Diagnosis serves as the foundation for all subsequent clinical decisions and a determining factor of treatment efficacy and patient outcomes. However, current diagnostic approaches rely on the clinician's intuition and personal experience to detect and interpret patient situations, which can lead to biased results or cause the overdiagnosis of certain conditions and the underestimation of others. To address this challenge, we propose to integrate computational algorithms into clinical settings to help improve the accuracy of diagnosis. We evaluated five algorithms, from the basic probability nomogram and logistic regression to more complicated machine learning models like LASSO, Support Vector Machine, and Random Forest. Several factors are considered for evaluating the performances of algorithms in clinical practices: predictive accuracy, generalizability, and level of complexity. We evaluated each model using our academic dataset (N = 550) as the training data and the community dataset (N = 511) as the testing data to mirror real-world research-to-clinic settings. Based on our evaluation results, each algorithm has shown some shrinkage on external validations, especially the machine learning models. Random Forest has demonstrated great potential in identifying the important factors for diagnosing pediatric bipolar disorder and obtained a high internal validation accuracy. However, considering the generalizability and the complexity of combining into clinical settings, we recommend using the probability nomogram to increase the diagnostic accuracy.