# MULTIMODAL FAKE NEWS DETECTION

ABSTRACT

- News articles and images together influence the judgement of the reader as to whether a particular article is fake.

- Current multimodal frameworks like EANN(event Adversarial Neural Network) and SpotFake rely on two separate deep learning models for text and image like BERT or Word2Vec(for text) and VGG-19(for image),concatenate features using various deep learning methods like adversarial networks,attention models etc and try to classify.

- In this model,we have come up with a different way to extract features that could optimize our model and that can be used simultaneously to train our model.

Flow-Chart
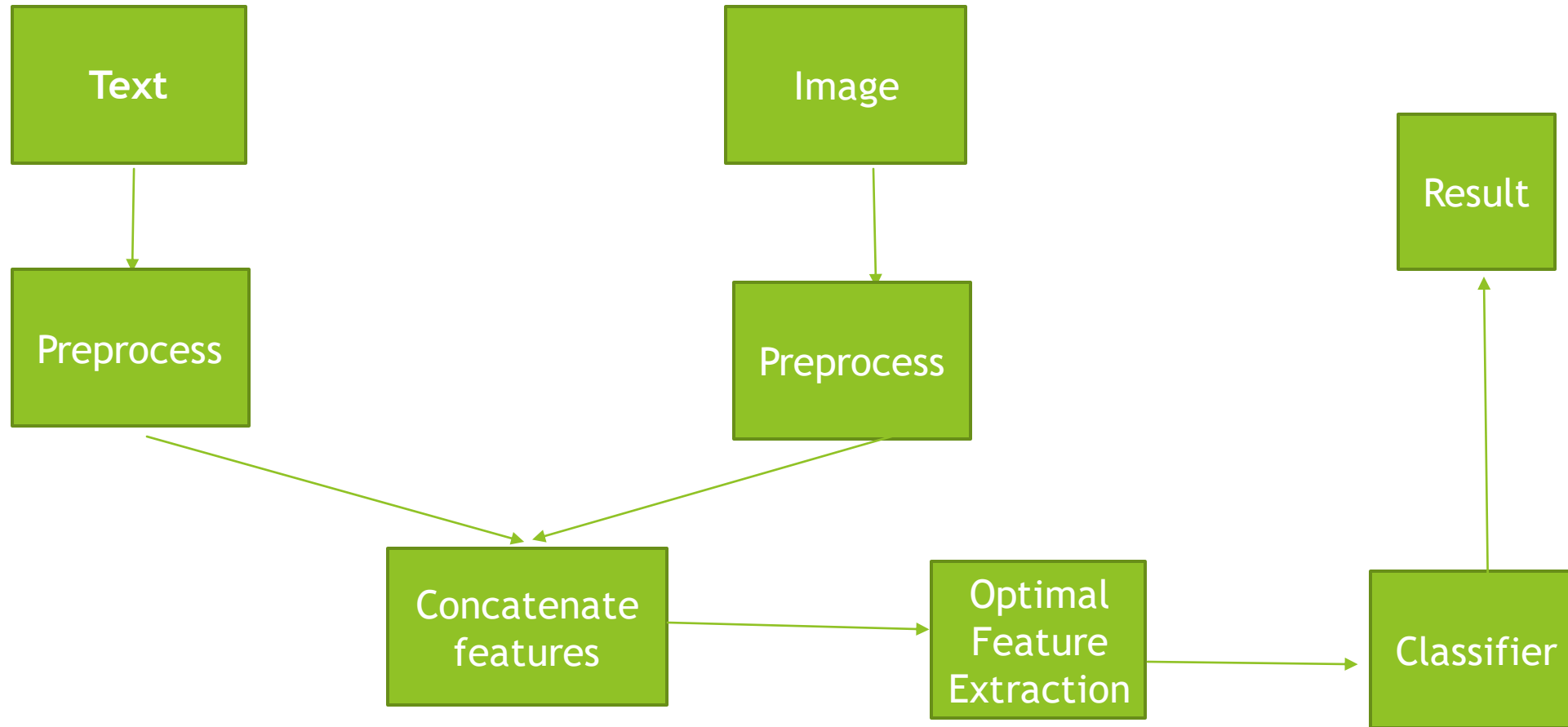
Text Preprocessing

- Most news articles are used directly,however not all words are relevant,especially stopwords(like "is","was" etc).The first step is removal of all

  Stopwords and punctuations since they do not add useful features to the text.

- Next,we can perform stemming to remove verb suffixes(Eg:- "ing","ed" etc)

- Finally we can use word embeddings or bert embeddings to represent our news articles.

Image Preprocessing

▶ Since extracting useful features from a complete RGB image,is a challenging task we propose the folloiwng preprocessing steps

1. Resize the image (eg:-100x100) and convert it to grayscale or 1-channel image.

2. We can then try to use Otsu's binarization method that finds the optimal threshold to separate the background and foreground.

3. We then suggest using image segmentation that assigns labels to image pixels and similar values help us to group similar pixels.The main idea is to change image representation into something easy to analyze.

4. We can use K-means to identify segments based on colour,shape and texture of image.

5. Other possible methods like DWT can give us features relating to edges and resolution.

Optimum Feature selection

▶ Once we have got both text and image features,we will concatenate both features.

▶ However since we have used a collection of different features from different methods,not all of them might be useful and consume valuable training time.

▶ So we need a model to extract useful features subset with minimum cost.There are two techniques:-filters and wrappers

▶ Filters:- uses some techniques such as mutual information to measure the information content of selected subset ignoring classifier algorithm.

▶ Wrappers:-the selected subset evaluates by a classifier algorithm.

▶ We propose a wrapper technique that could be effective(CA) along with SVM objective function.

CA algorithm

▶ Represent each feature subset as a vector of length n where n is no of features.The feature to be chosen has value 1 and rest 0

▶ Fitnesse score of a vector is calculated using the objective SVM function

▶ Initially choose a random set of vectors and calculate their scores.

▶ Updating the set can be done by two methods:- crossover and mutation

1. In crossover choose two vectors,such that probability of choosing one is directly proportional to their fitness score and interchange the substrings at crossover point.

2. In mutation,choose a vector similarly as previous and invert a particular position or set of positons.

▶ Repeat 1) and 2) until you get atleast n offsprings where n is population size.

▶ Continue for upto 50 generations.

CLASSIFICATION

▶ The optimal feature vectors obtained from CA are then used for classification of dataset.

▶ The calssifier that we have proposed is kernel SVM,like gaussian RBF or Laplacian