INSTITUTE OF COMPUTER SCIENCE
UNIVERSITY OF THE PHILIPPINES LOS BAÑOS

Identification of Fake GAN-Generated Images through Frequency Analysis and Machine Learning

CMSC 190 SPECIAL PROBLEM

Yanna Denise A. Hilario
BS Computer Science

Rodolfo Camaclang III


These people do not exist...

Images taken frome 140k fake and real images

Deception through AI

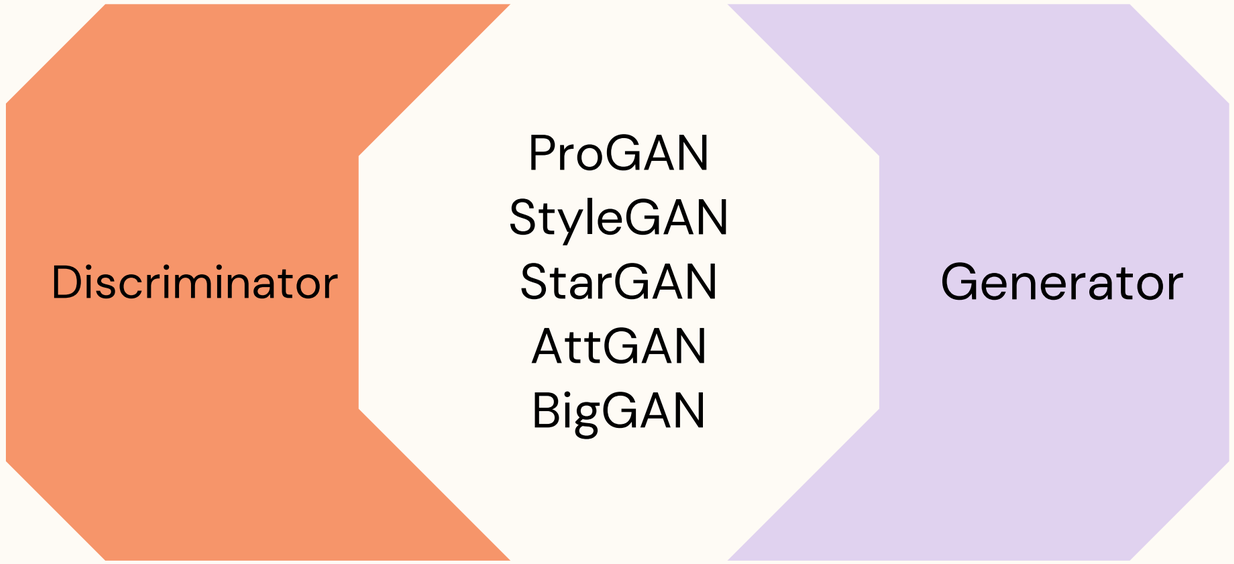
- Synthetic media generation through
  - Deepfake
  - Generative AI
- Fabrication of fake images, videos, and audios
- Deception of audiences
- Identity theft and defamation

Images taken frome 140k fake and real images

Anchors, reporters nagamit sa 'deepfake' ads

January 26, 2024 news report from TV Patrol

Generative Adversarial Networks



- Adversarial Training
- Hyper-realistic images
- Computationally expensive to train

---

*As GANs continuously improve, the race for **more advance detectors** to discriminate fake from real images also calls for **enhanced strategies**, **robust methods**, and **innovative approaches** to stay ahead in the ongoing battle between generative models and detection mechanisms.*

GAN detection

Deep Learning and the use of CNNs

- High computational resources
- Complex Architectures

Concentrated on Fake Faces

- Diverse Fake Face Dataset
- FaceForensics++
- 140k Real and Fake Faces
- CelebA

Significance of the Study

Frequency Analysis

Discrete Cosine Transform

Discrete Wavelet Transform

+

Machine Learning

Support Vector Machine

Random Forest

GAN traces in different object classes

- Animals
- Vehicles
- General objects

Research Objectives

- Collect real and synthetic images from the ProGAN dataset.
- Apply DCT and DWT as pre-processing methods on the input images.
- Determine which spectral features can be used to classify fake from real images.
- Implement two classification models for this problem namely: (1)Support Vector Machine (SVM) and (2)Random Forest.
- Evaluate the accuracy of the classification models using the proposed methods.

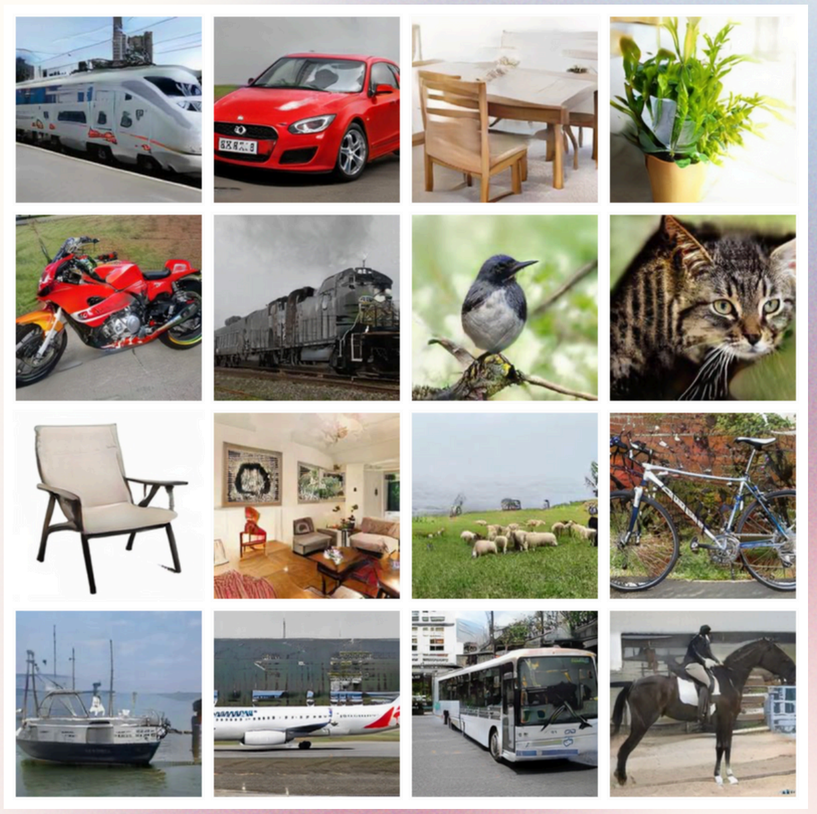
Methodology

Collect images from the **ProGAN** dataset with varying **subclasses**

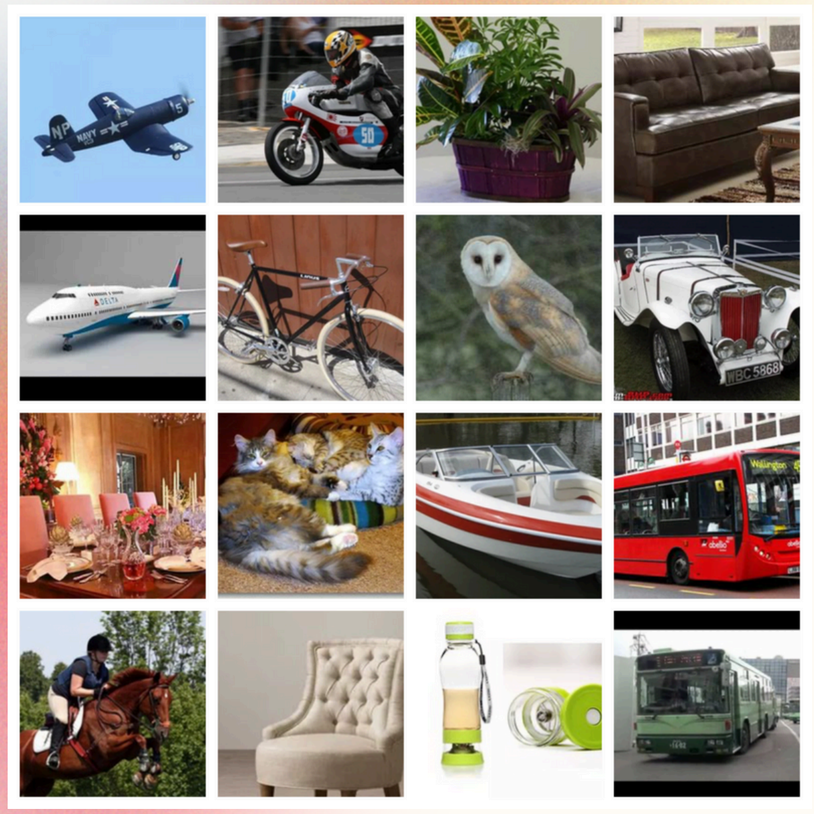
Apply **DCT** and **DWT** on images and determine which **spectral features** can be used to classify fake images

Classify the images using two classification models: **SVM** and **Random Forest**

Evaluate the **accuracy** of each classifcation model and the overall accuracy of the proposed solution

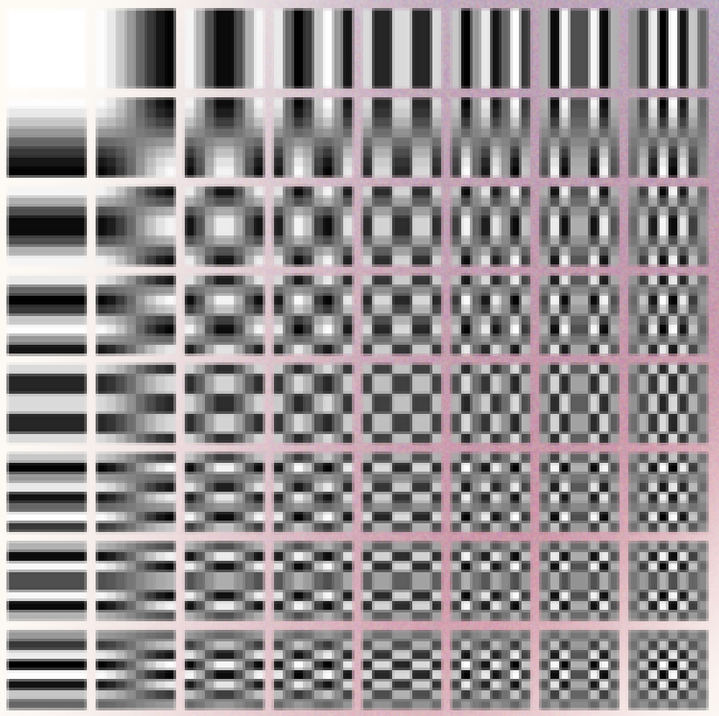


Fake Images from the ProGAN dataset



Real Images from the ProGAN dataset

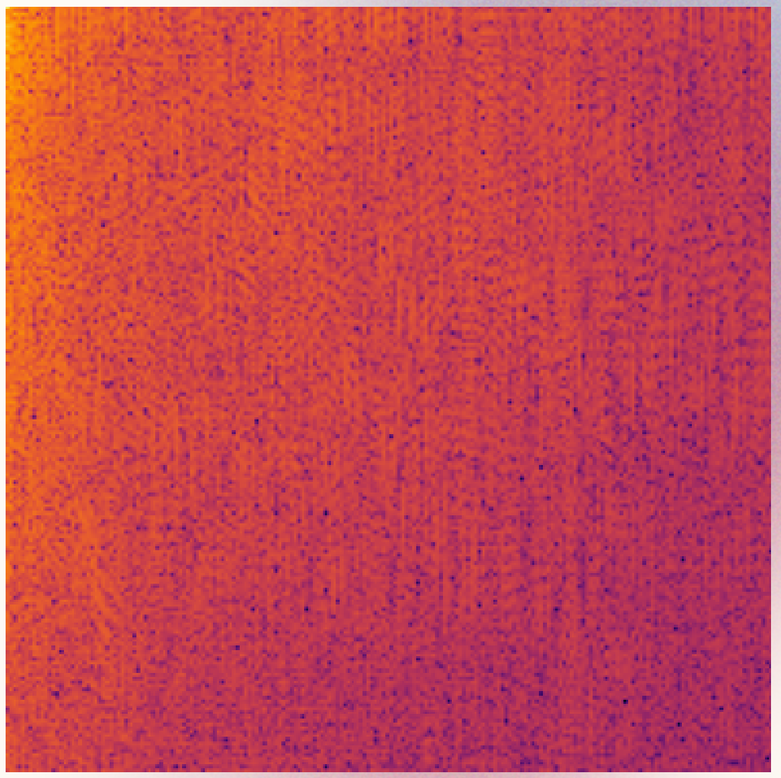
Discrete Cosine Transform



8 x 8 DCT Block

- Based on Fourier Transform
- Represent images as a sum of cosine waves
- Used in image compression algorithms
- Used in Frank et al. and Guidice et al.’s studies

Using SciPy for DCT



Log plot of DCT coeffiecients

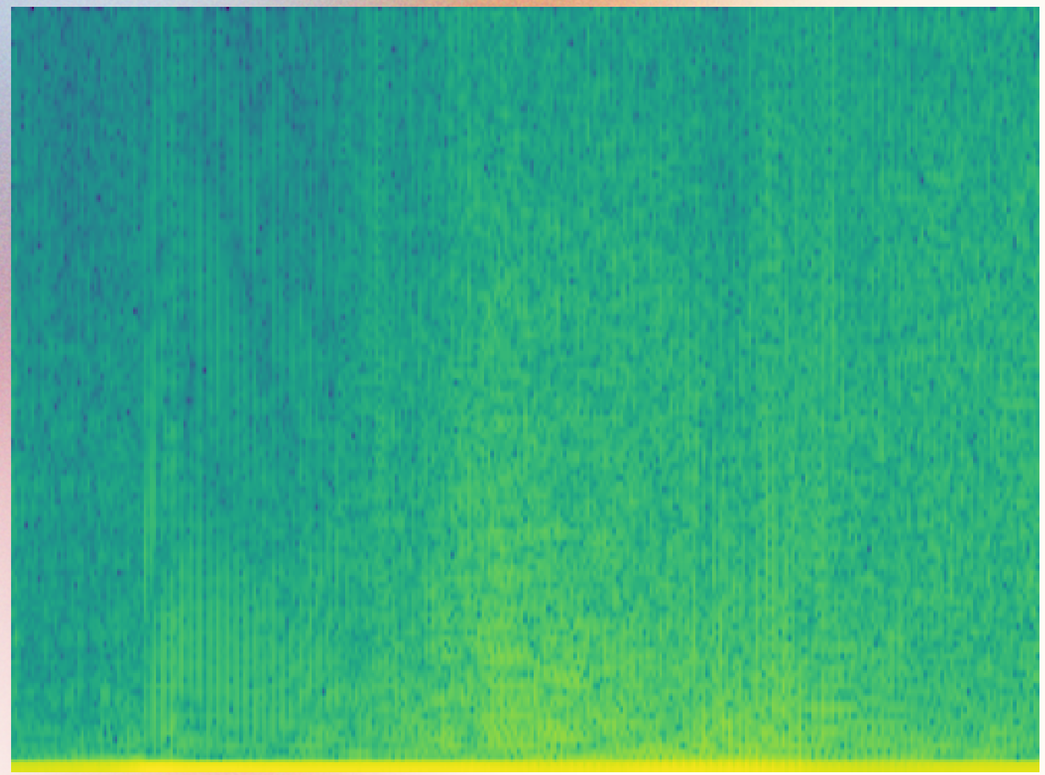
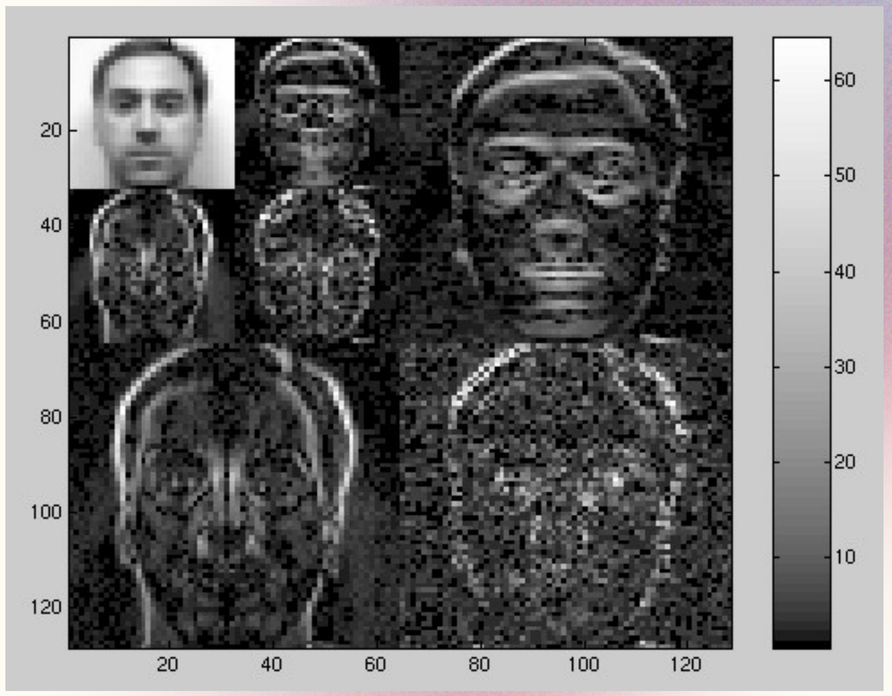


Image Spectrogram

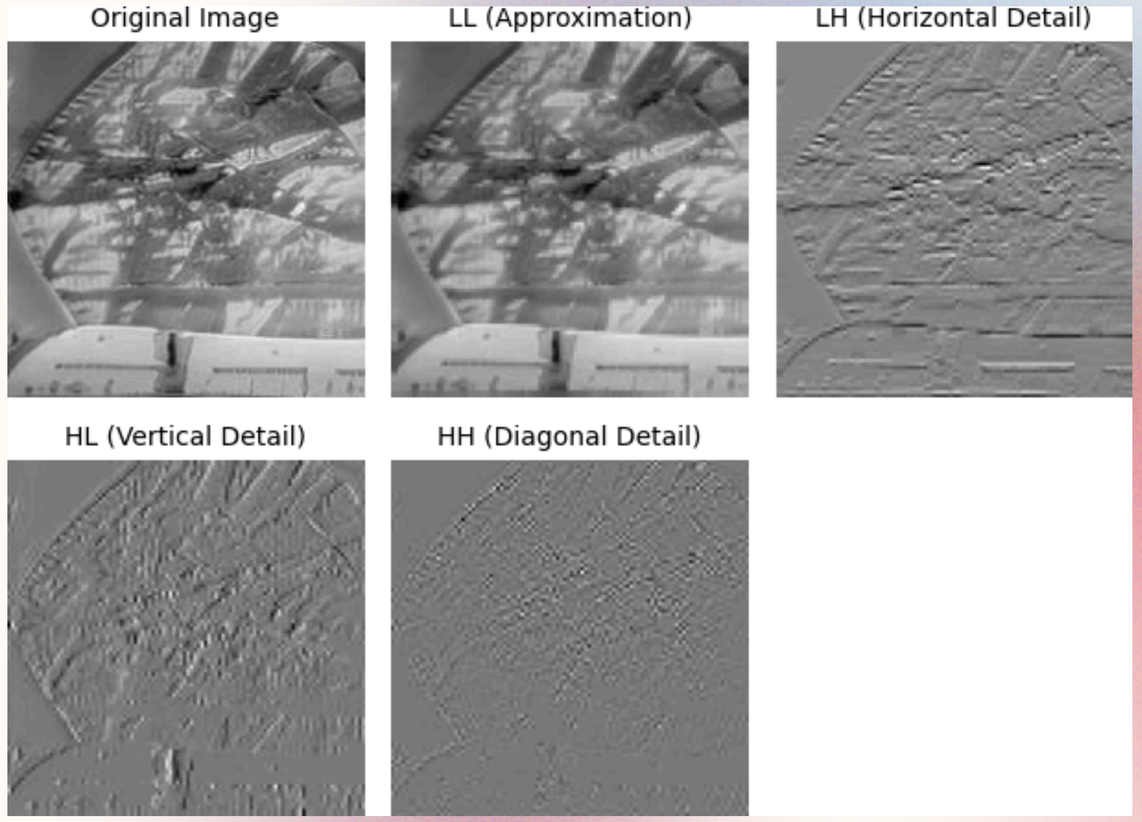
Discrete Wavelet Transform



Application of DWT on image processing

- Alternative to Fourier Transform
- Transforms image pixels to wavelets
- Multi-level frequency representation
- Used in Tang et al. and Liu et al.’s studies

DWT using PyWavelets



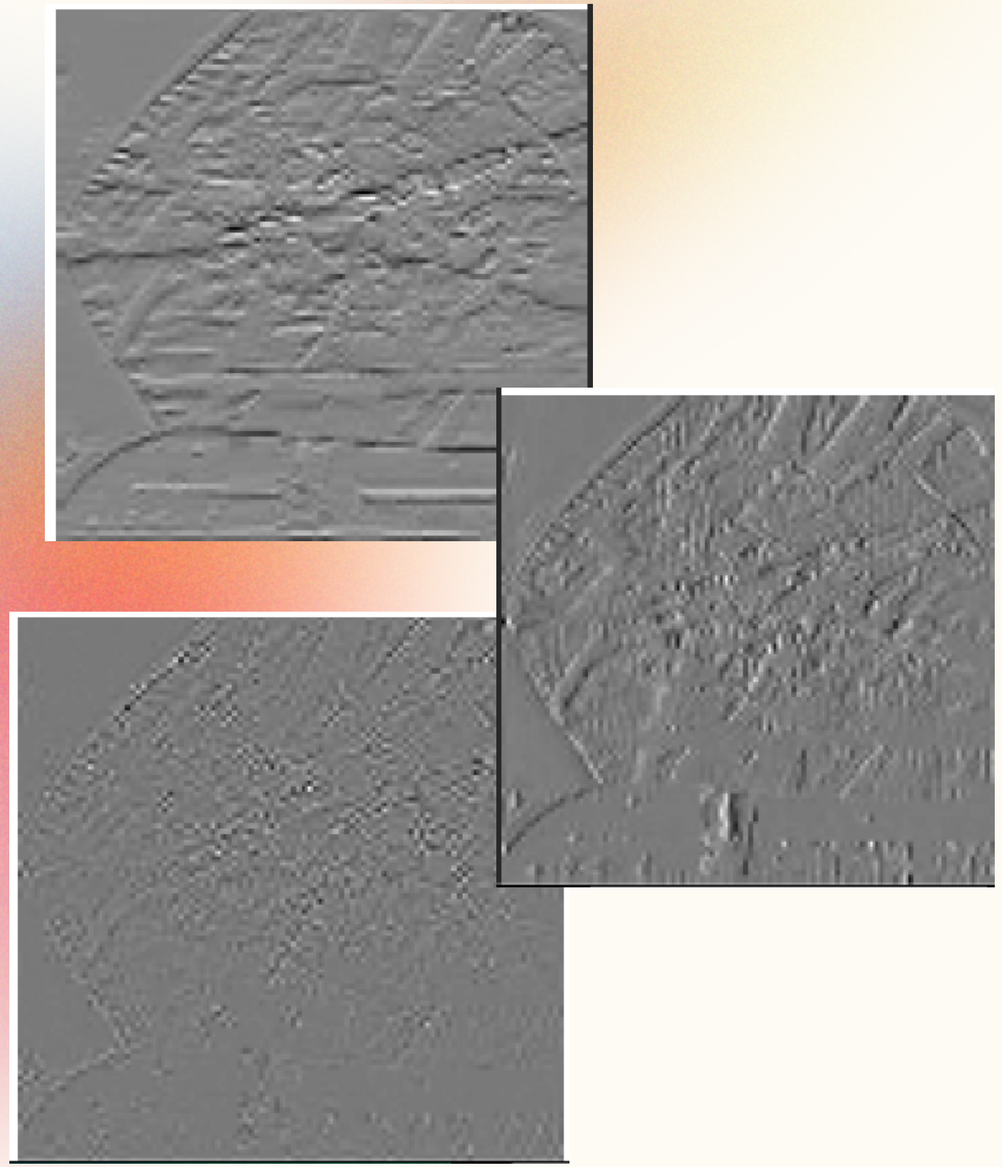
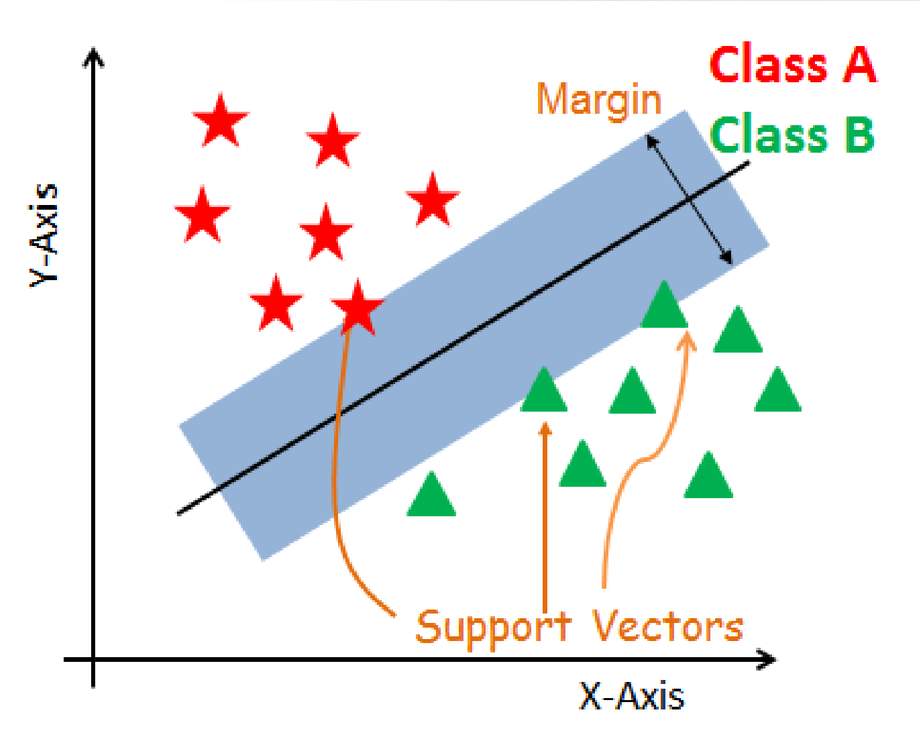
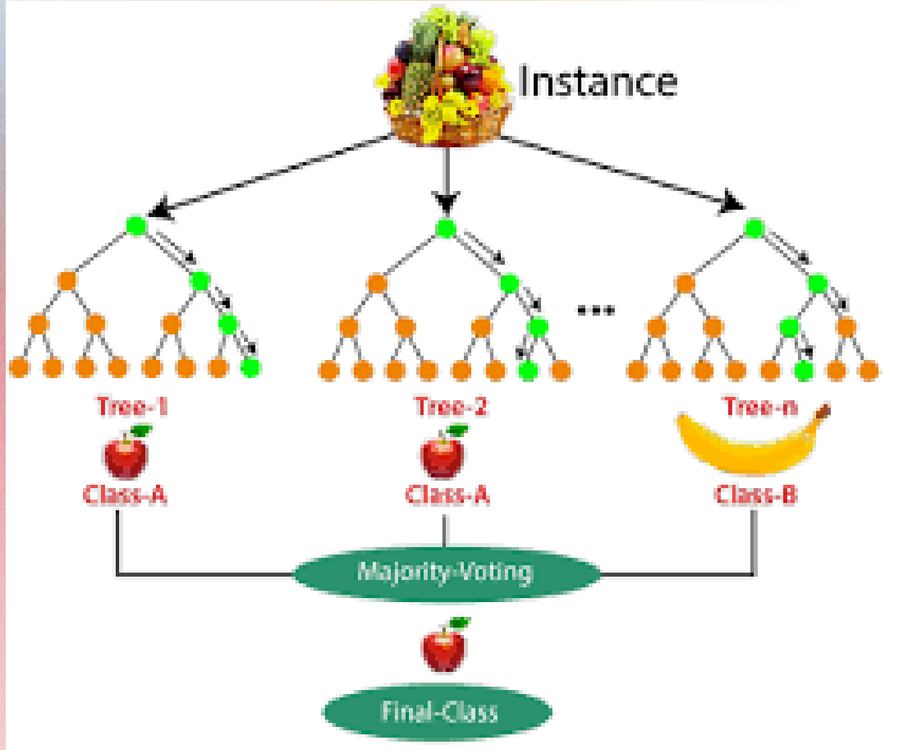


Image decomposition using DWT

Classification Models





Support Vector Machine

- High Dimensional Data
- Pattern Recognition
- Regression Analysis

Random Forest

- Decision Trees
- Overfitting
- Generalize

Sample Code & Initial Output