RESEARCH ARTICLE

Prediction model of Chinese medicine intervention effect for children with precocious puberty based on machine learning

Yanli Wang[1], Hao Liu[2], Wen Tang[3], Lili Shang[3, *]

[1]The first Clinical Medical College, The First Affiliated Hospital of Anhui University of Chinese Medicine, Hefei, Anhui, China. [2]Traditional Chinese Medicine Department, The Second People's Hospital of Anhui Province, Hefei, Anhui, China. [3]Department of Pediatrics, The First Affiliated Hospital of Anhui University of Chinese Medicine, Hefei, Anhui, China.

Received: November 19, 2024; **accepted:** March 25, 2025.

Precocious puberty in children is a growing health issue with its incidence rising globally, especially in urban areas due to lifestyle changes and pollution. This study used machine learning and traditional Chinese medicine (TCM) experience to build a prediction model for TCM's treatment efficacy on childhood precocious puberty. Data from multiple sources including medical records and international databases covered patient details like age (4 - 14 years, median 8), gender (60% female), Body Mass Index (BMI) (median 18.5), family history (28% with a history), and lifestyle habits. Rigorous data preprocessing and feature engineering were carried out. A deep - learning model based on residual network (ResNet) with an attention mechanism was constructed. Evaluated on training, validation, and test sets, the proposed model outperformed traditional models. The accuracy on the test set reached 90% and the area under the receiver operating characteristic curve (AUC - ROC) was 0.93 with 92% accuracy and 0.95 AUC - ROC on the training set, and 91% accuracy and 0.94 AUC - ROC on the validation set. The proposed model showed strong performance and generalization ability. This research provided new insights for TCM treatment of childhood precocious puberty and promoted machine learning application in TCM research.

Keywords: precocious puberty; children; traditional Chinese medicine; machine learning; prediction model; performance evaluation; data analysis.

[*]**Corresponding author:** Lili Shang, Department of Pediatrics, The First Affiliated Hospital of Anhui University of Chinese Medicine, Hefei 230031, Anhui, China. Email: shanglili188@outlook.com.

Introduction

Precocious puberty refers to a condition in which children demonstrate sexual development before puberty. With the changes in modern lifestyles and the influence of factors such as environmental pollution, the incidence of precocious puberty in children is on the rise worldwide. In the field of pediatric endocrinology, precocious puberty in children has attracted increasing attention [1, 2]. Precocious puberty is defined as the onset of sexual development before the age of 8 in girls and before the age of 9 in boys [3]. This condition not only affects the physical growth of children but also brings potential psychological and social adaptation problems [4, 5]. Over the past few decades, with the transformation of modern lifestyles and the influence of environmental factors such as pollution, the incidence of precocious puberty in children has been on the rise globally [6, 7]. According to the statistics from the World Health Organization, in some regions, the incidence of precocious puberty in children has reached 1%, especially in urban areas where the proportion is even higher due to factors like unreasonable dietary structures and increased exposure to chemical substances [8].

Traditional western medicine mainly uses hormone therapy to treat precocious puberty, which can control disease progression to a certain extent. However, it also has potential side effects [9]. In contrast, traditional Chinese medicine (TCM) has a long-standing history and rich experience in treating various diseases including precocious puberty with its unique theoretical system of syndrome differentiation and treatment [10, 11]. TCM believes that the occurrence of precocious puberty is related to internal factors such as insufficient kidney qi and liver depression transforming into fire, and treatment focuses on harmonizing yin and yang and soothing the liver and regulating qi [12]. Despite the application of TCM in treating precocious puberty, there is a lack of systematic and accurate methods to predict the efficacy of TCM intervention, which limits the optimization of treatment plans and the improvement of treatment effects [13].

To address this issue, this study used advanced machine learning technology, combined with a large amount of clinical data from multiple medical institutions, to build a prediction model for the efficacy of TCM intervention in children with precocious puberty [14]. The data covered multiple characteristics of patients including age, gender, body mass index, family genetic history, lifestyle, medical history, and laboratory test results. Through data preprocessing, feature engineering, and model selection, a prediction model based on a deep - learning framework was constructed. This research was of great significance to the scientific community, which not only provided a new perspective for the application of TCM in the treatment of precocious puberty in children but also promoted the integration of machine learning technology and TCM research, which might lead to more accurate and personalized treatment strategies in the future [15].

Materials and methods

Data collection

This study used multiple databases to ensure comprehensiveness and reliability of the data, which included National Health and Nutrition Examination Survey (NHANES) (https://www.cdc.gov/nchs/nhanes/index.htm). This database collected a large amount of health and nutrition-related data of the American people, providing rich basic information in this study. The data retrieved were collected from 2005 to 2020 and included more than 50,000 questionnaire survey data, more than 30,000 physical measurement data, and more than 25,000 sets of laboratory test data. In addition, relevant data from medical professional database platforms were also integrated such as PubMed (https://www.ncbi.nlm.nih.gov/pmc/), which brought about 1,500 articles on precocious puberty and related research to provide theoretical and references for this study. These multi-source data laid a solid foundation for building a comprehensive and accurate prediction model [16, 17]. In addition, the electronic medical records (EMR) of cooperative hospitals and clinical trial data on the treatment of precocious puberty with traditional Chinese medicine were obtained, which were important data sources for this study [18].

Research subjects

A total of 500 patient’s records were included in this study with female to male ratio of 60% to 40%, the age range of 4 -14 years old, of which 4 - 8 years old patients accounted for 35%, 8 - 12 years old patients accounted for 45%, and 12 - 14 years old patients accounted for 20% [19]. Only 28% of the patients had a family genetic history of precocious puberty. According to the theory of traditional Chinese medicine, the patients were classified into syndrome groups with "liver depression and fire" as the most common syndrome accounting for 40% followed by "kidney deficiency" accounting for 20% and the remaining "other" syndromes accounting for 40%. All patients’ data were collected anonymously to ensure patient privacy and from multiple locations with 180 records from Beijing Children's Hospital (Beijing, China), 150 records from Shanghai Children's Medical Center (Shanghai, China), 100 records from the First Affiliated Hospital of Guangzhou University of Chinese Medicine (Guangzhou, Guangdong, China), 70 records from hospitals and clinics located in Chengdu, (Sichuan, China), Wuhan (Hubei, China), Xi'an (Shaanxi, China), and other places. All research procedures were strictly following the ethical standards and were approved by the Ethics Committees in related hospitals and clinics.

Data preprocessing

To ensure that the data set used for modeling was accurate and reliable, a comprehensive quality check and sorting of the collected raw data were conducted during the data preprocessing stage. Box plots and Z-score methods were used to detect outliers for each data in the dataset. For numerical data such as age, height, weight, laboratory test indicators, *etc*., a reasonable threshold range was set. If the data point exceeded this range, it would be judged as an outlier. For these outliers, other relevant information about the patient was first combined to make reasonable inferences and corrections. If the data still could not be corrected, it would be marked as a missing value. Duplicate record was processed by using hash table data structure to avoid duplicating records of the same patient. By comparing key fields such as patient ID, date of visit, medical record number, *etc*., duplicate records were identified and deleted to ensure the uniqueness of each data and improve data quality and analysis efficiency. Data from different data sources were standardized to ensure the consistency of expression of the same attributes. All numerical data were normalized to eliminate the impact of dimensions and make different features in the same order of magnitude, which was convenient for subsequent model learning and analysis. For categorical data, unique hot encoding, label encoding, and other methods were used to convert them, so that they could be effectively processed by the model. For data such as laboratory test results that were expressed in different units, they were converted to the international system of units (SI) to ensure data consistency and comparability. For a small amount of missing numerical data, interpolation was used to fill in the missing values. Missing values were predicted by calculating the mean or median of the nearest neighbor samples or using linear regression and other methods. For a large amount of missing data, if the missing ratio of a feature exceeded 30%, the feature or the entire record containing the missing value would be deleted to avoid the bias of incomplete data on model training. In the process of processing missing values, the distribution and processing methods would be recorded for subsequent analysis and evaluation.

Feature engineering

This study carefully selected a series of feature variables closely related to the treatment effect of traditional Chinese medicine and improved the prediction ability of the model through reasonable feature conversion and combination. Feature variables were selected from multiple dimensions, covering demographic characteristics, physiological indicators, living habits, family genetic information, laboratory test results, TCM diagnostic information, and TCM syndrome differentiation. To enable the model to better capture the intrinsic relationship between features, some of the original features were transformed and combined, which included standardization/normalization to standardize or normalize numerical features to eliminate the impact of dimensions and make different features in the same order of magnitude; feature derivation by creating new features based on domain knowledge, calculating the BMI index, evaluating nutritional status, and calculating the ratio of daily calorie intake to consumption to reflect the energy balance state; feature interaction by considering that some features might have nonlinear relationships, introducing product terms or ratio terms between features to capture complex interactions, creating a product term of age and BMI to explore the impact of BMI on the development of precocious puberty in

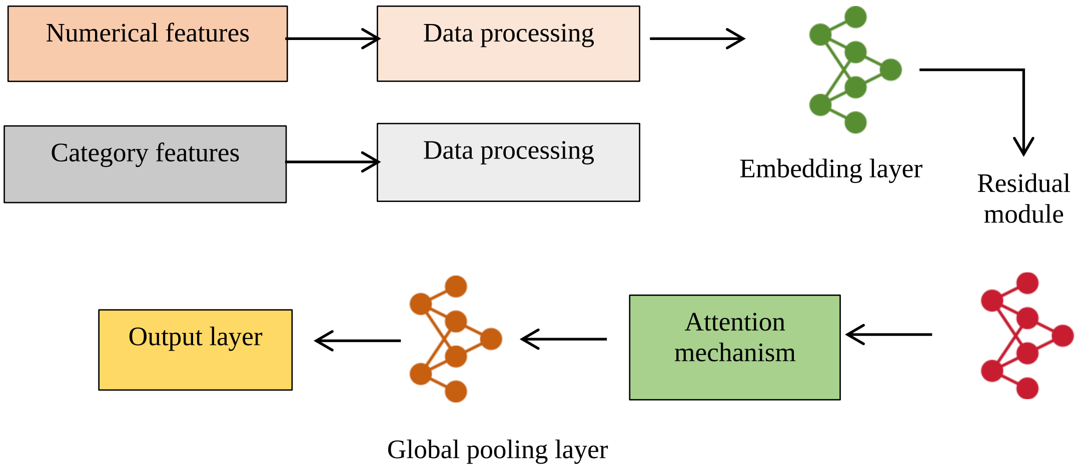

Figure 1. Proposed model framework.

children of different ages, and introducing a ratio term of hormone levels and family genetic history to evaluate the impact of genetic factors under different hormone levels. These screened and processed feature variables provided abundant information for the model and helped to improve the prediction accuracy and generalization ability of the model.

Model selection and training

The collected data were divided into training set, validation set, and test set on a ratio of 70%, 15%, and 15%, respectively. Residual network (ResNet) (https://arxiv.org/abs/1512.03385), TensorFlow (https://www.tensorflow.org/), and PyTorch (https://pytorch.org/) were employed in this research, while the Softmax function was used for multi-classification problems. Keras (https://keras.io/api/layers/activations/) was used to elaborate its application in the model, while linear activation functions were used to maintain the linear relationship between input and output in deep learning. Coursera (https://www.coursera.org/) was applied in the proposed model as well. To build an efficient model that could handle complex variable relationships, a deep learning architecture based on the ResNet was used and combined with the attention mechanism to enhance the model's ability to capture important features. The proposed deep learning model was designed to effectively handle complex feature relationships. The model started with two types of input features including numerical features and categorical features. Numerical features went directly into the data preprocessing stage, while categorical features were first converted into continuous vector form through the embedding layer to facilitate the model to understand and utilize their potential information. Subsequently, the core part of the model adopted the design idea of residual modules, which helped to solve the gradient vanishing problem in deep networks and maintain the effectiveness and stability of the model. In addition, an attention mechanism was added, which allowed the model to dynamically focus on the most relevant features, thereby improving the overall prediction accuracy. After the core processing was completed, the model compressed the feature map to a fixed size through a global pooling layer, simplifying the complexity of subsequent operations. The fully connected layer was then responsible for integrating and refining these features to form a higher-level expression, providing a basis for the final decision. At the output layer, the model applied the corresponding activation function according to the type of task such as Softmax for multi-classification tasks or linear activation function for regression analysis to complete the entire prediction process (Figure 1). The input layer received all selected feature variables as input, and the number of input nodes was equal to the number of feature variables. The dimension of each feature depends on the type and number of specific variables. For numerical features, their original values could be used directly, while, for categorical features, encoding conversion was required for processing in the network.

Assuming categorical feature C with V different category values (vocabulary size), an embedding matrix E with shape $V \times D$ could be defined, where D was the dimension of the embedding vector, also called the embedding size. For each category i, the i-th row in the embedding matrix was a vector of length D, representing the embedding representation of the category. If the category feature C corresponded to an integer index i, the corresponding embedding vector e_i could be obtained by using equation (1) below.

$$e_i = E[i,:] \tag{1}$$

The role of the embedding layer was to map discrete category information into a continuous space to better capture the interaction between features as shown in equation (2).

$$\mathbf{e}_i = \mathrm{Embed}(x_i) \tag{2}$$

where $\mathbf{e}_i$ was the embedding vector of the i-th feature. x_i was the corresponding original eigenvalue. The residual module contained multiple residual blocks, each of which contained two convolutional layers with a batch normalization (BN) layer and a ReLU activation function in between. The input was directly transmitted to the next convolutional layer through a skip connection to solve the gradient vanishing problem. This structure could effectively solve the gradient vanishing problem in deep networks and allowed the network to be stacked deeper to capture more complex relationships as follows.

$$\mathbf{y} = \mathrm{ReLU}(\mathrm{BN}(\mathrm{Conv}(\mathbf{x}))) + \mathbf{x} \tag{3}$$

where $\mathbf{x}$ was the input feature. $\mathbf{y}$ was the output of the residual block. The dense module contained multiple dense blocks, each of which contained multiple convolutional layers. The output of each convolutional layer was connected to all subsequent layers to form a dense connection, which enhanced the reuse of internal features of the network, thereby improving the learning ability of the model as shown below.

$$\mathbf{y} = \mathrm{Concat}(\mathrm{Conv}(\mathbf{x}), \mathrm{Conv}(\mathbf{y}), \ldots) \tag{4}$$

where $\mathbf{x}$ and $\mathbf{y}$ were the input and output feature vectors, respectively. An attention module was added after the backbone network to learn the importance of different features and weight the features. The attention mechanism could help the model focus on the most relevant parts of the input features, thereby improving the accuracy of the prediction as shown in equation (5).

$$\alpha_i = \frac{\exp(\mathbf{w}^{\cdot}\,\mathbf{h}_i)}{\sum_{j=1}^{N} \exp(\mathbf{w}^{\cdot}\,\mathbf{h}_j)} \tag{5}$$

where α_i was the attention weight of the i-th feature. $\mathbf{h}_i$ was the hidden layer representation of i features. $\mathbf{w}$ was the weight vector. The global pooling layer used the global average pooling (GAP) layer or the global max pooling (GMP) layer to compress the feature map into a vector of fixed size. The role of the global pooling layer was to reduce the dimension of the feature, while retaining important information, thereby reducing the computational complexity as shown below.

$$\mathbf{z} = \mathrm{GAP}(\mathbf{H}) \tag{6}$$

where $\mathbf{H}$ was the hidden layer representation of all features. $\mathbf{z}$ was the vector after global pooling. Fully connected layers included one or more fully connected layers, which were used to further extract features and ultimately perform classification or regression prediction. Fully connected layers could integrate the features extracted by the previous layers to form a higher level of abstract representation as follows.

$$\mathbf{o} = \mathrm{FC}(\mathbf{z}) \tag{7}$$

where $\mathbf{o}$ was the output of the fully connected layer. $\mathbf{z}$ was the vector after global pooling. The output layer selected an appropriate activation function based on the task requirements. If it was a classification task, the Softmax activation function was used. If it was a regression task, the linear activation function was used. The function of the output layer was to convert the prediction results of the model into an interpretable form for easy evaluation and application as shown in equation (8).

$$p = \mathrm{Softmax}(\mathbf{o}) \tag{8}$$

where p was the predicted probability vector.

Model validation and testing

To verify the effectiveness of the proposed model, comparative experiments with other existing benchmark models including traditional logistic regression, random forest, and support vector machine (SVM) (https://scikit-learn.org/stable/modules/svm.html), as well as more popular deep learning models such as convolutional neural network (CNN) and long short-term memory network (LSTM) were conducted. To comprehensively evaluate the predictive performance of the model, a series of evaluation indicators were examined including accuracy that measured the proportion of correctly predicted samples to the total number of samples, AUC-ROC to measure the performance of the classifier under the imbalance of positive and negative samples, the mean square error (MSE) and the mean absolute error (MAE) that measured the average squared difference and the average absolute difference between the predicted value and the true value for regression tasks, respectively, and F1 score for unbalanced datasets, a harmonic mean of precision and recall, to reflect the performance of the classifier. Confusion matrix provided a detailed classification report, showing the true positive (TP), false positive (FP), true negative (TN), and false negative (FN) of each category to gain a deeper understanding of the performance of the model in different categories. Through the comprehensive use of these indicators, the predictive ability and reliability of the model were evaluated.

Data analysis

To gain a comprehensive and in-depth insight into the characteristics of the dataset, basic statistical analysis on each feature in the dataset was conducted. For numerical features, the minimum value was the smallest value among all the values in the dataset, representing the lower limit of the feature in the sample. The first quartile (Q1) was the value at the 25% position after sorting the data from small to large, which reflected the lower quartile level of the data distribution and reflected the central tendency of the lower part of the data. The median was the value in the middle position after the data was arranged in ascending order (if the number of data was odd, the median was the middle number. If it was even, it was the average of the two middle numbers). The third quartile (Q3) was at the 75% position of the data, showing the central tendency of the higher part of the data. The maximum value represented the upper limit of the feature in the sample. By calculating and analyzing these key statistics, the distribution range, central tendency, and degree of dispersion of the data were identified, laying a solid foundation for subsequent data modeling and analysis.

Results and discussion

Statistical analysis of dataset features

The results showed that the age distribution ranged from 4 to 14 years old with a median of 8 years old, indicating that most of the samples were concentrated in the early childhood stage.

Table 1. Statistical analysis of dataset features.

Feature	Minimum	Q1	Median	Q3	Maximum
age	4.0	6.0	8.0	10.0	14.0
BMI	12.0	15.0	18.5	22.0	35.0
Gender (female = 0, male = 1)	0.0	0.0	0.0	1.0	1.0
Family history (no = 0, yes = 1)	0.0	0.0	0.0	0.0	1.0
Sleep duration (hours)	4.0	6.0	8.0	10.0	12.0
Average daily exercise volume (minutes)	0.0	15.0	30.0	45.0	120.0
Hormone levels (µg/dL)	0.1	0.5	1.0	1.5	2.5
Height (cm)	100.0	120.0	140.0	160.0	180.0
Weight (kg)	15.0	25.0	35.0	45.0	60.0
Regular exercise (number of times/week)	0.0	2.0	3.0	4.0	7.0
Fruit intake (servings/day)	0.0	1.0	2.0	3.0	5.0
Sugar intake (g/day)	0.0	20.0	40.0	60.0	100.0

The median BMI of 18.5 was within the normal range, but the maximum value of 35.0 indicated the existence of overweight or obese individuals. In terms of gender, there were slightly more female than male. Family history demonstrated that most patients had no family genetic history. The median sleep duration of 8 hours met the health standard, but there was a large individual difference. The average daily exercise volume indicated that most children had a certain amount of physical activity. The height range of 100 - 180 cm and the weight range of 15 - 60 kg reflected the different growth and development stages of children. The data related to living habits reflected the diversity of children's daily behavior patterns. These dataset characteristics reflected the diversity and complexity of the samples, providing rich information for subsequent model training and analysis. The potential relationships between factors such as age, BMI, living habits and precocious puberty needed to be considered during model construction. The distribution characteristics of gender and family genetic history in the samples provided clues for studying the influencing factors of precocious puberty.

Comparison of model performance

The changing trends of the two performance indicators, accuracy and AUC-ROC values, during multiple iterations of the model showed that, with the increase in the number of iterations, both the accuracy and AUC-ROC values demonstrated an upward trend. Specifically, the accuracy grew faster in the early stage, then gradually slowed down, and finally stabilized. The AUC-ROC value also showed a similar upward trend throughout the iteration process, but the growth rate was slightly faster than that of the accuracy. In addition, both indicators had been significantly improved at certain specific iteration points such as at the 6th and 8th iterations, which indicated that, in these two iterations, the adjustment of the model or the optimization of parameters had achieved good results (Figure 2). On the training set, the proposed model had the highest accuracy of 92% and an AUC - ROC of 0.95, surpassing logistic regression, random forest, and support vector machine, while, on the validation set, the model maintained an accuracy of 91% and an AUC - ROC of 0.94. On the test set, the accuracy was 90% and the AUC - ROC was 0.93, both leading other models (table 2). The excellent performance of the proposed model on various datasets demonstrated its strong fitting ability and good generalization ability. It could effectively learn data features during the training stage and make reliable predictions when facing new data, providing a more accurate tool for predicting the efficacy of TCM intervention in childhood precocious puberty. The comparison results with other models highlighted the advantages of the proposed model in addressing this issue, which was expected to provide strong

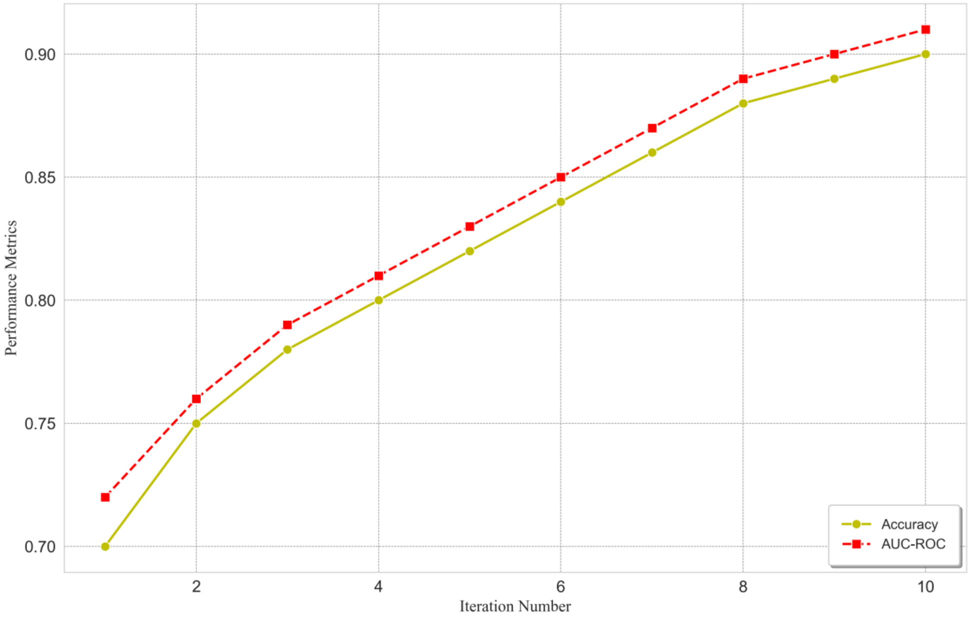

Figure 2. Model performance.

Table 2. Comparison of model performance.

Dataset	Logistic regression	Random forest	SVM	Proposed model
Training set	0.85 / 0.88	0.88 / 0.90	0.83 / 0.85	0.92 / 0.95
Validation set	0.84 / 0.87	0.87 / 0.89	0.82 / 0.84	0.91 / 0.94
Test Set	0.83 / 0.86	0.86 / 0.88	0.81 / 0.83	0.90 / 0.93

decision - making support for clinical practice.

The research found that the sample characteristics were diverse. Age was mainly concentrated in the early childhood stage, but there were also some adolescents. The BMI statistics showed that most children had normal weight, but there were some cases of overweight or obesity. The gender distribution was slightly more females than males, which was consistent with the higher incidence of precocious puberty in females. The characteristics of living habits reflected the diversity of children's daily behavior patterns, emphasizing the importance of a healthy lifestyle. The analysis of family genetic information indicated that, although most patients had no family genetic history, genetic factors still could not be ignored in the occurrence of precocious puberty. The results of syndrome differentiation based on TCM theory showed that different syndromes were distributed in the patients with "liver depression transforming into fire" being the most common one followed by "kidney qi deficiency", which was consistent with the understanding of the causes of precocious puberty in TCM theory. These comprehensive characteristics provided important background information for evaluating the efficacy of TCM intervention in childhood precocious puberty and laid the foundation for subsequent model training and verification. The relationships between different characteristics and precocious puberty were complex. Therefore, the model needed to fully consider these factors to improve prediction accuracy. The results of TCM syndrome differentiation provided a theoretical basis for the formulation of TCM treatment plans and the evaluation of treatment efficacy.

Conclusion

Precocious puberty in children not only affects physical health but also may lead to psychological and social adaptation problems. In recent years, its incidence has been gradually increasing worldwide. This study proposed a new prediction model using a deep - learning architecture based on ResNet by combining with the attention mechanism to evaluate the efficacy of TCM intervention in childhood precocious puberty through machine learning technology combined with the rich practical experience of traditional Chinese medicine. A series of feature variables closely related to the treatment effect of traditional Chinese medicine were selected through feature engineering, and the prediction ability of the model was improved through reasonable feature conversion and combination. The results indicated that the proposed model performed excellently on the training set, validation set, and test set. Especially on the test set, the accuracy rate reached 90% and the AUC - ROC was 0.93, verifying the model's fitting ability on the training data and its generalization ability on unknown data. The model also showed good stability, providing a reliable decision-support tool for clinicians when formulating personalized treatment plans.

Acknowledgements

This study was supported by National Famous Traditional Chinese Medicine Experts Inheritance Studio Construction Project (National Chinese Medicine Education letter (2022) No. 75); Anhui Leading Talents Plan of Traditional Chinese Medicine (Development of traditional Chinese Medicine (2018) No.23); and Anhui Key Clinical Construction Specialist Project (Wanwei Zhuan (2019) 91).

References

1. Han XX, Zhao FY, Gu KR, Wang GP, Zhang J, Tao R, *et al*. 2022. Development of precocious puberty in children: Surmised medicinal plant treatment. Biomed Pharmacother. 156:113907.
2. Polat R, Çokluk E, Budak Ö, Tuncer FB. 2022. Effect of propolis on precocious puberty in female rats. J Clin Res Pediatr Endocrinol. 14(4):415-421.
3. Aftab S, Manzoor J, Mahmood Q, Shaheen T. 2020. Precocious puberty: The clinical profile and the etiological classification of children presented at a tertiary care children's hospital. Pak J Med Sci. 38(4):955-959.
4. Gangat M, Radovick S. 2020. Precocious puberty. Minerva Pediatr. 72(6):491-500.
5. He YY, Han XH, Sun W, Yu J. 2019. The therapeutic effect of a traditional Chinese medicine mixture in rat models with precocious puberty through Lin 28/Let 7 pathway. Horm Res Paediatr. 91:235-236.
6. Lee HL, Lee YB, Choi JY, Lee JA. 2018. Herbal medicine for idiopathic central precocious puberty: A protocol for a systematic review of controlled trials. Medicine. 97(13):e0267.
7. Jiang Y, Yang L, Chen H, Chen J, Yang L, Wang Z, *et al*. 2024. Network pharmacology combined with lipidomics to reveal the regulatory effects and mechanisms of Kangzao granules in the hypothalamus of rats with central precocious puberty. J Pharm Biomed Anal. 242:116059.
8. Eugster EA. 2019. Update on precocious puberty in girls. J Pediatr Adolesc Gynecol. 32(5):455-459.
9. Señaris JAA, Diago CAA. 2020. Male precocious puberty described in literature of the Spanish Golden Age. Hormones-Int J Endocrinol Metab. 19(3):449-450.
10. Wu X, Wang L, Xue P, Tang J, Wang H, Kong H, *et al*. 2024. Association of screen exposure/sedentary behavior and precocious puberty/early puberty. Front Pediatr. 12:1447372.
11. Aguirre RS, Eugster EA. 2018. Central precocious puberty: From genetics to treatment. Best Pract Res Clin Endocrinol Metab. 32(4):343-354.
12. Chen X, Fu S, Chen C, Yuan Y, Dai Z, Chen A, *et al*. 2024. Association of Traditional dietary pattern with early and precocious puberty: A population-based cross-sectional study. Pediatr Res. 96(1):245-252.
13. Tsai MM, Chen MH, Chien YH, Tung YC. 2023. Precocious puberty in patients with Pompe disease. Front Endocrinol. 14:1150498.
14. Fuqua JS, Eugster EA. 2022. History of puberty: Normal and precocious. Horm Res Paediatr. 95(6):568-578.
15. Gohil A, Eugster EA. 2020. Delayed and precocious puberty: Genetic underpinnings and treatments. Endocrinol Metab Clin North Am. 49(4):741-754.
16. Chen H, Mo CY, Zhong LY. 2023. Central precocious puberty secondary to peripheral precocious puberty due to a pineal germ cell tumor: a case and review of literature. BMC Endocr Disord. 23(1):237.
17. Chen XH, Zheng M, Ma XL, Fei XH. 2023. Analysis of the efficacy of Dabuyin pill combined with gonadotropin-releasing hormone analogue in the treatment of central precocious puberty girls based on network pharmacology. Transl Pediatr. 12(3):364-374.
18. Palhano ACM, Kim LJ, Moreira GA, Coelho FMS, Tufik S, Andersen ML. 2018. Narcolepsy, precocious puberty and

obesity in the pediatric population: A literature review. Pediatr Endocrinol Rev. 16(2):266-274.

19. Chen GY, Wang LZ, Cui Y, Liu JC, Wang LQ, Wang LL, *et al*. 2023. Serum metabolomic analysis reveals key metabolites in drug treatment of central precocious puberty in female children. Front Mol Neurosci. 15:972297.