Exploring the Relationship between Parenting Style and College Students' Mental Health based on the SVM Algorithm

DAI HU*, MINGSHAN CHANG
Krirk University, Bangkok, 10220,
THAILAND

*Corresponding Author

Abstract: - This study aims to explore the relationship between parental parenting style and the mental health of college students based on the support vector machine (SVM) algorithm. Collect data on parenting style and mental health status through a questionnaire survey of college students. Using the SVM algorithm to analyze and process data, to reveal the impact mechanism of different parenting styles on the mental health of college students. The research results contribute to a deeper understanding of the role of parenting style in the development of college student's mental health, providing the scientific basis and practical guidance for college students' mental health education, and promoting their comprehensive and healthy growth.

Key-Words: - SVM algorithm, Parenting style, College students' mental health, Psychological resilience, Prediction accuracy, Recall.

Received: May 29, 2024. Revised: December 26, 2024. Accepted: January 25, 2025. Published: May 5, 2025.

1 Introduction

In today's society, college students' mental health problems are an increasingly widespread concern. Mental health not only affects college students' academic achievement, social life, and future development but also has far-reaching significance for the talent training and stability of the whole society. At the same time, parental rearing patterns, as the key environmental factors in the process of individual growth, play a fundamental role in shaping personality, values, and psychological traits. At present, China's college student population is generally faced with many mental health problems such as high learning pressure, social difficulties, emotional disorders, and so on. According to relevant surveys, about 40% of college students have different degrees of mental health problems, of which depression, anxiety, and other conditions are more common, [1]. College students are at a critical stage of personality formation and value establishment, and good psychological quality is crucial for the healthy growth of college students. Studies have shown that parenting style has a significant impact on children's mental health, [2]. The authoritative parenting style, characterized by the combination of care and reasonable requirements, is often associated with children's positive psychological development, such as high self-esteem, good social skills, and strong psychological toughness. On the contrary, an authoritarian parenting style, which emphasizes strict control and high requirements, may lead to children's anxiety, inferiority complex, and other adverse psychological conditions. However, the indulgent parenting style, due to excessive care and few requirements, makes it easy for children to lack independence and self-control ability, which may cause psychological distress in the face of various challenges in the University. Neglectful parenting style and parents' indifference to their children's emotions and needs may make their children feel inadequate support and guidance in the process of growth, and then affect their level of mental health. Therefore, it is of great theoretical and practical significance to explore the intrinsic relationship between parenting style and college students' psychological health in order to formulate targeted psychoeducational measures.

In terms of research methods, traditional statistical analysis methods were used to explore the relationship between parental rearing patterns and college students' mental health. However, with the continuous expansion of data scale and the increasing complexity of data structure, traditional methods have some limitations in dealing with multi-dimensional and non-linear relationships. As a supervised machine learning algorithm, Support Vector Machine (SVM) has been widely used in the fields of image recognition and speech processing, [3]. In recent years, the SVM algorithm has also been gradually applied in the field of education, such as student performance prediction, teaching quality assessment, etc. SVM algorithm provides a new technical means to analyze educational data by virtue of its excellent classification performance, [4]. Therefore, this study aims to use the SVM algorithm to explore the internal relationship between parental rearing patterns and college students' mental health, in order to provide a more accurate and scientific basis for college students' mental health education and intervention, and further enrich and expand the research results in this field.

2 Research Subjects and Methods

2.1 Research Object

In this study, publicly-funded teacher training students in Hunan Province, China, were selected between June 2023 and August 2023, and the study participants signed an informed consent form. A total of 300 questionnaires were distributed in this study, and 263 valid questionnaires were returned, with a validity rate of 87.7%.

2.2 Research Tools

2.2.1 Symptom Checklist 90 (SCL-90)

The psychological assessment scale consists of 90 items, with scores ranging from 1 to 5 for each item. The scale covers 10 factors, including somatization (12 items), obsessive-compulsive symptoms (10 items), depression (13 items), anxiety (10 items), paranoia (6 items), psychoticism (10 items), interpersonal sensitivity (9 items), phobia (7 items), hostility (6 items) and other (7 items), and is used to assess the individual's mental health status. The scale is rated on a five-point scale, with any factor score above 2 being considered positive for mental health, with higher scores representing poorer mental status. The reliability coefficient of the total scale is 0.954, and the average reliability coefficient of each factor is about 0.8, which is a good indicator of reliability and validity, [5].

2.2.2 Parental Bonding Instrument (PBI)

The scale is divided into two parts, the father's version and the mother's version, with 21 questions in each part, totaling 42 questions with a high degree of consistency between the questions, and a total of three dimensions: rejection, emotional warmth and overprotection, [6]. The items cover six factors, father encouraging autonomy (6 items), father caring (11 items), father controlling (6 items), mother caring (11 items), mother encouraging autonomy (6 items), and mother controlling (6 items), and are used to assess individuals' perceived self-reports of parenting style up to the age of 16. The scale had good internal consistency (Cronbach's alpha coefficients of 0.745 to 0.858) and retest reliability (0.746 to 0.941).

2.3 Research Methods

2.3.1 Support Vector Machine Algorithm

SVM, a very important branch in the field of machine learning, is a supervised learning algorithm that is mainly used to solve classification problems, especially binary classification problems, [7]. It finds an optimal hyperplane in a multidimensional space by finding an optimal hyperplane based on a given training set, which makes the hyperplane able to classify the test set more easily. At the same time, SVM has many advantages such as strong classification ability, the ability to handle unbalanced data, and obtain a more stable model, which is especially outstanding in the case of small sample size, data that is nonlinear, or high dimensionality, [8]. It has been widely used in both data mining and pattern recognition fields.

SVM is a powerful method for constructing classification models, which can better handle high-dimensional, nonlinear mental health data than traditional models such as linear regression and neural networks. SVM is also relatively more stable to noise and outliers and is able to accurately capture complex relationships between samples, [9]. In addition, the SVM model itself has good interpretability, which helps psychologists to deeply understand the key factors affecting mental health. The main idea of SVM is to be able to predict labels from one or more feature vectors by finding the optimal decision boundary between two categories and to make the distance between the sample points of the two categories that are close to the decision boundary as large as possible on both sides of the boundary, so as to make the classification model's generalization ability stronger. This decision boundary is known as the hyperplane and is positioned as far away as possible from the nearest data point of each category so that the interval of the decision boundary is maximized. The points closest to the decision boundary are called support vectors.

Given a labeled set of training samples M = {(x1, y1), (x2, y2), ..., (xn, yn)}, where yi ∈ {-1, +1}, xi denotes the feature vectors and yi denotes the category labels, the optimal hyperplane for classification learning is defined by the following mathematical model:

$$w^T x + b = 0 \quad (1)$$

The weight vector w determines the direction of the decision hyperplane,. The input feature vector x, on the other hand, represents the attribute information of the sample, and the intercept b determines the position of that hyperplane with respect to the coordinate origin. For any sample point, its distance to the classification hyperplane can be computed by representing the feature vector x and model parameters w and b for that point:

$$r = \left| w^T x + b \right| / \| w \| \quad (2)$$

The core task of a Support Vector Machine (SVM) is to find the optimal weight vectors and intercept terms to construct hyperplanes that maximise the classification boundaries and separate the dataset efficiently. Its basic mathematical form can be expressed as:

$$\begin{gathered} \min_{w,b,\xi} \frac{1}{2} \| w \|^2 + C \sum_{i=1}^{m} \xi_i \\ s.t. y(wx_i) + b \geq 1 - \xi, ... \xi_i \geq 0, ... i = 1,2,...,m \end{gathered} \quad (3)$$

C denotes the coefficient of penalty strength used for the penalty term and ξ represents the relaxation factor. The original problem can be converted to its dual form by applying the Lagrange multiplier method:

$$\begin{gathered} \max_{a} \sum_{i=1}^{m} a_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} a_i a_j y_i y_j x_i^T x_j \\ s.t. \sum_{j=1}^{m} a_i y_i = 0, 0 \leq a_i \leq C, i = 1,2,...,m \end{gathered} \quad (4)$$

In order to achieve effective differentiation of data in high-dimensional space for classification or regression purposes, it is necessary to map the original data to the high-dimensional space by means of a kernel function. The kernel function is expressed in the following four forms:

(1) Linear kernel, whose expression is $k(x_i, x_j) = x_i^T x_j$, is suitable for solving linear problems.

(2) Polynomial kernel, the expression is $k(x_i, x_j) = (x_i^T x_j)^d$, when the number of times d is 1, it is equivalent to the linear kernel function; when d to take a higher number of times, it can be used to solve non-linear problems.

(3) Gaussian kernel with expression $k(x_i, x_j) = \exp(-\frac{\| x_i, x_j \|}{2\sigma})$, which can be applied to nonlinear problems.

(4) The sigmoid kernel has an expression of $k(x_i, x_j) = \tanh(\beta x_i^T x_i + \theta)$ and can also be used for solving nonlinear problems.

There is no fixed criterion for the selection of kernel functions in support vector machine models. Usually, researchers need to select the kernel function that can make the model performance optimal by means of model training. The introduction of the kernel function can achieve the inner product operation of two samples in a high-dimensional space, and the final classification function is obtained as follows:

$$f(x) = \sum_{i=1}^{m} a_i y_i k(x, x_i) + b \quad (5)$$

2.3.2 Model Evaluation Indicators

Model evaluation is a key step in the machine learning process that visualises the strengths and weaknesses of a model. This study aims to predict the mental health status of college students, which is a dichotomous problem. The following will briefly introduce the commonly used evaluation metrics for dichotomous models.

For dichotomous classification problems, a confusion matrix is usually used to describe the classification results. Based on the combination of true labels and model-predicted labels, the samples can be classified into four categories; True Positive (TP), i.e., actually positive and predicted to be positive as well; False Positive (FP), actually non-positive but predicted to be positive; False Negative (FN), actually positive but predicted to be non-positive; and True Negative (TN), actually non-positive and predicted to be non-positive as well. These four cases form the confusion matrix shown in Table 1.

Table 1. Confusion matrix of binary classification results

Actual results	Forecast results	
	1(Positive)	0(Non positive)
1(Positive)	TP	FN
0(Non positive)	FP	TN

The following rate metrics can be derived from the confusion matrix based analysis:

True Positive Rate (TPR), also known as sensitivity, indicates the proportion of the number of samples predicted to be positive by the classification model to the actual number of positive samples. Its formula is calculated as follows:

$$TPR = \frac{TP}{TP + FN} \tag{6}$$

The False Positive Rate (FPR) refers to the proportion of non-positive samples that the model incorrectly predicts as positive. Its formula is:

$$FPR = \frac{FP}{FP + TN} \tag{7}$$

The False Negative Rate (FNR) is the proportion of true positive samples that are predicted as negative by the classification model. Its formula is:

$$FNR = \frac{FN}{FN + TP} \tag{8}$$

The true negative rate (TNR) represents the proportion of the number of non-positive samples correctly predicted as non-positive by the model to the actual number of non-positive samples, with the formula:

$$TNR = \frac{TN}{FP + TN} \tag{9}$$

Accuracy is a metric used to assess the performance of a classification model. It reflects the proportion of samples correctly predicted by the model, with higher values indicating better model predictions. Its calculation formula is:

$$accurary = \frac{TP + FN}{TP + FP + TN + FN} \tag{10}$$

Precision is a measure of the classification accuracy of positive samples, i.e. how many of the samples predicted by the classification model to be positive are true positive positive samples. It is calculated by the formula:

$$precision = \frac{TP}{TP + FP} \tag{11}$$

Recall refers to the proportion of samples that are actually positive in real data that are correctly predicted as positive by the classification model, and is complementary to precision. It is calculated by the formula:

$$recall = \frac{TP}{TP + FN} \tag{12}$$

The F1-score is the result obtained by combining precision and recall. The F1 value is calculated as:

$$F1 - score = \frac{2 * precision * recall}{precision + recall} \tag{13}$$

2.3.3 ROC Curve

The ROC curve is commonly used to evaluate the performance of binary classification models. The curve describes the relationship between the true positive rate and the false positive rate. The curve is plotted using FPR as the horizontal coordinate and TPR as the vertical coordinate. The closer the curve is to the upper left corner, the better the model's classification performance is; conversely, if the curve is closer to the lower right corner, the model's classification performance is less good. Overall, the ROC curve provides an intuitive way for researchers to assess the overall performance of the constructed binary classifiers.

The AUC value represents the area under the ROC curve, and this metric is a more intuitive reflection of the performance of the classification model. Its value range is generally between 0.5 and 1, with larger values implying a stronger classification ability of the model, [10]. In practice, the performance of the model can be roughly judged based on the interval of the AUC value: when the AUC is between 0.5 and 0.7, the classification performance of the model is weak; between 0.7 and 0.9, the classification performance of the model is strong; and when it is higher than 0.9, the classification performance of the model is excellent. In view of the fact that this study involves binary classification problems, combining the research objectives and data characteristics, the accuracy, F1 value, precision rate, recall rate, and AUC value are finally selected as the evaluation index system.

3 Results

3.1 Basic Information

Of the 263 students selected by the questionnaire, 74 (28.1%) were male and 189 (71.9%) were female. Among them, the maximum age was 24 years old, the minimum age was 17 years old, and the mean age was (22.37±1.12) years old. Self-acceptance questionnaire mental health status was positive in 92 cases (35.0%) and non-positive in 171 cases (65.0%). The distribution of parenting style questionnaire scores is detailed in Table 2.

Table 2. Distribution of Scores on the Parenting Style Questionnaire

Variable	Minimum value	Maximum value	Average value	Standard deviation
Mother's care	0	33	16.42	2.36
Mother encourages autonomy	0	18	12.78	2.86
Mother's control	0	18	4.56	1.15
Father's care	0	33	17.48	2.88
Father encourages autonomy	0	18	12.13	2.47
Father's control	0	18	4.25	1.65

3.2 Establishment and Results of Support Vector Machine Model

In the experimental study using Jupyter Notebook, we use Support Vector Machine (SVM) algorithm to construct the classification model. The core parameters of the SVM model include the penalty term coefficient, C, and the kernel function parameter, gamma. the C value reflects the model's tolerance of errors; when C is high, the model is stricter about the occurrence of errors; on the contrary, when C is low, the model can accept the occurrence of errors more easily. The gamma value, on the other hand, affects the number of support vectors; a larger gamma means fewer support vectors, which is prone to overfitting; a smaller gamma is prone to underfitting.

The experimental data processed in this study has a total of 20 feature variables, and the target variable is the mental health status of college students. We divided the dataset into a training set and a validation set, and used the training set to tune the parameters of the SVM model. Specifically, we used grid search combined with cross-validation to traverse different combinations of values of C and gamma and select the parameter configuration with the highest accuracy rate. The final SVM model evaluation metrics obtained are shown in Table 3, including accuracy, recall and F1 value. Through the careful adjustment of the core parameters, a college student mental health prediction model with excellent performance was successfully constructed. The specific parameter adjustment results are shown in Table 3.

Based on the results in Table 3, it can be seen that this support vector machine model performs optimally when the RBF kernel function is used and the penalty parameter C is set to 7 and the gamma parameter is set to 0.05. We constructed the support vector machine model with this optimal parameter configuration and evaluated it on the test set. The test results are shown in the confusion matrix in Table 4, and the ROC curve shown in Figure 1 is also plotted.

Table 3. Optimal Parameters of SVM Algorithm under Grid Search and Cross-Validation

Parameter	Parameter meaning	Parameter range	Result
Kernel	kernel function	[linear，sigmoid，poly]	rbf
C	Punishment coefficient of the objective function	[5，6，7，8，9]	7
gamma	The coefficients of the kernel function	[0.01，0.03，0.05，0.07，0.1]	0.05

Table 4. Confusion Matrix of SVM Algorithm

Actual results	Forecast results	
	1(Positive)	0(Non positive)
1(Positive)	65	27
0(Non positive)	17	154

The sample size of this study was 263 cases, of which 92 were individuals with positive mental health status and 171 were non-positive individuals. The prediction results showed that the prediction accuracy of individuals with positive mental health status was 70.7%, i.e., 65 cases were correctly predicted and 27 cases were incorrectly predicted, while the prediction accuracy of individuals with non-positive mental health status was 90.1%, i.e., 154 cases were correctly predicted and 17 cases were incorrectly predicted. When the proportion of non positive and positive mental health cases is close to 2:1, it has many effects on the performance of the model. For example, during training, the model will tend to learn its characteristics due to more non positive cases, and it is easy to over fit the non positive mode, resulting in insufficient learning of the characteristics of positive cases. This makes the model may have more errors in identifying positive cases, and the false negative rate increases.

In total, the model correctly predicted 219 samples and incorrectly predicted 44 samples. The overall prediction accuracy was 83.3%. Further, ROC curve analysis was plotted (Fig1). The calculated AUC value is 0.9067, indicating that the SVM classification model has high recognition performance.

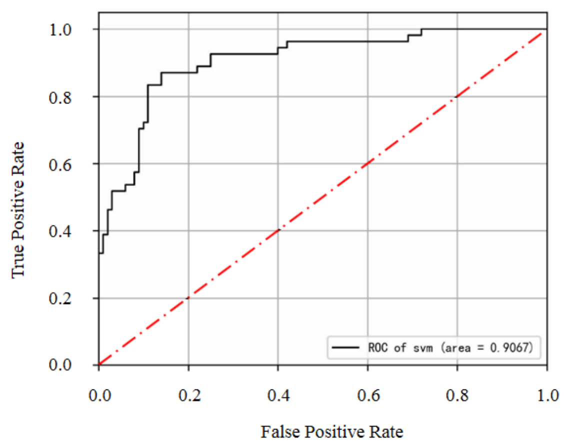

Fig. 1: ROC curve of SVM algorithm

4 Discussion

With the continuous expansion of college enrollment and the frequent reports of self-harm and even suicide caused by poor mental health among college students in recent years, the mental health status of college students urgently needs the attention of schools and society, [11]. In order to prevent negative psychological states from further developing into emotional disorders such as depression that seriously affect daily life, early identification and diagnosis of the mental health status of college students and timely implementation of corresponding intervention measures are of great significance. There are studies showing that machine learning has developed rapidly in various fields such as healthcare and industry in recent years, and has gradually been used for identifying and predicting mental health states. However, there is relatively little research in the field of psychology in China at present. In this study, 263 normal college students in Hunan Province were selected as the research object, and their family upbringing and mental health status were counted by questionnaire. The mental health status was predicted by the SVM model. The results showed that 92 cases were positive and 171 cases were nonpositive. The prediction accuracy of the SVM model for positive individuals with mental health status was 70.7%, and the prediction accuracy of nonpositive individuals with mental health status was 90.1%.

Which accurately identifies the mental health status Therefore, based on the collected data, this study establishes an SVM model for predicting mental health status, which accurately identifies the mental health status of publicly funded normal students in Hunan Province to a certain extent, provides new ideas for large-scale screening of mental health, and provides a theoretical basis for early identification and timely intervention. It has practical significance for improving the mental health level of college students.

5 Conclusion

This study used SVM algorithm to investigate 263 normal students in Hunan Province, and analyzed the relationship between parental rearing patterns and college students' mental health. The results showed that the gender ratio of the sample was unbalanced, and the scores of each dimension of parental rearing style were different. The optimal parameters of SVM were determined by grid search and cross validation. The overall prediction accuracy of the constructed model in the test set was 83.3%, and the AUC value was 0.9067, with good recognition performance. This study provides new ideas and theoretical support for college students' mental health screening.

6 Future Research Directions and Suggestions

The selected samples are only from publicly funded normal students in Hunan Province, and there are large differences in the number of male and female samples, and the survey scale is relatively single, which may affect the results of the study to a certain extent. In future research, we will expand the sample size and the sampling range of different subjects, and pull parents into the survey, so that we can grasp the way of parent-child communication, more comprehensively discuss parental rearing patterns, and more carefully and deeply explore the relationship between College Students' mental health and parental rearing patterns.

Declaration of Generative AI and AI-assisted Technologies in the Writing Process

The authors wrote, reviewed and edited the content as needed and they have not utilised articial itelligence (A) tools. The authors take full responsibility for the content of the publication.

References:

[1] Iparraguirre-Villanueva O, Paulino-Moreno C, Epifanía-Huerta A, Torres-Ceclén C. Machine Learning Models to Classify and Predict Depression in College Students. *International Journal of Interactive Mobile*

Technologies, 2024, 18(14): 148. DOI: 10.3991/ijim.v18i14.48669.

[2] Hancock JT, Khoshgoftaar TM. CatBoost for big data: an interdisciplinary review. *Journal of Big Data*, 2020, 7(1): 94. DOI: 10.1186/s40537-020-00369-8.

[3] Luo H, Meng X, Zhao YCM. Exploring the impact of sentiment on multi-dimensional information dissemination using COVID-19 data in China. *Computers in human behavior*, 2023, 144(Jul.): 1.1-1.21. DOI: 10.1016/j.chb.2023.107733.

[4] Zhang T, Zhong Z, Mao W, Zhang Z, Li Z. A New Machine-Learning-Driven Grade-Point Average Prediction Approach for College Students Incorporating Psychological Evaluations in the Post-COVID-19 Era. *Electronics*, 2024, 13(10): 1928. DOI: 10.3390/electronics13101928.

[5] Song SI, Hong HT, Lee C, Lee SB. A machine learning approach for predicting suicidal ideation in post stroke patients. *Scientific Reports*, 2022, 12(1): 15906. DOI: 10.1038/s41598-022-19828-8.

[6] Gino F, Brooks AW, Schweitzer ME. Anxiety, advice, and the ability to discern: feeling anxious motivates individuals to seek and use advice. *Journal of Personality & Social Psychology*, 2012, 102(3): 497-512. DOI: 10.1037/a0026413.

[7] Kleef GAV, Anastasopoulou C, Nijstad BA. Can expressions of anger enhance creativity? A test of the emotions as social information (EASI) model. *Journal of Experimental Social Psychology*, 2010, 46(6): 0-1048. DOI: 10.1016/j.jesp.2010.05.015.

[8] Giarrusso R, Feng, Du, Silverstein M, Bengtson VL. Grandparent—adult grandchild affection and consensus: Cross-generational and cross-ethnic comparisons. *Journal of Family Issues*, 2001, 22(4): 456-477. DOI: 10.1177/019251301022004004.

[9] Arkaprabha S, Ishita B. Screening of anxiety and depression among seafarers using machine learning technology. *Informatics in Medicine Unlocked*, 2019, 16: 100149. DOI: 10.1016/j.imu.2019.100228.

[10] Bhadra S, Kumar CJ. An insight into diagnosis of depression using machine learning techniques: a systematic review. *Current Medical Research and Opinion*, 2022, 38(5): 749-771. DOI: 10.1080/03007995.2022.2038487.

[11] Priya A, Garg S, Tigga NP. Predicting Anxiety, Depression and Stress in Modern Life using Machine Learning Algorithms. *Procedia Computer Science*, 2020, 167: 1258-1267. DOI: 10.1016/j.procs.2020.03.442.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

The authors equally contributed in the present research, at all stages from the formulation of the problem to the final findings and solution.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

No funding was received for conducting this study.

Conflict of Interest

The authors have no conflicts of interest to declare.