

Postgraduate Seminar Series

Speaker Information

Shawon Kumar Saha (Student no: 0422052082) is a full time M.Sc. student in the department of CSE, BUET. He completed his undergraduate studies at Rajshahi University of Engineering & Technology (RUET) in 2021. His research interest lies in the fields of Multi-objective Optimization and Machine Learning. He is currently doing his postgraduate thesis under the supervision of Dr. Muhammad Ali Nayeem. He will be speaking about his ongoing research in this talk.



HANDLING IMBALANCED DATASET USING DEEP LEARNING AND EVOLUTIONARY ALGORITHMS

Classification, a supervised machine learning approach, employs models to predict accurate labels or classes for given datasets. Adequate data availability is essential for effective classification tasks. However, a prevalent challenge in data collection for such tasks is the imbalance in data distribution, where certain classes are significantly underrepresented compared to others. Imbalanced datasets, characterized by skewed class proportions, are frequently encountered in real-world classification scenarios, including but not limited to fraud detection, intrusion detection, and text classification. Several classes of techniques, including undersampling and oversampling, have been proposed to mitigate the challenges posed by imbalanced datasets. Oversampling, a widely adopted approach, aims to rectify imbalanced class distributions by augmenting the instances of the minority class within a dataset. Conversely, undersampling methods seek to address class imbalances by reducing the number of instances belonging to the majority class. Applying undersampling techniques to address data imbalance may entail a loss of information and consequently diminish model performance due to the reduction in data volume. Conversely, oversampling strategies can induce overfitting, wherein the model captures noise from the oversampled data rather than discerning the underlying patterns, potentially leading to the inclusion of duplicate instances. Recently, there has been a notable trend towards leveraging the combined strengths of deep learning and evolutionary algorithms to address this challenge effectively. Nevertheless, the exploration of this approach has not received adequate attention thus far. To tackle this problem, we introduce a novel way of integrating evolutionary algorithms with deep learning to address data imbalance issues effectively for binary classification. This integration will be able to solve the problem created due to imbalanced data.