



A MACHINE LEARNING APPROACH TO PROTEIN FOLD RECOGNITION

DON BOSCO INSTITUTE OF TECHNOLOGY , KURLA WEST 400 070

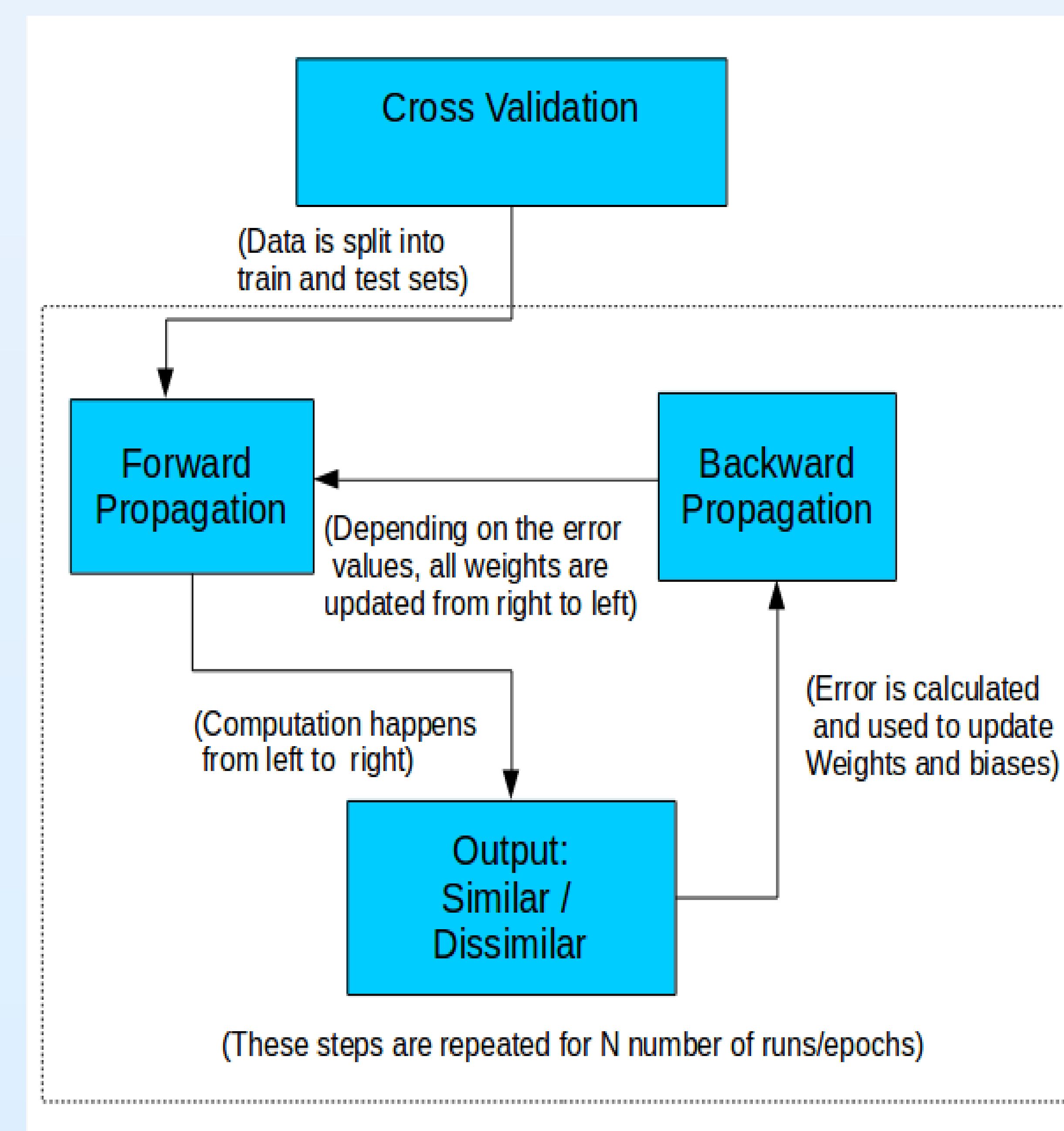
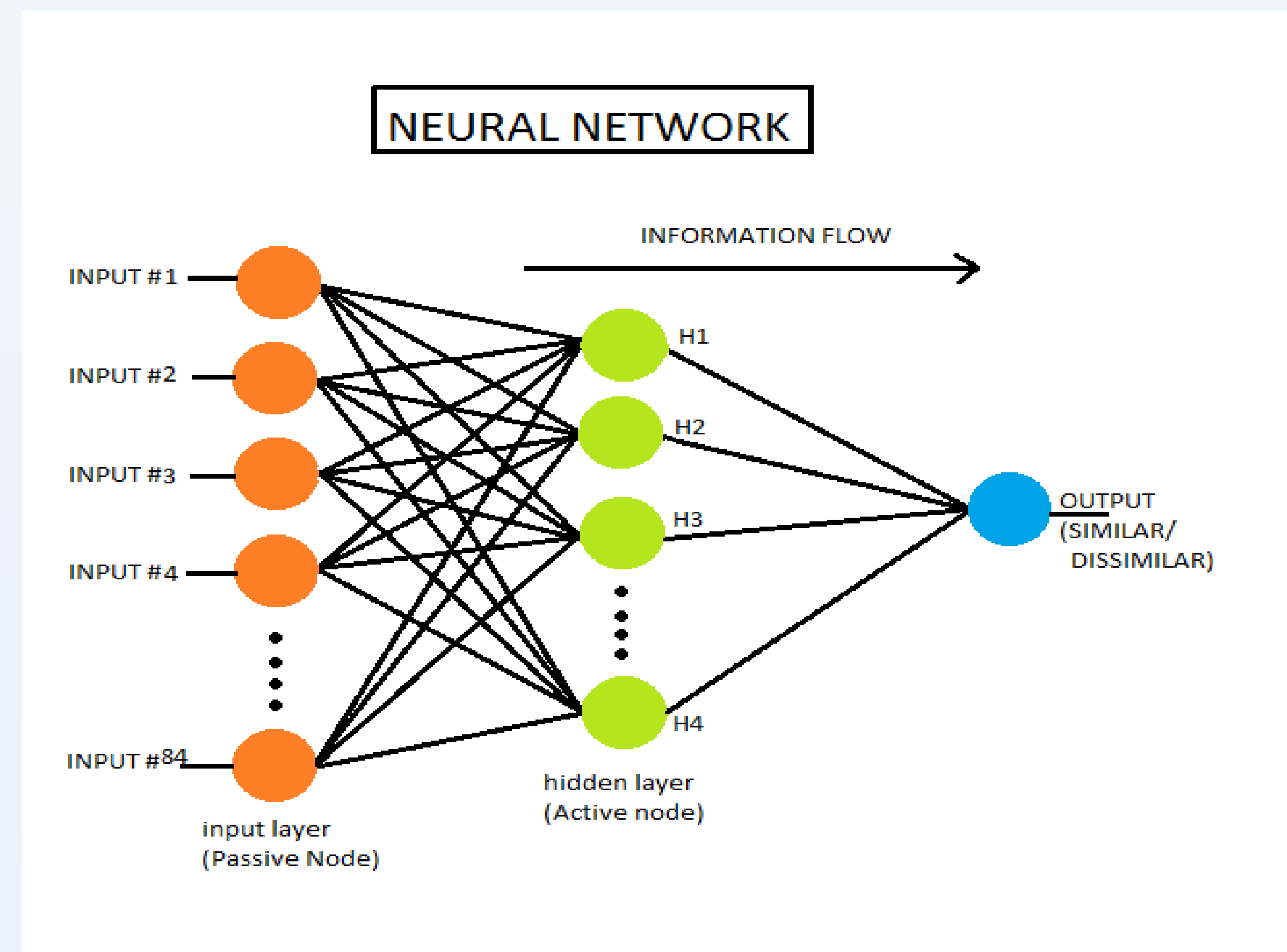
ABSTRACT :

- Protein folding is the physical process by which a protein chain acquires its native 3-dimensional structure, a conformation that is usually biologically functional, in an expeditious and reproducible manner.
- An important task in protein structure prediction is to identify proteins that have similar tertiary structures (from among those that have already been determined experimentally).
- By identifying such proteins, their structures can be used as a template to model the unknown structure of another protein.
- For accurate recognition of protein folds, a neural network is developed to predict whether a given protein pair belongs to the same structural fold.*

KEY POINTS :

- For each pair of proteins in the dataset, 84 pairwise similarity features are used to characterize the protein pair.
- These similarity features used were obtained through five types of sequence alignment and/or protein structure prediction tools (i.e., sequence-sequence alignment, sequence-family information, sequence-profile alignment, profile-profile alignment and structural information) and selected based on their use in prior works.
- The dataset consists of proteins from the SCOP dataset (version 1.37) such that the sequence identity between any pair of proteins was <40%.
- Two proteins pertaining to the same structural fold is a problem that can be readily addressed as a binary classification problem. In this context, a protein pair which shares the same structural fold is labelled as 1 and pairs from dissimilar folds are labelled as 0.

MECHANISM :



POSSIBLE FUTURE DIRECTIONS :

- Creating a dynamic system which takes input in the FASTA format and generates similarity scores by accessing tools(data pre-processing).

CONCLUSION :

- Neural networks give better accuracy as compared to other machine learning methods like SVM, regression etc.

REFERENCE :

- <http://www.nature.com/articles/srep17573#abstract>
- <http://iris.rnet.missouri.edu/dnfold/>

PROJECT GUIDE :

UDAY NAYAK

PROJECT TEAM :

KARAN MANGHI
KAVYA KOTIAN
CHIRAAG LIMAYE
ASHWIN FERNANDES