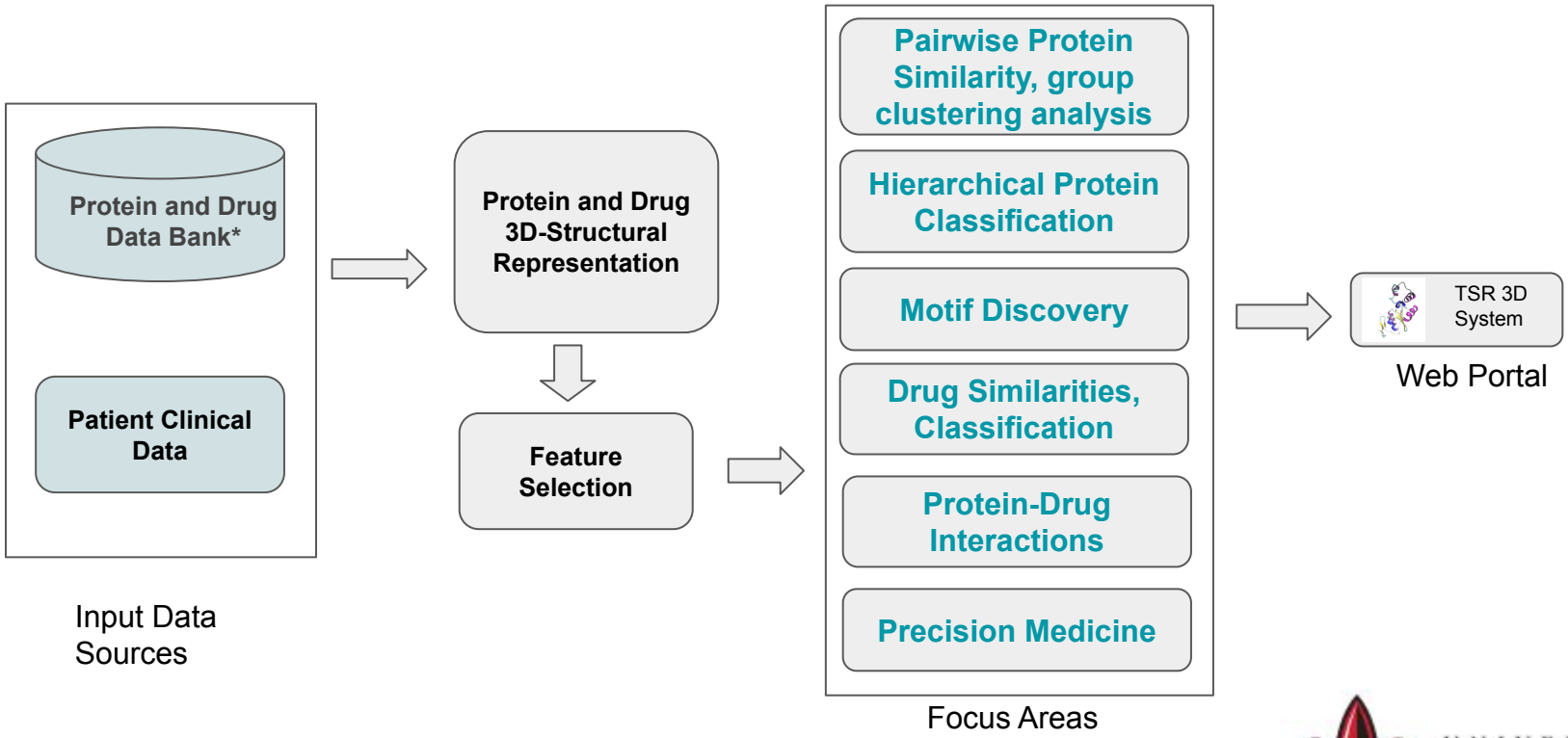
# Hierarchical Classification using deep transfer learning for Protein Functions based on TSR-3D representation

Sarika Kondra

# Good to have

- Solved and understood Machine Learning classification problems
- Python coding and debugging skills
- Deep Learning knowledge( For Proposal 2)

# Project Overview

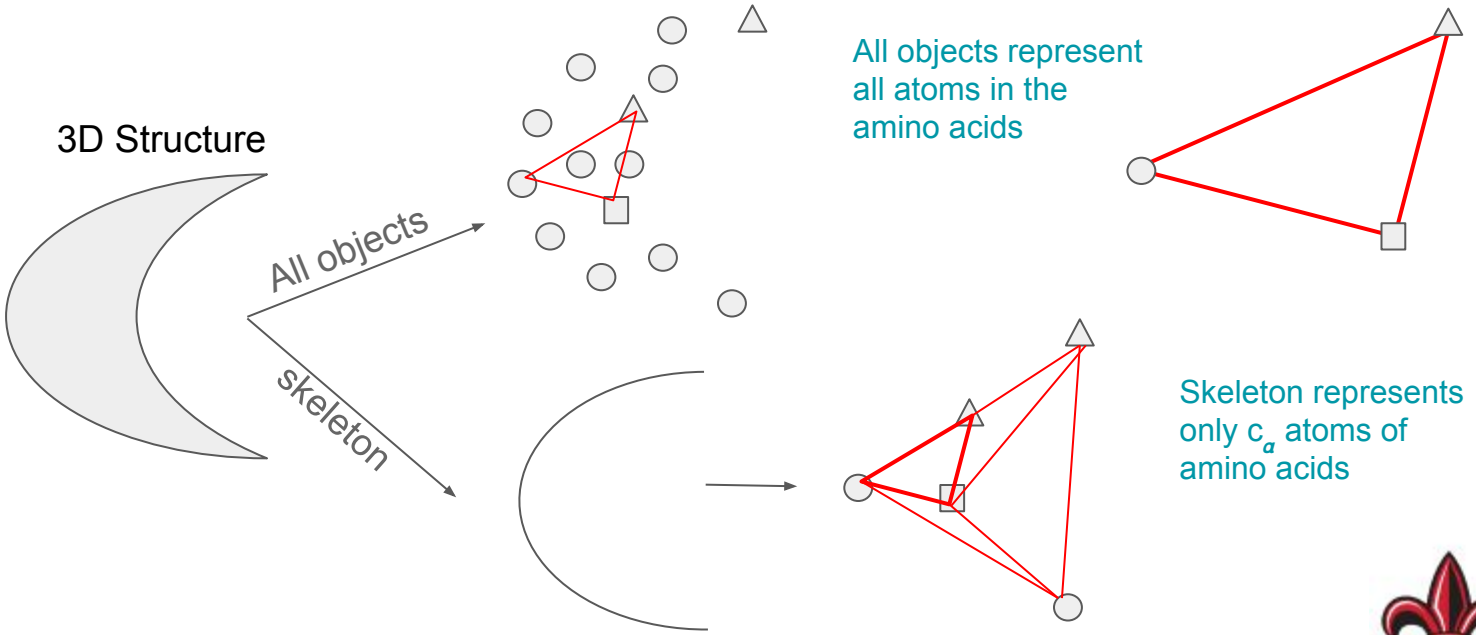



*https://www.rcsb.org/,KEGG BRITE, BRENDA, SUPERTARGET, DrugBank

UNIVERSITY of LOUISIANA LAFAYETTE

# Our Method: Protein and Drug 3D Representation

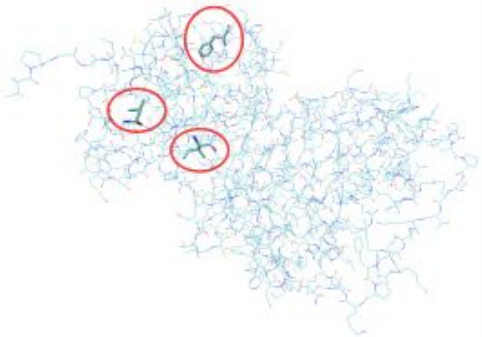
Triangular Spatial Relationship (TSR)



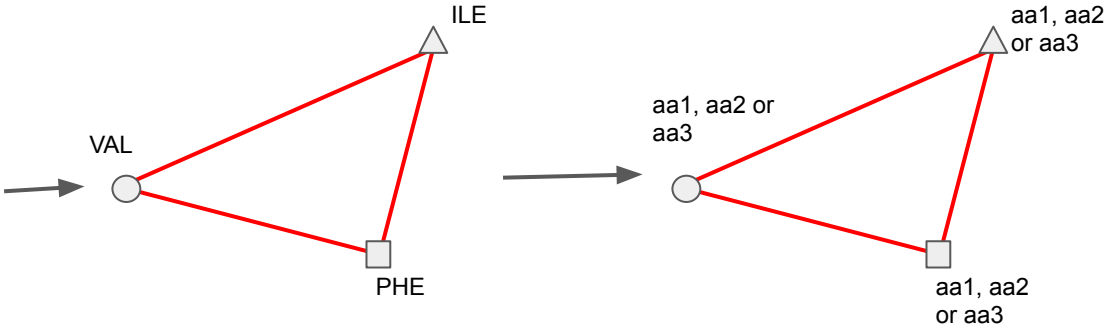

UNIVERSITY of LOUISIANA LAFAYETTE

# Our Method: Protein and Drug 3D Representation

## Motivation for Rule based-label arrangement

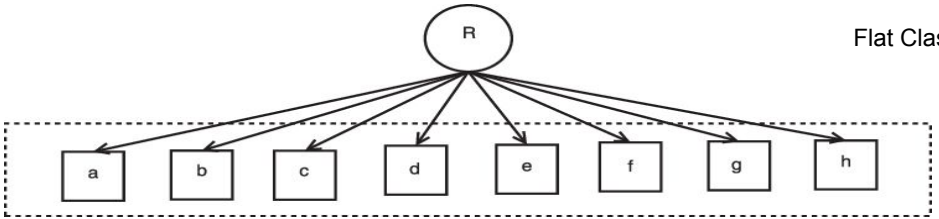


Protein: 2ZOQ
With circled amino acids considered as 3 vertices

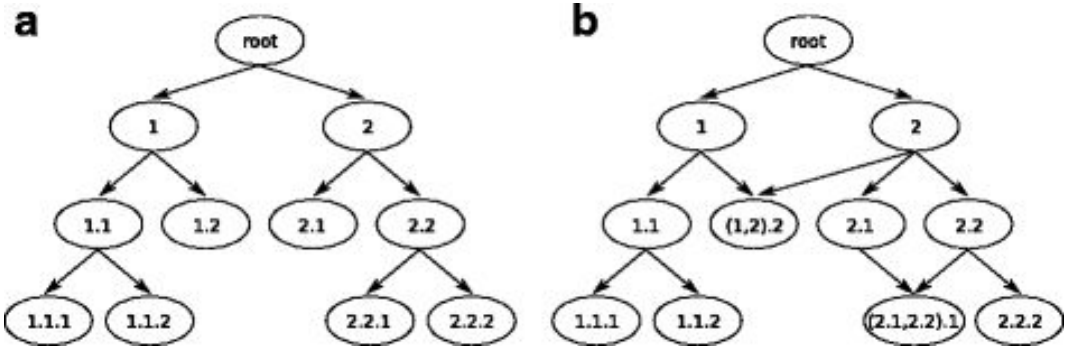



UNIVERSITY of LOUISIANA LAFAYETTE

# Classification: Flat Vs Hierarchical



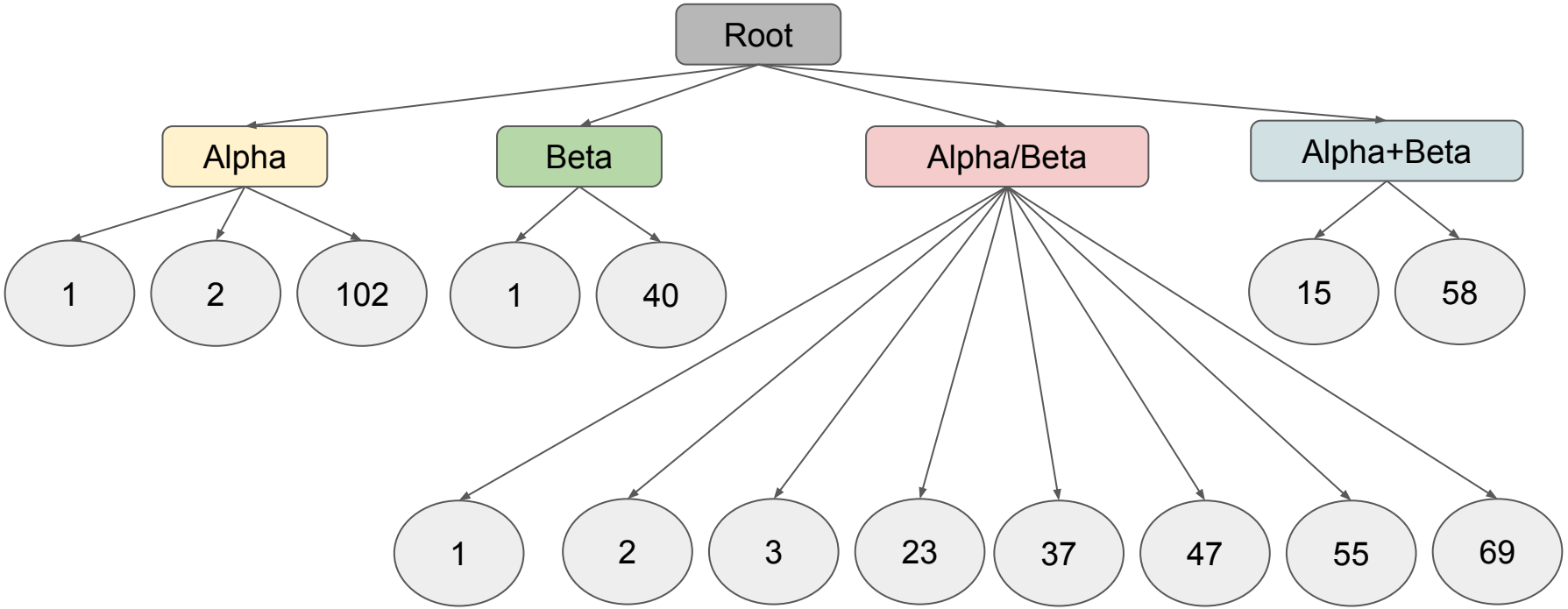

Flat Classification




Tree class structure

Directed Acyclic Graph class structure

# Datasets Hierarchy



# Method

- Start at ROOT node
- Perform Resampling to balance the dataset at the current node
- Perform Feature Selection
- Select best classifier at every node using training and validation sets
- Perform Classification
- Recursively repeat the steps for all parent nodes

Some Classifiers used:

Decision Tree, Logistic Regression, Nearest Neighbors, Linear SVM, Random Forest, Neural Net

# Proposal 1

- Try 3 different datasets
- Analyse different set of features for classification: Intra, Inter and all features
- Perform Flat and Hierarchical Classification and report results
- Analyse the results
- Work on LONI supercomputer
- Code is ready
- Be a co-author on my paper.

# Proposal 2

## Transfer Learning

Transfer learning is a machine learning technique where a model trained on one task is re-purposed on a second related task.

### Pre-trained Model Approach

1. **Select Source Model**. A pre-trained source model is built on level 1 with all available data.
2. **Reuse Model**. The pre-trained model can then be used as the starting point for a model on the second task of interest i.e level 2. This may involve using all or parts of the model, depending on the modeling technique used.
3. **Tune Model**. Optionally, the model may need to be adapted or refined on the input-output pair data available for the task of interest.

# Proposal 2

- Build Deep Transfer Learning models for the hierarchical Classification of proteins.
- Analyse different set of features for classification: Intra, Inter and all features
- Compare results with Flat Classification and report results
- Work on LONI supercomputer
- Basic code is available.
- Be a co-author on my paper.

# Thank You

Questions??