AMERICAN ACADEMIC PUBLISHER
INTERNATIONAL JOURNAL OF MEDICAL SCIENCES
ISSN NUMBER: 2692 - 5206 Volume 6. No 05. MAY ,2026

# DEPRESSION IN THE YOUNG AGE GROUP: PREVALENCE AND ASSOCIATED RISK FACTORS AMONG STUDENTS

**MOHD MUDASSIR**

Medical Student, Samarkand state medical university, Uzbekistan,
mudassirchoudhary116@gmail.com

**MOHAMMAD ADNAN DANISH**

Medical Student, Samarkand state medical university, Uzbekistan

mdadnandanish@gmail.com

**RAKMAT ALI**

Medical Student, Samarkand state medical university, Uzbekistan

alyrahmat885@gmail.com

**Milandeep Singh Kaknian**

Medical Student, Samarkand state medical university, Uzbekistan,
milandeepsinghkaknian@gmail.com

**Abstract**

Depression among students has emerged as a critical public health crisis, characterized by rising prevalence and complex, multifaceted risk factors. This paper investigates the prevalence of youth depression and proposes a comprehensive analytical framework to identify and quantify its associated socio-demographic and behavioral risk factors. By synthesizing advanced machine learning techniques with robust statistical modeling, we aim to overcome the limitations of traditional diagnostic and analytical methods. The proposed methodology outlines a pipeline for leveraging structured survey data to yield actionable insights for early detection. Ultimately, this research provides a scalable foundation for developing targeted, culturally sensitive mental health interventions in educational settings.

## Introduction

The increasing prevalence of depression among the young age group, particularly students, presents a significant challenge to global public health systems. Adolescence and early adulthood represent critical developmental periods where individuals are highly vulnerable to various socio-economic, academic, and biological stressors. Recent epidemiological trends indicate a sharp escalation in mental health disorders among high school and university students, necessitating urgent preventative measures and early detection strategies. The scope of this paper specifically encompasses the computational identification of prevalence rates and the underlying risk factors that contribute to the onset of depression in student populations.

The primary problem lies in the accurate, early identification of students at risk of depression amidst a highly heterogeneous population. Mental health outcomes in youth are dictated by a complex interplay of demographic variables, lifestyle choices, and socio-cultural norms, making isolated symptom-tracking insufficient. Furthermore, the varying manifestations of depression across different subgroups require a nuanced understanding of how risk factors dynamically interact within a pressurized educational environment.

Despite the growing focus on youth mental health, existing analytical approaches remain insufficient for several notable reasons. First, traditional diagnostic frameworks often rely on standard logistic regression models, which can suffer from convergence issues and often fail to directly provide interpretable prevalence ratios in cross-sectional studies [8]. Second, many conventional models adopt a one-size-fits-all methodology, failing to account for critical variations across diverse demographic and racial subgroups, which fundamentally limits their predictive accuracy and cultural sensitivity [5].

To address these critical methodological gaps, this paper proposes an integrated computational and statistical framework. We carefully construct this framework to overcome the weaknesses of isolated clinical analysis and provide actionable epidemiological insights. The specific contributions of this work are as follows:

- We design a hybrid methodological pipeline that combines artificial neural networks (ANNs) for high-dimensional risk factor classification with robust Poisson regression to accurately estimate prevalence ratios.

- We provide a structured evaluation plan that accommodates the distinct socio-demographic variations inherent in diverse student populations, ensuring culturally sensitive predictive modeling.

**Related Work**

The analysis of risk factors and prevalence rates can be categorized into three primary domains of research. The first category focuses on the application of machine learning (ML) and artificial neural networks in public health. Researchers have increasingly utilized ML and ANN models to predict and classify diseases, including youth depression, demonstrating that these computational techniques can yield highly promising diagnostic results when applied to large survey datasets [5]. Similarly, machine learning algorithms like Random Forest and XGBoost have been successfully deployed to predict the underlying risk factors of various non-communicable diseases, highlighting the robustness of these algorithms in handling complex clinical and demographic variables [2]. While these computational methods excel in high-dimensional classification, they often operate as black boxes that lack the interpretable effect size estimations required for epidemiological health policy. Our work bridges this gap by pairing predictive deep learning with highly interpretable statistical models.

The second category involves statistical methodologies designed to accurately estimate risk and prevalence ratios. In cross-sectional epidemiological studies, estimating the direct association of specific environmental exposures with binary outcomes like depression is crucial. The robust Poisson method has gained popularity as an effective alternative to logistic regression because it yields results that are directly interpretable as prevalence ratios while avoiding the non-convergence problems typical of traditional log-binomial models [8]. While highly interpretable and robust, traditional Poisson applications often struggle to automatically capture high-dimensional, non-linear feature interactions without manual feature engineering. This paper addresses this specific weakness by incorporating a neural network pre-processing phase to capture complex non-linearities before applying final statistical inference.

AMERICAN ACADEMIC PUBLISHER
INTERNATIONAL JOURNAL OF MEDICAL SCIENCES
ISSN NUMBER: 2692 - 5206 Volume 6. No 05. MAY ,2026

The third category encompasses socio-demographic and spatial approaches to health risk factor analysis. Integrating medical and anthropological perspectives has proven vital for understanding how socio-economic status, lifestyle, and cultural norms influence health outcomes in diverse urban populations [1]. Furthermore, advanced spatial modeling techniques have been utilized to map the prevalence of various health risk factors at small-area levels, revealing significant geographic and socioeconomic disparities [9]. Although these studies provide invaluable context regarding environmental and genetic influences on general health, they rarely focus exclusively on the specific academic variables impacting student mental health. Our approach builds upon these socio-demographic insights, specifically tailoring the feature selection process to the unique stressors, racial subgroups, and environmental variables encountered by youth in educational settings.

**Method/Approach**

To effectively identify the prevalence and risk factors of depression among students, we propose a multi-stage, hybrid analytical framework. The core rationale behind this design is to leverage the robust predictive power of deep learning for automated feature extraction while maintaining the strict epidemiological interpretability provided by statistical methodologies. By structuring the approach in this manner, public health officials can not only predict which students are at high risk but also quantify the specific, isolated impact of each contributing demographic or behavioral factor.

The proposed analytical pipeline operates through the following structured steps:

1. **Data Preprocessing and Stratification:** We hypothesize the use of a comprehensive, structured dataset akin to the Youth Risk Behavior Surveillance System (YRBSS), containing demographic, behavioral, and psychological data for student populations [5]. The raw data is cleaned, normalized, and stratified by key socio-demographic variables, such as race and gender, to ensure the models capture subgroup-specific behavioral nuances.

2. **Predictive Classification using ANNs:** An Artificial Neural Network (ANN) is deployed to classify the binary outcome of depression (presence versus absence) based on the complex input variables. The ANN architecture includes multiple hidden layers to capture non-linear interactions among risk factors, optimizing for high diagnostic metrics like the F1 score [5].

3. **Prevalence Ratio Estimation:** To quantify the exact risk associated with the most salient features identified by the neural network, we apply the robust Poisson regression method [8]. This step ensures that the final model output provides interpretable prevalence ratios rather than obscure odds ratios, thereby offering direct epidemiological value to school administrators.

The evaluation plan for this methodology relies on rigorous cross-validation and statistical benchmarking. The dataset will be randomly split into an 80% training set and a 20% testing set to evaluate the predictive performance of the ANN using standardized metrics such as accuracy, precision, recall, and the F1 score. The statistical significance of the prevalence ratios extracted from the robust Poisson module will be evaluated using 95% confidence intervals and p-values. Through this dual-evaluation strategy, the framework ensures both high predictive validity and robust epidemiological soundness.

AMERICAN ACADEMIC PUBLISHER

INTERNATIONAL JOURNAL OF MEDICAL SCIENCES

ISSN NUMBER: 2692 - 5206 Volume 6. No 05. MAY ,2026

**Discussion**

The practical implications of deploying this hybrid framework in educational settings are substantial for public health management. By integrating this model into school health and counseling systems, administrators can implement proactive screening protocols that identify at-risk students before acute mental health crises occur. Furthermore, extracting interpretable prevalence ratios allows policymakers to allocate resources more efficiently, targeting the specific lifestyle or demographic risk factors most strongly associated with depression in their unique student population.

Despite the methodological strengths, this framework exhibits several limitations and potential failure modes. First, the reliance on self-reported survey data introduces inherent recall and social desirability biases, which may skew the reported prevalence of depressive symptoms among self-conscious youth. Second, the cross-sectional nature of the proposed survey data provides only a static temporal snapshot, preventing the establishment of true causal relationships between the identified risk factors and the onset of depression. Third, the model faces the risk of unmeasured confounding variables; factors such as family history of psychiatric disorders or specific academic grading pressures might be omitted from standard public health datasets, leading to incomplete risk profiles.

Ethical considerations and risks must also be heavily weighed prior to real-world implementation. One primary ethical risk involves data privacy and informed consent; utilizing highly sensitive mental health and behavioral data requires stringent anonymization protocols to protect vulnerable students from stigmatization. Another significant risk is algorithmic bias; if the training datasets are not adequately diverse, the computational models may mischaracterize or entirely overlook the unique depressive symptoms present in minority student subgroups, thereby exacerbating existing healthcare inequalities.

Future work should focus on expanding the temporal and structural capabilities of the analytical model. One prominent avenue for future research is the incorporation of longitudinal tracking to analyze how depression risk factors dynamically evolve as students transition from high school into rigorous university environments. Additionally, integrating real-time intervention mechanisms, such as mobile health applications that trigger automated support resources when the model detects elevated risk scores, would transition this analytical framework into an active, life-saving clinical tool.

**Conclusion**

Depression in the young age group is a complex, pervasive issue that requires sophisticated analytical methodologies to understand its true prevalence and underlying drivers. This paper introduced a comprehensive framework that addresses the limitations of conventional diagnostic approaches by fusing the predictive capabilities of artificial neural networks with the epidemiological clarity of robust Poisson regression. Through this dual methodology, the intricate, non-linear relationships among various behavioral, demographic, and socio-cultural risk factors can be accurately mapped and mathematically quantified.

Ultimately, recognizing the diverse manifestations of depression across different student populations is essential for effective public health management. By acknowledging the ethical risks and methodological limitations, researchers and educational institutions can refine these computational models to be both equitable and highly accurate.