

**Invitation to M.Tech. Thesis Defense of Tejas: August 18, 2023 (Friday): 19:00 PM - 20.00 PM IST**

In Partial Fulfilment of the Requirements for the Degree of

**M.Tech. CB**

**Tejas (MT21232)**

Will defend his thesis

Title: “Machine Learning And Deep Learning Models For Solvation Energy Prediction”

IIIT-D Faculty and Students are invited

**Date: August 18, 2023 (Friday)
Time:** **19.00 -20.00 IST
Online over Google meet (https://meet.google.com/zfo-ohht-krb)**

**Examiner: Internal:   Dr.Tarini** **Shankar Ghosh**

**~~External~~/Internal: Dr. Jaspreet Kaur Dhanjal**

**Advisor: Dr. N. Arul Murugan**

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**Abstract**

Drug discovery is divided into 4 phases. The first phase is in-silico processes beginning with target identification and validation, followed by the hit discovery process, assay development, high throughput screening, preparing lead, and lead optimization. Screening through all the hits is experimentally unfeasible in terms of both time and resources. Henceforth in-silico models have been developed to predict the properties such as protein-ligand binding affinity, ligand permeability and so on. Here in this thesis, the property in focus is Solvation Energy which provides us with ligand dissolution energy during the protein-ligand binding process in an aqueous medium. The classical mechanics and quantum mechanics-based deterministic approaches can be employed to predict the solvation energies, but these approaches are computationally very demanding. Machine learning and deep learning methods can be used, which can provide reliable results and can be computationally less demanding. Here ten Graph-based deep learning models have been trained using graph representations in combination with two featurizers which help in featurizing the input molecules. Various unsupervised mechanisms like convolution, attention mechanisms, and supervised mechanism interaction networks are implemented for solvation energy prediction. These algorithms work upon graph representations constructed from input molecules by mapping atoms to vertices and bonds to edges. Apart from these three, machine learning models are also trained on different types of descriptors. Weave and CIGIN models perform best in graph-based deep learning algorithms trained on the FreeSolv dataset. Among different machine learning models, the random forest model is found to work best on both datasets of FreeSolv and MNSol. In this thesis, we establish that reliable machine learning and deep learning models can be developed for predicting the solvation energies.