Improving Lepton Identification in CLAS12 Using Machine Learning

Mariana Tenorio Pita
On behalf of the CLAS Collaboration

Old Dominion University

Abstract

In this study, machine learning techniques are applied to improve lepton identification in CLAS12 physics experiments. The primary focus is on minimizing pion contamination, which is observed in both experimental and simulated datasets. Two machine learning models, Boosted Decision Trees (BDT) and Multilayer Perceptron (MLP), were developed, evaluated, and validated. Each model was trained using two sets of variables, and rigorous validation was conducted with simulated and experimental data to ensure reliability. The results demonstrated the effectiveness of these models in mitigating background contamination. By retaining around 90% of leptons, the models achieved a tenfold reduction in pion contamination. This approach has been successfully applied to studies of J/ψ photoproduction near threshold, showing significant improvements in signal detection.