

# Enhancing Autism Screening Classification using Feature Selection and Stacking Classifier

Ainie Hayati Noruzman  
Faculty of Computing  
University Malaysia Pahang Al-Sultan  
Abdullah  
Pekan, Malaysia  
ainiehayatinoruzman@gmail.com

Ngahzaifa Abd Ghani  
Faculty of Computing  
University Malaysia Pahang Al-Sultan  
Abdullah  
Pekan, Malaysia  
zaifa@ump.edu.my

Nor Saradatulakmar Zulkifli  
Faculty of Computing  
University Malaysia Pahang Al-Sultan  
Abdullah  
Pekan, Malaysia  
saradatulakmar@ump.edu.my

**Abstract:** The current screening process for early detection of autism spectrum disorders (ASD) is time-consuming and costly, requiring numerous questions about various aspects of a child's development. To address this issue, this study integrates the Recursive Feature Elimination (RFE) method into a stacking ensemble classifier, allowing to identify the most important and effective features from the autism screening tool. This approach is aimed to create a simplified version of the autism screening and to make the screening process faster and more efficient by reducing the number of questions in autism screening tool. The proposed model provides a more efficient and simplified alternative for autism screening, allowing for early decision-making based on consistent and precise results. With 0.9760% accuracy results in predicting ASD traits, it is hoped that these findings will be an alternative option to make the screening questions much simpler while also providing an alternative to parents in predicting autism earlier and with less time.

**Keywords—**Feature Selection, Recursive Feature Elimination, Stacking Ensemble, Autism, Screening Tool.

## I. INTRODUCTION

Autism Spectrum Disorder (ASD) is a common neurodevelopmental disorder in children, with varying prevalence rates. According to recent data from the CDC, approximately 1 in 44 children in the US are diagnosed with ASD. It is more prevalent in males, occurring four times more frequently than in females [1], [2]. Since ASD is a brain defect that results in behaviour and social communication deficits, a comprehensive diagnostic evaluation by clinical experts who understand the ASD symptoms in the screening tools is required [3]. For most current screening procedures, the instruments are often represented by items delivered in a questionnaire, such as the Autism Quotient (AQ) and the Modified Checklist for Autism in Toddlers (M-CHAT) [4]. The screening instruments and simple human expert rules with a scoring function accumulate scores connected with the questionnaire items to calculate the outcome of questions [5]. However, screening tools consist of specific criteria domain such as behaviour, communication, and social emotions, resulting in excessive questions and significantly lengthening the autism screening process [6]. Furthermore, it must also be interpreted by an expert, which can be in ambiguous results due to the varying clinical perspectives of each expert [7]. To address this, utilizing a feature selection method becomes essential [8]. This method automatically extracts relevant features, making the decision-making process easier for experts [9]. However, there has been little headway in investigating existing feature selection in autism screening tools, which contain many features, making it difficult to identify the most relevant ones [10]. By employing the feature

selection method, it might can assist experts in making informed decisions, even in cases where results might be ambiguous due to the differing clinical perspectives of each expert. This way, the feature selection process streamlines and optimizes the screening process, ensuring a more efficient and accurate autism assessment and ASD classification [11]–[13]

To address the above-mentioned issues in ASD classification, this study proposes a model that integrate feature selection method, the Recursive Feature Elimination (RFE) into a Stacking ensemble model in a process called RFE-SE to select the most relevant features in the Quantitative Checklist for Autism in Toddlers (QCHAT10). The QCHAT 10 is a quick reference guide for parents to complete about their toddler (18 – 24 months) with concerns about autism [14]. This model can help in the early detection by identifying the most important features associated with autism in autism screening tool. Additionally, these expected outcomes might can replace existing domain expert guidelines, thereby assisting them in recommending individuals with ASD symptoms for further examination.

This paper is structured as follows: Section II addresses the stacking ensemble with recursive feature elimination (RFE-SE) approach. Section III explains the material and methodology. Section IV then presents results and Section V explains the conclusion and highlights the use and future scope.

## II. RECURSIVE FEATURE ELIMINATION WITH STACKING ENSEMBLE (RFE-SE)

Today, ensemble learning has attracted the attention of researchers to develop methods for classification and prediction tasks [15]–[17]. The main idea of the ensemble learning method relies on multiple classifiers and combines the outcome to generate the final predictive results to improve accuracy and reduce the error rate. Since each type of classifier can produce different results, the ensemble method leverages the strengths of each classifier to achieve a better final ensemble result. One of the components of ensemble learning is Stacking [18].

Stacking or stacked generalization is an ensemble machine learning algorithm [19]. Stacking comprises two-level classification techniques, Level-0 or base-classifier and Level-1, or meta-classifier. Essentially, the Level-0 classifiers are trained using the original dataset and generate probability prediction output datasets for the Level-1 classifier, which acts as a meta-classifier that combines all probability predictions from Level-0 to give the final prediction result. Fig 1. show the stacking ensemble framework [20].