factor loading coefficients, we assume that the measurement variables pass the significance test (P<0.05) and the standardized loading coefficient value is greater than 0.56 can indicate that the measurement variables meet the requirements of factor dimensionality reduction.

3) Find the model path coefficients of the paired features through the model path coefficient table, and analyze whether there is an influence relationship between the model variables according to the significance test (P<0.05), if there is significance, it indicates that there is an influence relationship between the variables, and carry out an in-depth analysis of the efficiency of the influence through the standardized path coefficients.

4) Through the model fit index can be analyzed model fit, while not requiring all test results; the covariance table analyzes the nodes in the model without incidence, while it can be used to analyze the correlation between the path nodes.

Drawing the model path diagram is shown in Fig。 1:

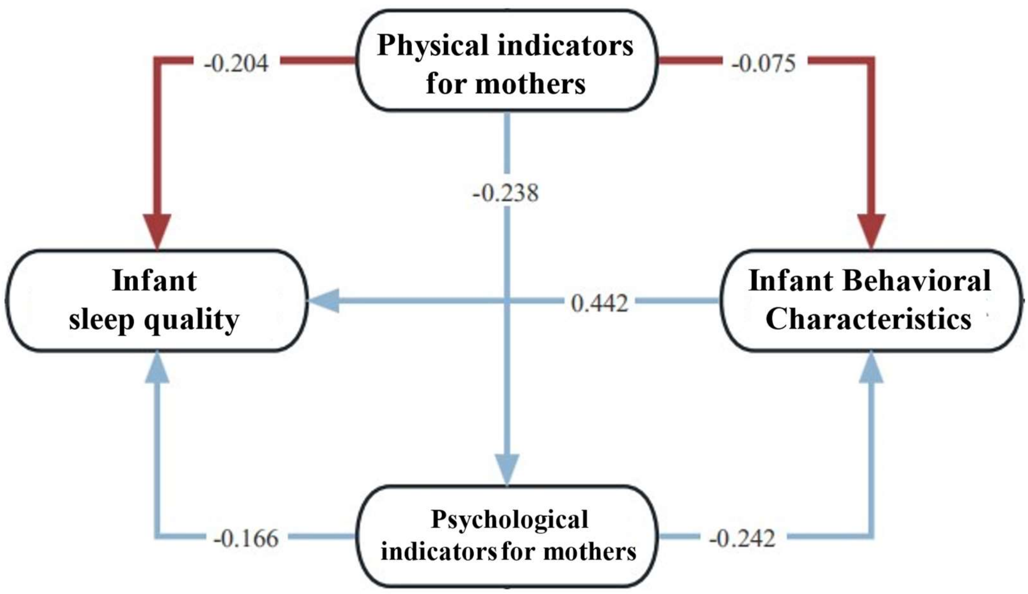

Fig. 1 Model Path Diagram

For the observed and latent variables loading factors are shown in Table 1.

The original hypothesis is rejected if it presents significance at the level based on educational attainment (significance p-value of 0.038**), while its standardized loading coefficients are all greater than 0.4, which can be considered as having enough variance explained to show that the variables can be presented on the same factor.

Based on the EPDS (significance p-value of 0.000***), and HADS (significance p-value of 0.000***) presenting significance at the level, the original hypothesis is rejected, while its standardized loading coefficients are all greater than 0.4, which can be considered to have enough variance explained to show that each variable can be presented on the same factor.

Table 1. Table of load factors for observed and latent variables

factor	variant	Non-standard load factors	Standardized load factor	z	S.E.	P
Physical indicators for mothers	Age of mother	1	0.325	-	-	-
	marital status	0.26	0.18	1.945	0.134	0.052*
	educational attainment	2.718	0.565	2.078	1.308	0.038**
	Duration of pregnancy (weeks)	0.099	0.044	0.551	0.18	0.581
	Mode of delivery	-0.026	-0.012	-0.15	0.173	0.881
Psychological indicators for mothers	CBTS	1	0.823	-	-	-
	EPDS	1.182	0.951	21.654	0.055	0.000***
	HADS	0.909	0.833	19.569	0.046	0.000***
Behavioral characteristics of young infants	Infant Behavioral Characteristics	1	0.607	-	-	-
Sleep quality of young infants	Sleeping time throughout the night	1	0.472	-	-	-
	Number of wakings	-1.097	-0.665	-4.698	0.233	0.000***
	How to fall asleep	1.456	0.404	4.691	0.31	0.000***
Note: ***, **, and * represent 1%, 5%, and 10% significance levels, respectively.						

Based on the quality of sleep of young infants (significance p-value of 0.000***) presented at the level of significance, the original hypothesis is rejected, while its standardized loading coefficients are all greater than 0.4, it can be considered that it has enough variance explained rate to show that each variable can be presented on the same factor.

Based on the mode of falling asleep (significance p-value of 0.000***) presents significance at the level of the original hypothesis is rejected, at the same time its standardized loading coefficients are greater than 0.4, it can be considered that it has enough variance explained rate to show that each variable can be presented on the same factor.

For the model regression coefficient table data are shown in Table 2:

Table 2. Model regression coefficient

Factor(latent variable)	→	Analytic terms (explicit variables)	Non-standardized coefficient	Standardized coefficient	standard error	Z	P
Physical indicators for mothers	→	Behavioral characteristics of young infants	-0.276	-0.075	0.523	-0.528	0.597
Psychological indicators for mothers	→	Behavioral characteristics of young infants	-0.233	-0.242	0.089	-2.601	0.009***
Physical indicators for mothers	→	Sleep quality of young infants	-0.389	-0.204	0.239	-1.631	0.103
Psychological indicators for mothers	→	Sleep quality of young infants	-0.083	-0.166	0.039	-2.131	0.033**
Behavioral characteristics of young infants	→	Sleep quality of young infants	0.229	0.442	0.065	3.523	0.000***
Physical indicators for mothers	→	Psychological indicators for mothers	-0.914	-0.238	0.405	-2.256	0.024**
Note: ***, **, and * represent 1%, 5%, and 10% significance levels, respectively.							

Based on the paired item Maternal Physical Indicators -> Young Infant Behavioral Characteristics, the significance p-value is 0.597, which does not present significance at the level, then the original hypothesis cannot be rejected, therefore this path is invalid.

Based on the paired item Mothers' Psychological Indicators -> Young Infants' Behavioral Characteristics, the significance p-value is 0.009***, which presents significance at the level, then the original hypothesis is rejected, so this path is valid with an impact coefficient of -0.242.

Based on the paired item Maternal Physical Indicators -> Sleep Quality of Young Infants, the significance p-value is 0.103, which does not present significance at the level, then the original hypothesis cannot be rejected, so this path is invalid.

Based on the paired item Maternal Psychological Indicators -> Sleep Quality of Young Infants, the significance p-value is 0.033**, which presents significance at the level, then the original hypothesis is rejected, so this path is valid with an impact coefficient of -0.166.

Based on the paired item Infant Behavioral Characteristics -> Infant Sleep Quality, the significance p-value is 0.000***, which presents significance at the level, then the original hypothesis is rejected, so this path is valid and its impact coefficient is 0.442.

Based on the paired term mother's physical indicators -> mother's psychological indicators, the significance p-value is 0.024**, which presents significance at the level, then the original hypothesis is rejected, so this path is valid and its impact coefficient is -0.238.

The data for the indicators fitted to the model are shown in Table 3.

Table 3. Model fit indicator data

χ^2	df	P	chi-square degree-of-freedom ratio	GFI	RMSEA	RMR	CFI	NFI	NNFI
-	-	>0.05	<3	>0.9	<0.10	<0.05	>0.9	>0.9	>0.9
110.626	48.000	0.000***	2.305	0.891	0.058	0.005	0.934	0.891	0.909
Note: ***, **, * represent 1%, 5%, and 10% significance levels, respectively.									

GFI (Goodness of Fit Index): The coefficient of determination and regression standard deviation are used to test the fit of the model to the sample observations. The value of GFI is between 0 and 1, and the closer it is to 0, the worse the fit is.CFI ≥ 0.9, the model is considered to be a good fit. The GFI value of this model is 0.891, which is close to 0.9 and can be considered as a good model fit.

RMSEA (Root Mean Square of Approximation Error): In general, RMSEA is below 0.08 (the smaller the better). The RMSEA value of this model is 0.058, which is below 0.08 and is eligible.

RMR (Root Mean Square Residual): this metric measures the fit of the model by measuring the average residual of the predicted correlation and the actual observed correlation. If the RMR is <0.1, the model is considered to be well-fitted. The RMR value of this model is 0.005, which is much less than 0.1, and therefore the model is well-fitted.

CFI (Comparative Fit Index): When comparing the hypothetical model with the independent model, the value is between 0 and 1. The closer to 0 means the worse the fit, and the closer to 1 means the better the fit. In general, a CFI ≥ 0.9 is considered a good model fit. The CFI value of this model is 0.934, which meets the condition, and therefore the model is well-fitted.

NNFI (non-normalized fit coefficient) and CFI (comparative fit index): the larger the value, the better the performance of the fitted model. The NNFI value of this model is 0.909, which is eligible and therefore the model fits better.

Based on the analysis of the significance of the coefficient magnitude above, the model fit performance is good.

For the above analysis of the table data relationship between observed variables and latent variables, we get the graph of the relationship between observed variables and latent variables as shown in Fig. 2.

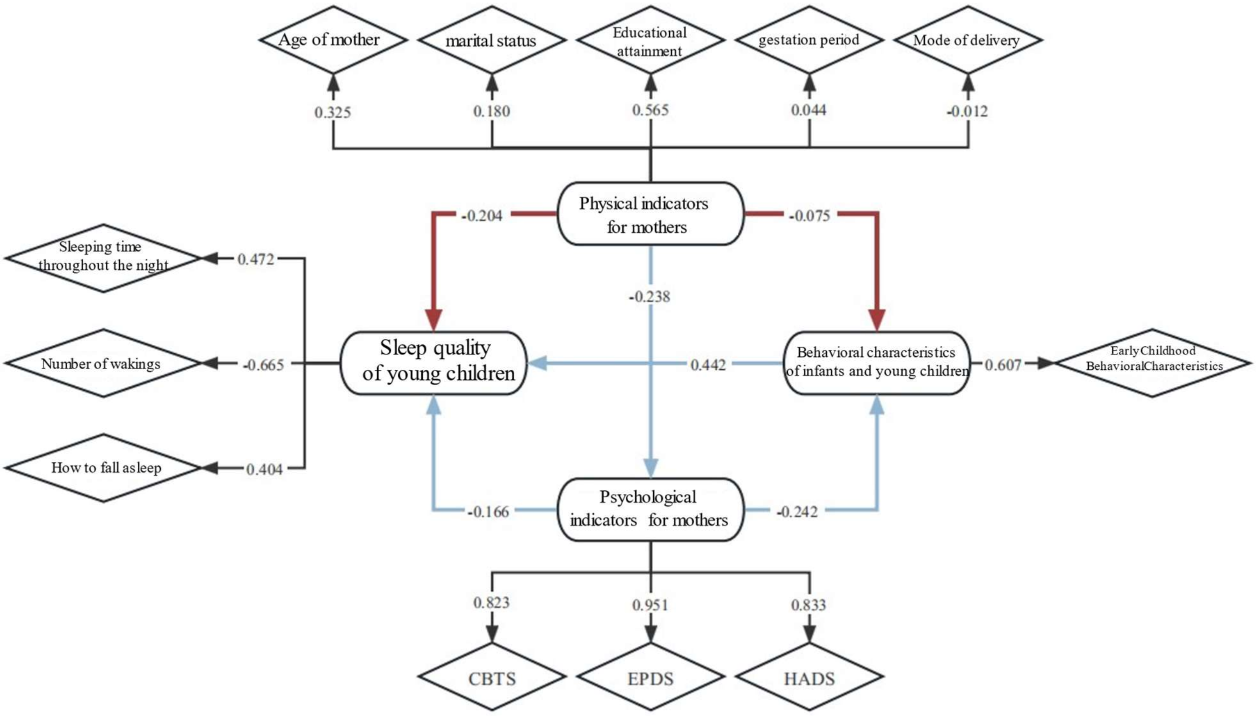

Fig. 2 Plot of observed variables versus latent variables

Mothers' physical indicators and psychological indicators, the mode of delivery and gestation time p-values are relatively high, and can not reject the original hypothesis so rounded, the other factor's p-values are smaller, most of them are less than 0.05 significant level, the credibility is higher, so the SEM structure is more stable can be used.

By assuming that there is a certain relationship between the variables, derived from the red arrow path in Figure 2 mother's physical indicators -> behavioral characteristics of young infants, mother's physical indicators -> sleep quality of young infants influence path is invalid;

The blue arrow path in Figure 2 is valid, in which the influence coefficient of infant behavioral characteristics -> infant sleep quality is positive, and infant behavioral characteristics are positively correlated with infant sleep quality; the influence coefficients of mother's psychological indicators -> infant sleep quality, mother's psychological indicators -> infant behavioral characteristics, and mother's physical indicators -> mother's psychological indicators are negative, and according to the requirements of the Appendix Table, it is known that the greater the value of the mother's psychological indicators, the more serious the symptoms of psychological disorders, and the more serious the symptoms of mental disorders. The higher the value of the psychological indicators, the more severe the symptoms of mental illness, and categorized as costly indicators;

Therefore, the negative impact coefficients of the above three are in line with common sense, and according to the absolute value of the standardized coefficients, they are divided into significant and large impacts; among them, the behavioral characteristics of young children have a significant impact on the quality of young children's behaviors, and the psychological indicators of mothers have a large impact on the behavioral characteristics of young infants and mothers, and the physical indicators of mothers -> psychological indicators of mothers. Therefore, there is a correlation between the status of mothers' physical and psychological indicators and the status of infants' behavioral characteristics and sleep quality.

ISSN: 2664-9640

3. Infant Behavior Prediction Model based on LightGBM

LightGBM (Light Gradient Boosting Machine) is a framework for implementing the GBDT algorithm, which supports efficient parallel training and has the advantages of faster training speed, lower memory consumption, better accuracy, and support for distributed can quickly deal with massive data.

The steps to build the LightGBM model are:

1) Build the LightGBM model by training set data.

2) Calculate the feature importance through the established LightGBM model.

3) Apply the established LightGBM model to the training and testing data to get the model evaluation results.

In this paper, the set model parameters are shown in Table 4.

Table 4. Model parameter

parameter name	parameter value
training time	0.173s
data slice	0.7
confuse data	NO
cross-validation	NO
base learning device (BLD)	gbdt
Number of base learners	100
learning rate	0.1
L1 regular term	0
L2 regular terms	1
sample rate	1
Tree feature sampling rate	1
Node splitting threshold	0
Minimum weight of samples in leaf nodes	0
Maximum depth of the tree	10
Minimum number of samples for leaf nodes	10

Running the LightGBM model, the ranking of feature importance can be obtained as shown in Fig. 3.

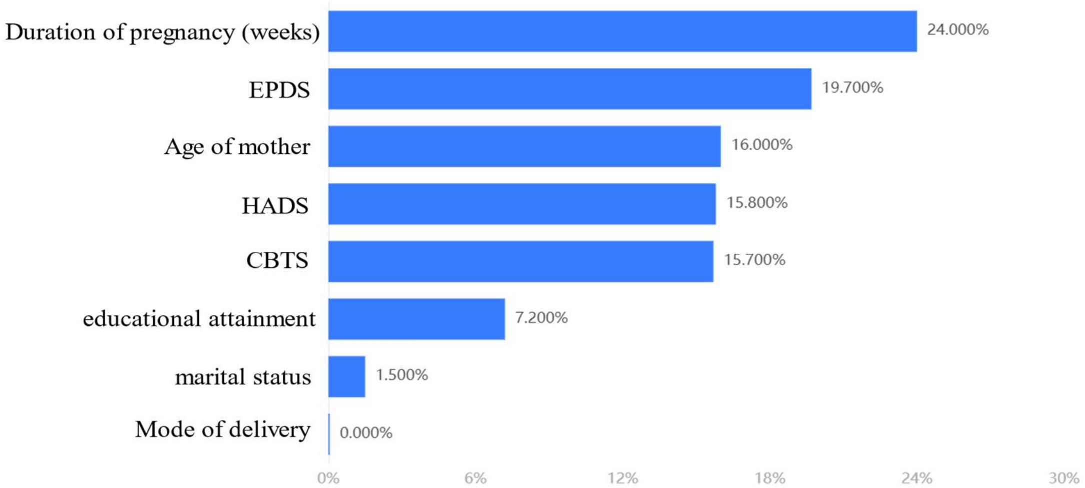

Fig. 3 Characteristic importance

ISSN: 2664-9640

As can be seen in Figure 3, gestation time (weeks) is the factor with the greatest influence of the mother's physical and psychological indicators on the infant's behavioral characteristics, with a share of 24.0%; followed by EPDS, with a share of 19.7%; the three factors of the mother's age, HADS, and CBTS are close to the same degree of influence, with a share of 16.0%; and the mother's education level is the next highest, with a share of 7.2%; Marital status and mode of delivery were the two factors in which the physical and psychological indicators of the mother had the least influence on the behavioral characteristics of the infant, with a percentage of 1.5% and 0.0%, respectively.

In order to quantify the prediction effect, MSE, RMSE, MAE, and MAPE were introduced to evaluate the prediction accuracy. The calculation formula is shown below.

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |\hat{y}_i - y_i| \tag{2}$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \tag{3}$$

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (y_i - \hat{y}_i)^2} \tag{4}$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N} \left|\frac{\hat{y}_i - y_i}{y_i}\right| \tag{5}$$

where N is the number of samples, y_i is the true value, and $\hat{y}_i$ is the predicted value.

Table V demonstrates the prediction evaluation metrics for the training and test sets, which measure the prediction effectiveness of LightGBM through quantitative metrics. The MSE, RMSE, and MAE of the model's test set and training set are all extremely small, and the MAPE is less than 24.1%, which indicates that the model performs well.

Table 5. Model evaluation results

	MSE	RMSE	MAE	MAPE
training set	0.06	0.244	0.198	9.842
test set	0.42	0.648	0.522	24.067

4. Summary

In this paper, a study was conducted on the relationship between mothers' physical and mental health and infant behavior. Using structural equation modeling (SEM) to analyze the standardized data, analyze the relationship between the influence of the path of each factor, from the global understanding of the composition of the path of each factor node, and through the factor loading coefficients of the factor within the quantitative variables of the filtering argument, record the standardized coefficients and the significance of the data test to eliminate the invalid paths, and then got the correlation law. At the same time, the relationship model between the behavioral characteristics of infants and the physical and psychological indicators of mothers was established using the LightGBM regression model for classification, so as to establish the relationship model and obtain a model with good performance.

It is worth mentioning that the research method designed in this paper can be applied not only to the study of the relationship between the mother's physical and mental health and the infant's behavior, but also to the rest of the complex correlation problems.

References

[1] Mary et al., "Interactional model for the association among maternal alcohol use, mother-infant interaction, and infant cognitive development," Infant Behavior & Development, 1993.

[2] S. B. Crockenberg and P. Smith, "Antecedents of mother–infant interaction and infant irritability in the first 3 months of life," Infant Behavior and Development, vol. 25, no. 1, pp. 2–15, 2002.

[3] N. Tokuda, H. Tanaka, H. Sawai, H. Shibahara, Y. Takeshima, and M. Shima, "Analyzing the relationship between feelings about pregnancy and mother-infant bonding with the onset of maternal psychological distress after childbirth: The Japan Environment and Children's Study," Journal of affective disorders, 2023.

[4] C. Burg, D. Hart, and Q. J. Rapoport, "Newborn minor physical anomalies and prediction of infant behavior," Journal of Autism&Developmental Disorders, 1978, doi: 10.1007/BF01538048.

[5] F. Aliabadi, M. Amini, M. Alizade, and M. Kalani, "Prediction of Infant Motor Performance Through Performance Evaluation of Behavior," Modern Rehabilitation, 2012, Accessed: Sep. 05, 2023.

[6] B. Chen, J. Liu, S. Wang, W. Yao, and L. Zhang, "Research on toilet pedal design using human factors engineering," in International Conference on Advanced Manufacturing Technology and Manufacturing Systems (ICAMTMS 2022), SPIE, Aug. 2022, pp. 237–244.

[7] R. Shilo, A. Weinsdrfer, H. Rakoczy, and G. Diesendruck, "Children's prediction of others' behavior based on group vs. individual properties," Cognitive Development, vol. 57, p. 100955, 2021.