

**Invitation to M.Tech. Thesis Defense of Trinita Roy: May 30, 2022 (Monday): 02:00 PM-02.45 PM IST**

In Partial Fulfilment of the Requirements for the Degree of

**M.Tech. CB**

**Trinita Roy (MT20221)**

Will defend her thesis

**Title: “Protein-based in silico models to predict infectious strains of Influenza A”**

IIIT-D Faculty and Students are invited

**Date: May 30, 2022 (Monday)
Time:** **02.00 PM-02.45 PM IST**

**Venue: Room No: A-320 (Meeting Room), R & D Building**

**Examiner: Internal:   Jaspreet Kaur Dhanjal**

**External/~~Internal~~: Manish Kumar (University of Delhi)**

**Advisor: G P S Raghava**

**Co-Advisor NA**

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

**Abstract**

Influenza A, an infectious viral disease affecting the lungs, is a significant public health concern. It has already caused four pandemics in the past, and some strains are now seasonal. Therefore, recognizing zoonotic strains that can cause an outbreak in humans and their origin is the need of the hour. In this analysis, we have devised a machine learning method to predict infectious strains of the Influenza A virus from avians/mammals to humans. The training and validation of the 15 protein sequence was conducted on data obtained from the Influenza Research Database. Random forest-based models using composition-based features attained maximum AUC for the 15 proteins ranging from 0.93 to 0.98 on the validation dataset. On training and validation datasets, the haemagglutinin (HA) protein has the highest AUC of 0.98. We have formulated an in-silico tool for the prediction of infectious strains from protein sequences as a service to the scientific community. The best models were incorporated in our web server named FluSPred which can be accessed freely at “https://webs.iiitd.edu.in/raghava/fluspred/”. We expect that this research will assist in prioritizing high-risk viral strains hereafter and analyze the risk of a novel influenza virus emergence. This tool can be integrated with early warning systems and is beneficial for pandemic preparedness, disease surveillance, and determining the overall public-health impact.