

Differential Analysis of Adolescent Mental Health Early Warning Levels via Machine Learning

ManLin Wang

CEFLS Senior High School, Tianfu New Area, Chengdu, 610213, China

ManlinWang123@163.com

Abstract. Drawing on the citywide survey of 95 545 adolescents reported by Zhao et al., we find that school bullying is a key driver of deteriorating youth mental health. The pression, anxiety, stress-related symptoms and psychotic-like experiences are all rising, while the age of first depressive episode keeps falling. To quantify these links, the current study conducts a differential analysis of mental health early warning levels across middle and high school, systematically comparing alternative baseline distribution and multi-dimensional features to produce an evidence-based metric linking risk influential factors to graded alert thresholds. These results demonstrate that proactive bullying prevention and early intervention strategies are critical to reversing the trajectory of declining adolescent mental health and reducing the incidence of early-onset depression.

Keywords: Psychological factors; feature selection; data analysis; machine learning

1. Introduction

Adolescence is a pivotal window for physical and mental maturation. Probing this issue in depth will allow timely, evidence-based intervention for youth at risk, fostering the steady, healthy growth of the nation's future backbone. However, there is currently insufficient discussion and research on the factors affecting the mental health of adolescents. As a result, with respect to the Research Questions and Objectives, the key question is to identify which factors influence adolescent psychological early warning performance and to what extent these factors affect the final outcomes.

Current empirical evidence, as demonstrated in Zhao et al.'s citywide survey of 95,545 adolescents, reveals a critical mental health crisis with school bullying driving rising rates of depression, anxiety, and psychotic-like symptoms, while the age of first depressive episode continues to decline [1]. This pattern aligns with the cumulative risk framework, where multiple risk factors such as adverse social environments [2], early childhood adversities [3], and family risk profiles across socioeconomic strata [4-6] significantly contribute to the deterioration of adolescent mental health. To address this urgent situation, schools and communities should implement evidence-based anti-bullying curricula informed by comprehensive understanding of bullying dynamics and integrate routine mental health screenings into educational systems [7, 8]. This integrated approach, supported by both large-scale survey data and longitudinal risk factor analyses, represents a critical pathway toward reversing the trajectory of declining adolescent mental health. For that, the two primary objectives are conducted in this study. First, we preliminarily explore the strength and direction of the relationship between each potential influencing factor and different psychological warning levels through alternative baseline distribution. And then, we quantify the specific impact weights of different factors on mental health based on machine learning models.

2. Dataset Description

The dataset covers 68 300 students aged 12–22 enrolled in 73 middle schools (including nine-year schools, 43 124 students) and 14 high schools (25 203 students). It is obtained from the citywide survey reported by Zhao et al. [1] by excluding the students in primary school. We use preprocess pipeline to improve the data quality, and rescale the numerical variables. The dependent variable is the mental health early warning levels, of which the values is ranged from 0 (healthy) to

4 (very strong). Figure 1 presents the distributions of dependent variable. It shows that among 43,124 middle-school students, Level-1 (3,938) and Level-4 (3,616) alerts account for the bulk of risk, warranting urgent attention; by contrast, high-school risk is concentrated in Level-1 alone, with Levels 2–4 contributing only a minor share and yielding a markedly milder overall profile. Most strikingly, Level-4 prevalence in middle school (8.4 %) is almost 70 % higher than in high school (5 %), flagging junior-high populations as the critical intervention target.

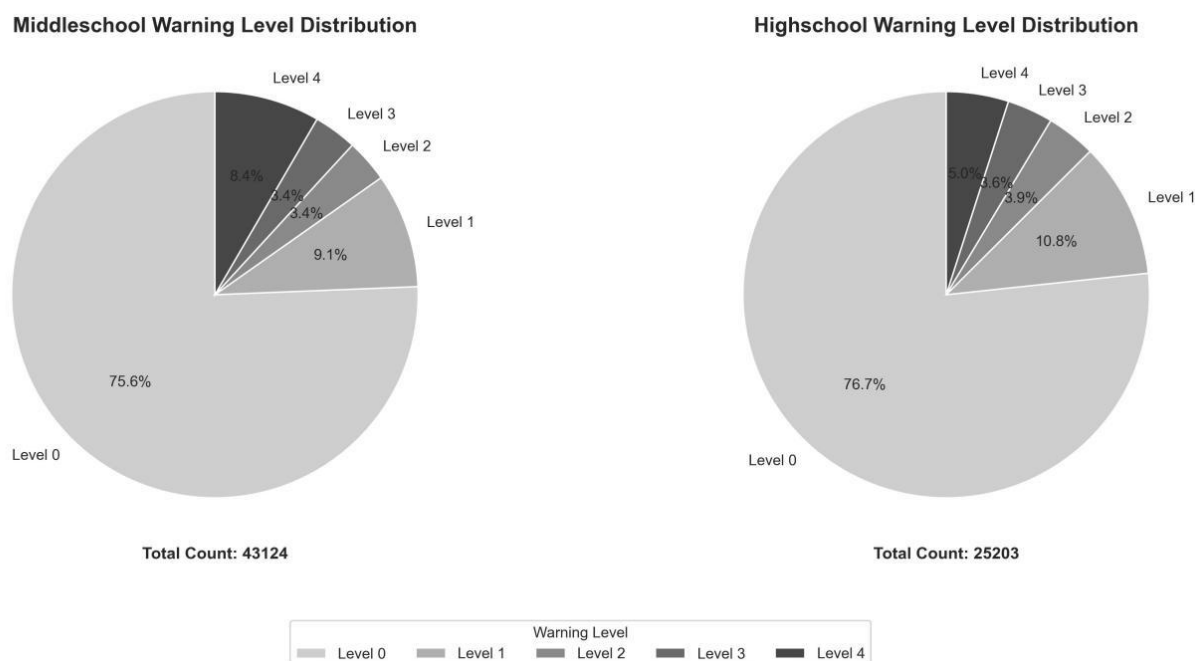


Figure 1 Distributions of mental health early warning levels in middle and high school.

It also includes two classes of influential factors (i.e., independent variables), 11 family-context factors and 21 individual factors. The first class has, high/low parental education, parental alcoholism/smoking, only child, twins, left-behind status, reconstituted-family, single-parent household, long-term living environment and grand-parent rearing. The second class has academic stage, age, stress response, psychological resilience, gender, emotional volatility, emotional behavior, smartphone dependence, depression, aggression/rebellion, inattention, anxiety, sleep quality, social avoidance, positive mental health, psychotic-prodrome risk, internet addiction, online gaming disorder, exposure to social manipulation, verbal abuse, physical victimization. According to the statistical analysis in Figure 1, we can see that despite heavier academic pressure in high school, depression is more severe among the city's middle-schoolers—likely driven by the perfect storm of early-puberty upheaval, immature emotional coping, and the simultaneous shock of new campuses, teachers, peers and dormitories. Thus, the following differential analysis focuses on the 8 family-context factors and 8 individual mental factors.

3. Differential Analysis of Key Influential Factors

We firstly make a differential analysis of 8 family-context factors for 43,124 middle-school student and 25,203 high-school student adolescents. The differential analysis in Figure 2 yields the following concise findings: 1) The left-behind (ratio 0.341 and 0.390), grandparent-raised (ratio 0.387 and 0.369) and long-term-living-environment (ratio 0.329 and 0.306) factors have a significant impact on the mental health in both middle and high school. As the age increases, the high-school students are slightly reduced by the lack of parental care, yet scarred more deeply by the graver harm of being left behind. 2) The parental-alcoholism (ratio 0.165 and 0.183) and only-child (ratio 0.223 and 0.207) factors also properly affect the mental health of both middle- and

high-school students indiscriminately. 3) The reconstituted-family (ratio 0.079 and 0.072) and twin (ratio 0.094 and 0.06) non-trivial affect warning level. 4) Parents with high education slightly release expectations and barely affects warning level.

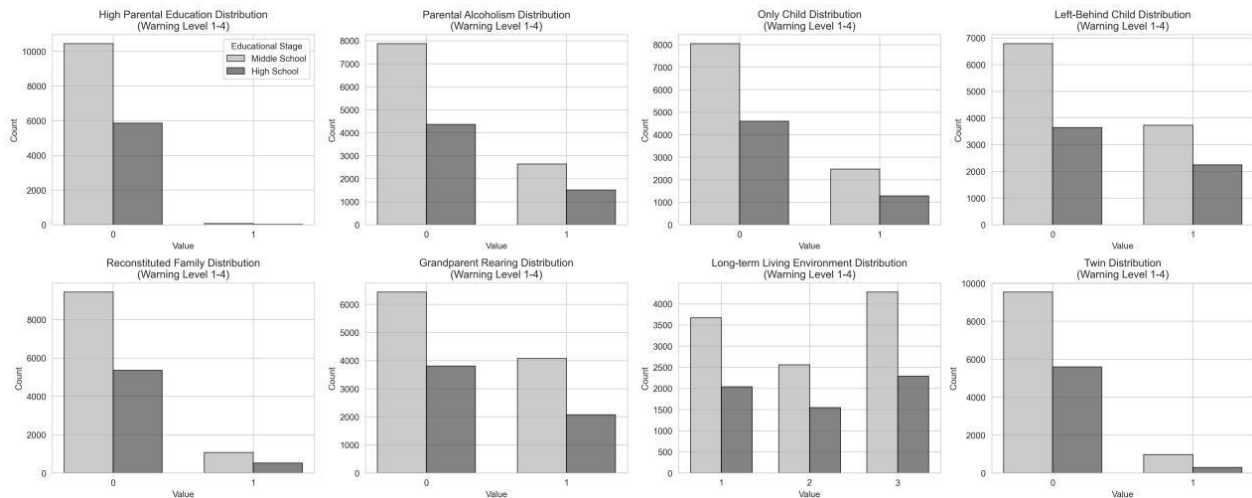


Figure 2: The distribution of family environmental characteristics in middle and high school.

We then show that 8 psychological health indicators affect the warning levels 1-4 across middle- and high-school students. Sleep Quality is skewed toward the healthy end with the healthy ratio 0.93 and 0.94), yet the right tail contains 262 middle-school and 111 high-school cases flagged for severe sleep disorder. Social Avoidance represents a difference in no social behavior (ratio 0.914 and 0.937), which means Social avoidance issues are relatively rare among both middle- and high-school students. Proportion of students who have positive mental health (ratio 0.70 and 0.72), which means under the influence of this factor, the gap between different academic stages is not significant. With respect to Psychotic Risk, the ratio 0.82 and 0.86 indicates that prodromal risk of psychosis has a significant impact on both academic stages. Severe ratios of depression with values 0.087 0.073 and of anxiety with values 0.041 and 0.030 show that the prevalence of depression and anxiety is more pronounced in middle school compared to high school.

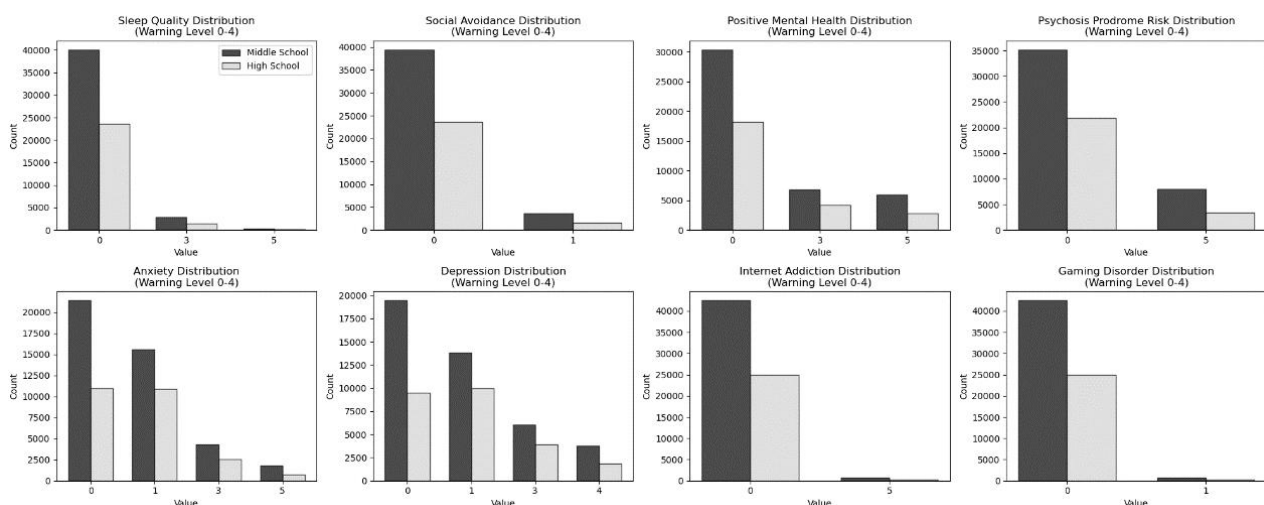


Figure 3: The distribution of psychological health indicators in middle and high school.

Thus, we can see that the significant correlating factors for a high warning level may include parental alcohol abuse, being a single child, being a left-behind child, being raised by grandparents, prolonged residence in rural areas, depression and anxiety.

4. Feature Importance Analysis for Early Warning Level Prediction

Firstly, we employ a Random Forest (RF) regressor to quantify the relative importance of candidate influential factors in classifying early warning levels. As a non-parametric ensemble learning method, RF captures non-linear effects and higher-order interactions, thereby furnishing an independent validation of the regression estimates. For each influential factor, compute the total reduction in Gini impurity achieved at every split that uses it across all trees, then take the average. Its importance is defined:

$$Importance(f) = \frac{1}{T} \sum_{t=1}^T \sum_{n \in nodes_t(f)} \left(\frac{N_n}{N} * \Delta Gini(n) \right) \quad (1)$$

T is the number of trees in RF model. $nodes_t(f)$ is the set of nodes in tree t where feature f is used for splitting. N_n is the number of samples reaching node n . N is the total number of samples. The resultant importance shown in Figure 4 were computed as the mean decrease in node impurity (Gini importance) and normalized to unit sum for interpretability.

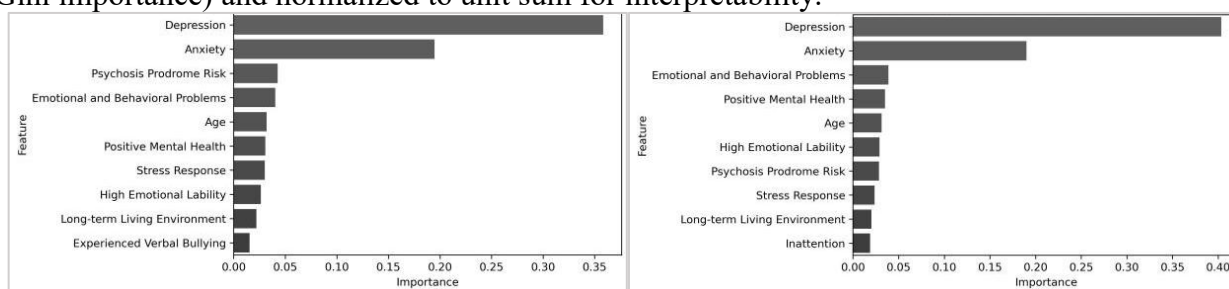


Figure 4: The assessment of feature importance in middle (left panel) and high (right panel) school.

We can see that the significance of the characteristics of middle school and high school is dominated by depression and anxiety. However, for middle school students, the impact of the risk of prodromal symptoms is more severe than that for high school students. The primary reason for this is their physiological development is still weak: middle school students are in a critical period of rapid brain development, particularly in the prefrontal cortex, which is responsible for emotional regulation, judgment, and self-awareness, and has not yet fully matured. This physiological immaturity makes it more challenging for them to understand and cope with prodromal symptoms such as hallucinations, delusions, or thought disorder, leading them to easily fall into fear or confusion.

Summary

The mental health issues of adolescents are increasingly significant, and from a temporal perspective, middle school is the stage that requires the most attention and intervention. From a factor perspective, family factors have a greater impact than individual factors, thus a supportive family plays an indispensable role in the healthy development of students. Furthermore, in analyzing bullying types, verbal abuse is more prevalent and subtle, necessitating careful assessment and prompt handling from all parties involved. Although individual factors are less common, their effects, particularly concerning depression and anxiety, can be substantial and should be taken seriously if they arise.

References

- [1] Zhao N., Yang S., Zhang Q., et al. School bullying results in poor psychological conditions: evidence from a survey of 95,545 subjects. *Frontiers in psychology*, 2024, 15: 1279872.
- [2] Stepanous J, Munford L, Qualter P. Social environment and brain structure in adolescent mental health: A cross-sectional structural equation modelling study using IMAGEN data. *PLoS One*, 2023, 18(1): e0280062.

- [3] Oeri N, Roebbers CM. Adversity in early childhood: Long-term effects on early academic skills. *Child Abuse Negl*, 2022, 125: 105507.
- [4] Lin Y C, Seo D C. Cumulative family risks across income levels predict deterioration of children's general health during childhood and adolescence. *PLoS One*, 2017, 12(5): e0177531.
- [5] Evans G W, Li D P, Whipple S S. Cumulative risk and child development. *Psychol Bull*, 2013, 139(6): 1342-1396.
- [6] Bodovski K. Parental practices and educational achievement: social class, race, and habitus. *Br J Sociol Educ*, 2010, 31(2): 139-156.
- [7] Smokowski P R, Kopasz K H. Bullying in school: An overview of types, effects, family characteristics, and intervention strategies. *Children & Schools*, 2005, 27(2): 101–110.
- [8] Lin D., Zhang L. An Analysis of the phenomenon of campus bullying and the Education of Adolescents' Values and Moral Boundaries. *Educator*, 2024(24): 12-13.