

## RESEARCH OVERVIEW

Amlan Kar

---

**Homepage:** [home.iitk.ac.in/~amlan](http://home.iitk.ac.in/~amlan)

**Email:** [amlan@iitk.ac.in](mailto:amlan@iitk.ac.in)

**Institution:** Indian Institute of Technology, Kanpur

## ABOUT

---

# WHO AM I ?

- Junior Undergraduate Student in the Department of Computer Science and Engineering at IIT Kanpur
- **Research Interests:**
  - Computer Vision
  - Machine Learning
  - Multi-modal models

# WHY ML AND COMPUTER VISION?

I strongly believe that this is the path that will lead us to a *deep understanding of the perceptive basis of the Human Mind*. I also have a lot of interest in working to build *assistive technologies* to empower people with limited perceptive abilities.

To put it very simply, these fields *excite me beyond measure*.

## PROJECTS

---

# Affective Computing Using a Triad Based Method

## **Problem Statement:**

To classify the sentiment in a given video segment using the visual, auditory and spoken text cues into neutral and the 6 basic Ekman emotions (i.e. Happiness, Sadness, Surprise, Fear, Disgust and Anger)

<http://home.iitk.ac.in/~amlan/resources/triad.pdf>

# AFFECTIVE COMPUTING USING A TRIAD BASED METHOD

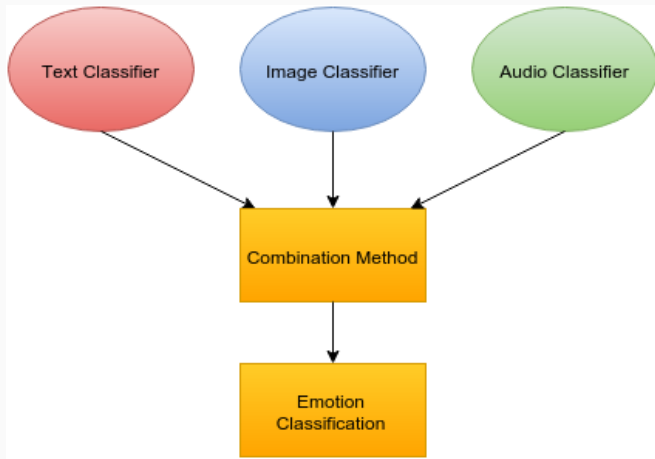


Figure: What we did



- Studied and implemented the Viola-Jones object detection framework to detect a face, eye, nose and mouth(fixed sized boxes) in an image using numerous Haar-like features.
- Took the concatenation of these regions (after dimensionality reduction using PCA)(very, very naively) as the feature to train a multi-class SVM, which achieved only 72% accuracy.

# AFFECTIVE COMPUTING USING A TRIAD BASED METHOD

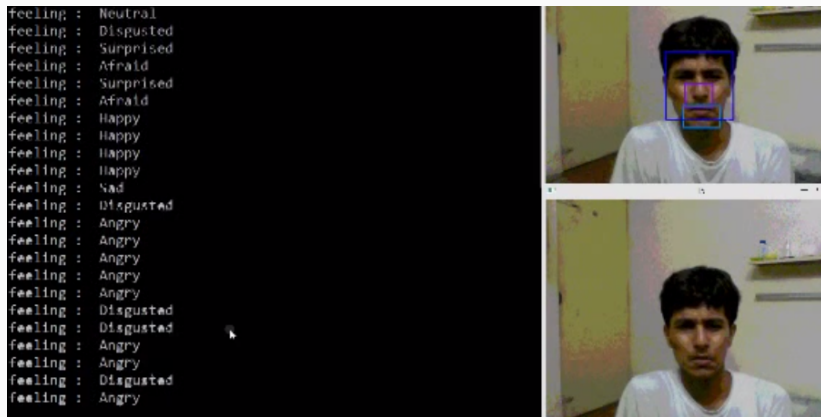


Figure: The image classifier at work

- Learnt various NLP techniques and we basically tried them all while trying to do better
- The best results were achieved by a simple Naive Bayes classifier from a Bag of Words after using the NLTK POS tagger for marking the parts of speech and using the word, POS tuple as a pseudo word in the BoW approach.
- This method achieved around 79% accuracy on the binary classification task (positive/negative).

- We built the Audio vector by clustering small windows of the data represented by it's Mel Frequency Cepstrum Coefficient vector and assigning the cluster number to each window in the output vector.
- This vector was used to train a K-Nearest Neighbours classifier and a multi-class SVM which could achieve 64% accuracy. Interestingly, the authors of the database claim that they experimentally found 67% accuracy by humans for the same task.

# Music Classification using Deep Neural Networks

# MUSIC CLASSIFICATION USING DEEP NEURAL NETWORKS

## **Problem Statement:**

To classify music segments into genre and artists. We used 10 genre from the GTZAN dataset and 10 Blues artists on a self created Dataset.

<http://home.iitk.ac.in/~amlan/cs365/project/>

# MUSIC CLASSIFICATION USING DEEP NEURAL NETWORKS

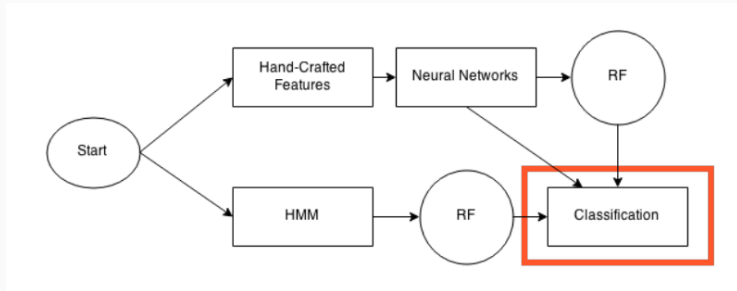
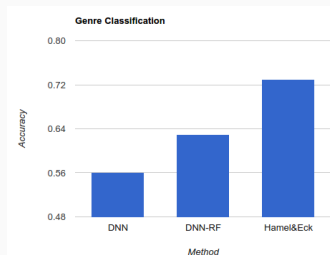


Figure: Methodology

- Used Mel Frequency Cepstrum Coefficients (MFCC) and FFT (Fast Fourier Transform) features of the input data.
- Trained a NN architecture on the data.
- Used the neural net embedding of the data to train a Random Forest classifier, which consistently gave nearly 10% increase in classification accuracy.
- Experimented with modeling the data using a Hidden Markov model, with discouraging results.



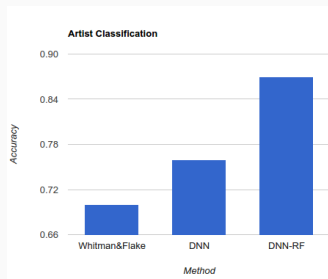
# RESULTS



**Figure:** Results on Genre Classification Task

- Suffering from Ill ground truth ? - Need much more training ? (104 hours(Hamel&Eck) vs 9 hours)

# RESULTS



**Figure:** Results on Artist Classification Task

- Overfitting ? Confusion matrices show high confusion between Clapton and Cream!
- Possibility of "The Producer Effect" in self-created dataset for some artists

## CURRENT PROJECTS

---

# Generating Unique Inversions from a Deep Model of Motion

## **Problem Statement:**

Generate reconstructions of a robot given its joint angles by using only visual data.

## **Why?**

To try a new way of robot path planning which requires minimal prior knowledge of the robot. We aim to use the reconstructions to perform the local planner task in the Probabilistic road map algorithm.

We took inspiration from Dosvitskiy's convolutional chair generation model to perform our reconstruction task by learning a deep embedding of the joint angle representation that can be used to generate the robot image.

We have been playing with their Caffe code and have modified it to our requirements.

We have shifted our focus for now to create an example dataset of robot arms(no such dataset exists to the best of our knowledge, as of now) for our model.

- Completed the background subtraction part to obtain segmentation masks for the robot.
- Completed the key-point mapping between two robot images using the OpenCV implementation of ORB features.
- Working on the local planner step that decides object obstruction given a single RGB-D image of the object. We plan to extend it to the case where a single RGB image is given using depth estimators such as the one shown by Saxena et al.

# Deep Learning for Document Classification using Skip-Thoughts

- Submitted project proposal for CS671(Natural Language Processing) on October 4, 2015.
- I intend to model documents using the Skip-Thought model for sentences and train a Convolutional Neural Network to take this matrix and perform standard Document Classification tasks.
- <http://home.iitk.ac.in/~amlan/cs671/project/proposal.pdf>



# THINGS I DO FOR FUN

- I love implementing Machine Learning algorithms to understand them better. My first Neural Network is probably the most memorable one yet.
- The IITK Family Tree  
([http://iitk.ac.in/counsel/family\\_tree/](http://iitk.ac.in/counsel/family_tree/))
- Making simple and elegant websites for people.
- Learning cool Linux tricks.
- Playing the guitar and going on random photo walks.
- Helping people out as a core team member at the Institute Counseling Service.

DONE!