GRAVITATIONAL WAVE DETECTION

Data Mining and Analysis(Course Project ID : 5DAMCP07)

Team Number:A03

Submitted by:

Name	USN
Yashasvi A R	01FE19BCS017
Rohan Reddy S V	01FE19BCS027
Sharath S Shanbhag	01FE19BCS029
Anu B Herakal	01FE19BCS038

Academic year 2021-2022

Contents

1 **Introduction** 3
- 1.1 Overview of the project 3
- 1.2 Motivation 3
- 1.3 Current Status of Work 3

2 **Problem statement** 4

3 **Methodology** 5
- 3.1 1.Data Preprocessing: 5
- 3.2 Learning Models 6,7

4 **Results and Discussion** 8
- 4.1 Dataset used 8
- 4.2 Performance Evaluation metric 9
- 4.3 Exploratory Data Analysis 9
- 4.4 Results 10

5 **Conclusion** 11

6 **References** 11

1 Introduction

1.1 Overview of the project

G2Net is a network of Gravitational Wave, Geophysics and Machine Learning. Via an Action from COST (European Cooperation in Science and Technology), a funding agency for research and innovation networks, G2Net aims to create a broad network of scientists. From four different areas of expertise, namely GW physics, Geophysics, Computing Science and Robotics, these scientists have agreed on a common goal of tackling challenges in data analysis and noise characterization for GW detectors.

1.2 Motivation

Data mining techniques are used in the process of converting raw data into useful data and finding out the patterns in the data. Gravitational waves interact very weakly with matter (unlike EM radiation, which can be absorbed, reflected, refracted, or bent), they travel through the Universe virtually unimpeded, giving us a clear view of the gravitational-wave Universe. The waves carry information about their origins that is free of the distortions or alterations suffered by EM radiation as it traverses intergalactic space. The gravitational waves that LIGO detects are caused by some of the most energetic events in the Universe—colliding black holes, merging neutron stars, exploding stars, and possibly even the birth of the Universe itself. Detecting and analyzing the information carried by gravitational waves is allowing us to observe the Universe in a way never before possible, providing astronomers and other scientists with their first glimpses of literally unseeable wonders. LIGO has removed a veil of mystery on the Universe and in so doing, has ushered in exciting new research in physics, astronomy, and astrophysics.

1.3Current Status of Work

With the dataset provided, Using the most efficient model (EfficientNet-B7), The model yielded an accuracy of 0.8751.

2 Problem statement

The challenge had provided a data set containing the time series data from a network of 3 earth-based detectors (LIGO Livingston, LIGO Hanford, VIRGO). The main aim is to detect GW signals from the mergers of binary black holes. Specifically, a model is built to analyze simulated GW time-series data from a network of Earth based detectors

·

· We are provided with set of time series data containing simulated gravitational wave measurements from a network of 3 gravitational wave interferometers (LIGO Hanford, LIGO Livingston, and Virgo).

· Each time series contains either detector noise or detector noise plus a simulated gravitational wave signal. The task is to identify when a signal is present in the data (target=1)

· The parameters that determine the exact form of a binary black hole waveform are the masses, sky location, distance, black hole spins, binary orientation angle, gravitational wave polarization, time of arrival, and phase at coalescence (merger).

· These parameters (15 in total) have been randomized according to astrophysically motivated prior distributions and used to generate the simulated signals present in the data, but are not provided as part of the competition data.

· Each data set contains 3 time series (1 for each detector) and each spans 2sec and is sampled at 2048 Hz.

· Train users data contains 560000 rows and 2 columns , first column represents the id of the time series and second column represents the target value

(Target value is the probability of the presence of gravitational wave signal in the time series containing detector noise)

- The 3 time series of each datasets are measured by network of three gravitational wave interferometers
- Therefore the shape of time series is (3,4096) that is each series of length 2^12

2.1 Exploratory Data Analysis

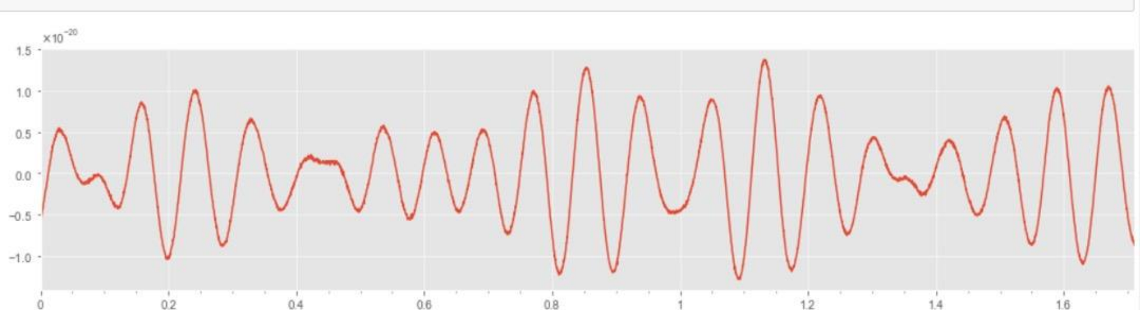
For the competition, we are provided with a training set of time series data containing simulated gravitational wave measurements from a network of 3 gravitational wave interferometers (LIGO Hanford, LIGO Livingston, and Virgo). Each time series contains either detector noise or detector noise plus a simulated gravitational wave signal. The task is to identify when a signal is present in the data (target=1).

Each data sample (npy file) contains 3 time series (1 for each detector) and each span 2 sec and is sampled at 2,048 Hz.

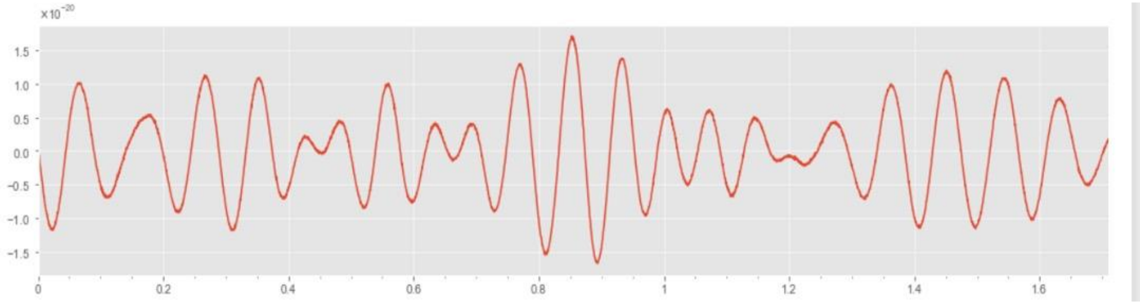
3. Methodology

The data in the form of numpy array is transformed to signal form.

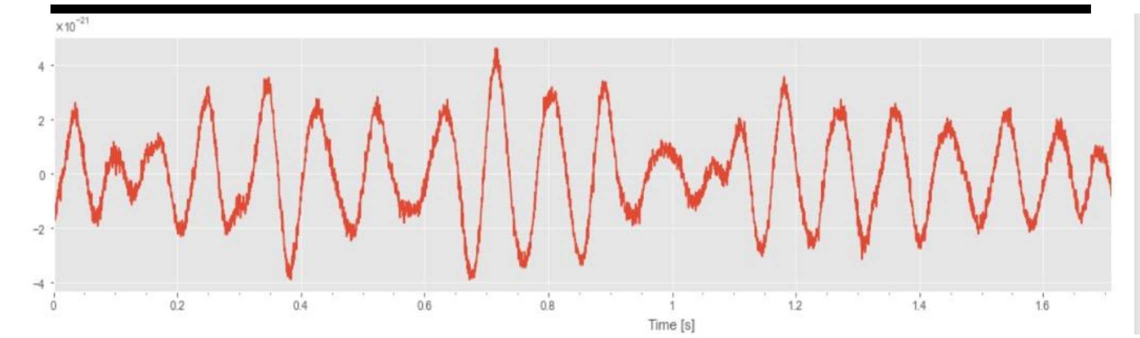
Signal from Ligo Hanford:



Ligo Livingston:



Virgo



Classifying the data into 2 classes of target 1 and target 0

- 0 – 280070
- 1 – 279930
- We can infer from the above data that we have almost same probabilities of getting gravitational wave signal in the time series containing detector noise

KDD process:

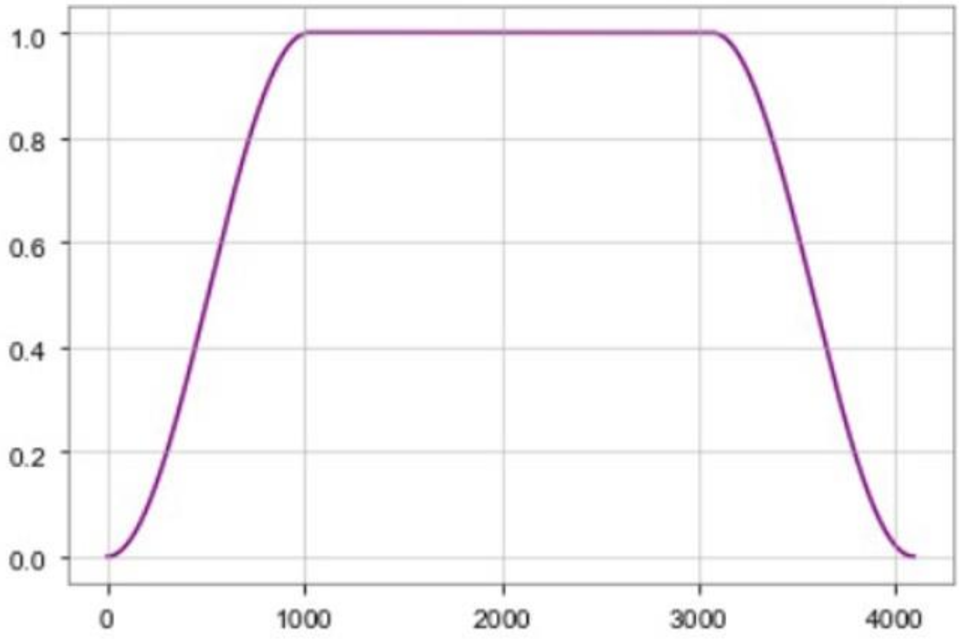
Iteration 1:
Preprocessing:

1.Tukey Window:

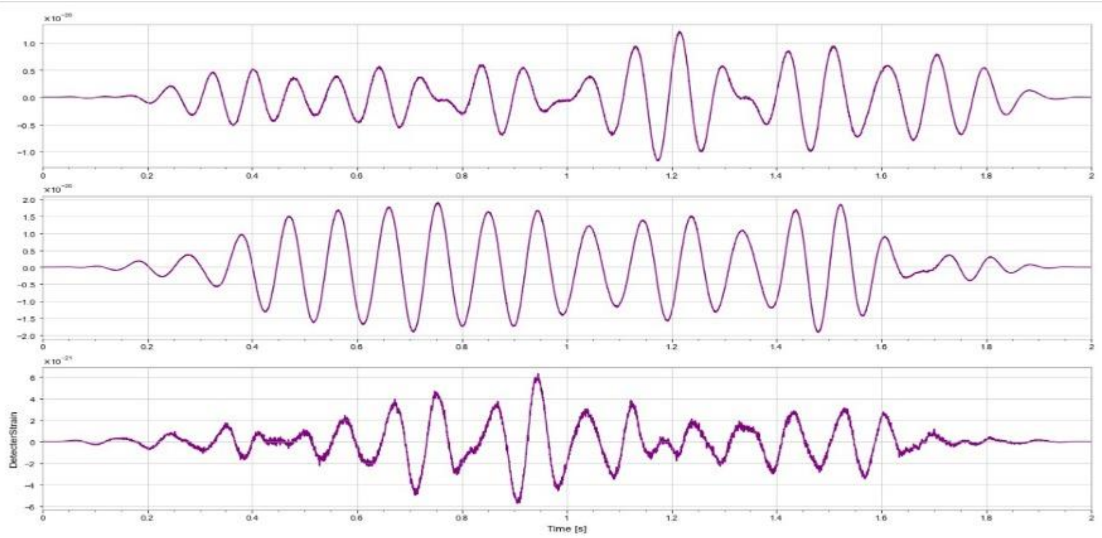
Apply a window function (Tukey - tapered cosine window) to suppress spectral leakage.

The Tukey window is a rectangular window with the first and last r /2 percent of the samples equal to parts of a cosine.

- ➔ w= tukeywin (L, r) returns an L-point Tukey window with cosine fraction r.
- ➔ w=>Tukey window, returned as a column vector.
- ➔ L=>Window length, specified as a positive integer.
- ➔ r=>Cosine fraction, specified as a real scalar.

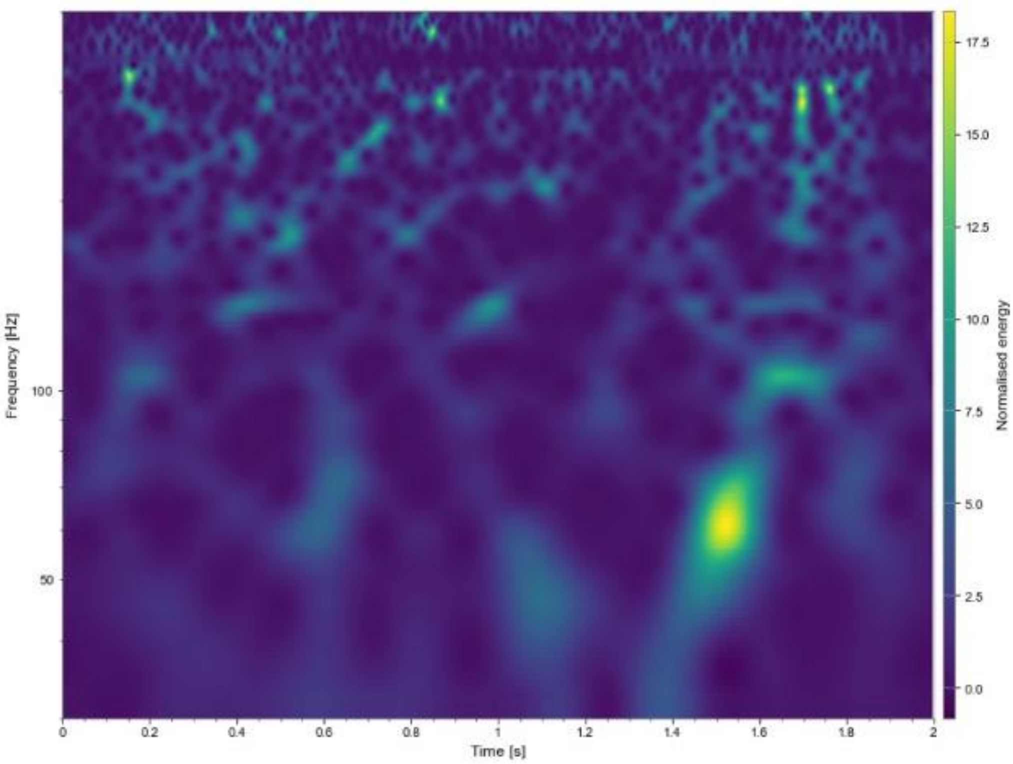


Resulting time series after applying tukey window function:



2. Q-Transform:

➔ Fourier transform is a means of mapping a signal, in the time or space domain into its spectrum in the frequency domain.
➔ DFT (Discrete Fourier Transform): Calculates the spectrum of a finite duration signal.
➔ FFT (Fast Fourier Transform): An algorithm for fast and efficient computation of the DFT.



3.Combining three channels into one RGB image:

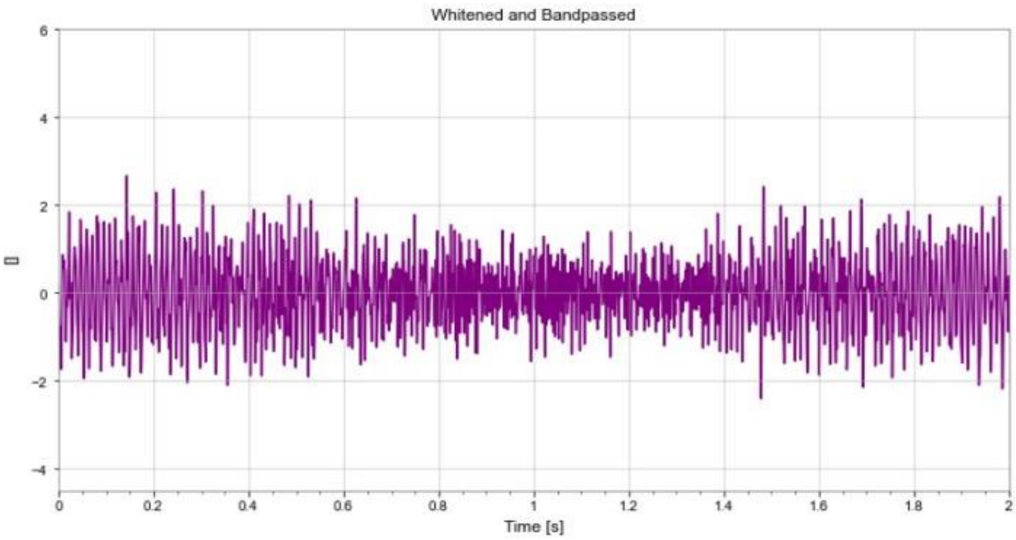
➔ Since we have 3 detectors, we can combine the Q-Transforms as RGB channels into one color image.

4.Band passing

- A good way to downweigh the Noise is through a spectral analysis. Basically, we have performed a Fourier

Transformation/Analysis to record the frequency of occurrence of signals with different frequencies.

- Then, we are to downway the more common frequencies by 'making them more silent' and make all the frequencies equally likely to occur.
- We have used range as 24Hz – 500 Hz, and have observed that the gravitational wave is dominant at 34Hz.



3.2 Learning Models

After preprocessing the data, we have worked on various learning models. The model is checked for accuracy before selecting and the model giving best accuracy is selected.

The models are:

4.2.1 ResNet 50d model

ResNet, short for Residual Networks, is a classic neural network used as a backbone for many computer vision tasks. ResNet are deep convolutional neural networks using residual connections.

4.2.2 EfficientNet-B7 model

EfficientNet is a convolutional neural network architecture and scaling method that uniformly scales all dimensions of depth/width/resolution using a compound coefficient. Unlike conventional practice that arbitrarily scales these factors, the EfficientNet scaling method uniformly scales network width, depth, and resolution with a set of fixed scaling coefficients. The compound scaling method is justified by the intuition that if the input image is bigger, then the network needs more layers to increase the receptive field and more channels to capture more fine-grained patterns on the bigger image.

4.2.3 EfficientNet-B0 model

EfficientNet-b0 is a convolutional neural network that is trained on more than a million images from the ImageNet database. The network can classify images into 1000 object categories, such as keyboard, mouse, pencil, and many animals. As a result, the network has learned rich feature representations for a wide range of images. The network has an image input size of 224-by-224.

4. Results and Discussion

4.1Dataset used

The problem data has

1. train:

•The training set files, one npy file per observation; labels are provided in a separate file
•Total npy Files = 5,60,000

2. test:

•The test set files; to predict the probability that the observation contains a gravitational wave
•Total npy Files = 2,26,000

3. training_labels.csv:

•The training set files, one npy file per observation; labels are provided in a separate file

• total rows and columns= 560000 rows x 2 columns

4. Sample_submission.csv:

•A sample submission file in the correct format - id & target

4.2 Performance Evaluation metric

Submissions are evaluated on area under the ROC curve between the predicted probability and the observed target

4.3 Results

method 1

Model	Before Tuning	After Tuning
Decision Tree	0.5036	0.5049
Random Forest	0.5168	……
Gradient boosting	0.5067	0.5135

method 2

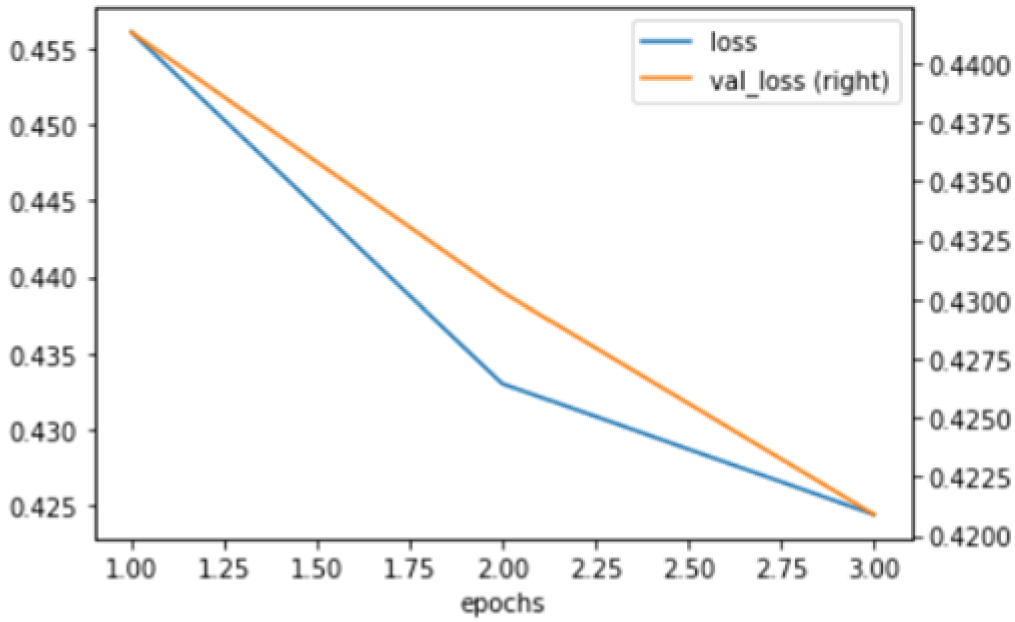
Model	Before Tuning	After Tuning
Decision Tree	0.509	0.528
Random Forest	0.514	……
Gradient boosting	0.545	0.873

As per the methodology two different preprocessing methods and three different classifier algorithms are used to build our model, where Gradient boosting on the second method of preprocessing gave the higher accuracy. Decision tree Classifier and Random forest are also used which have given lesser

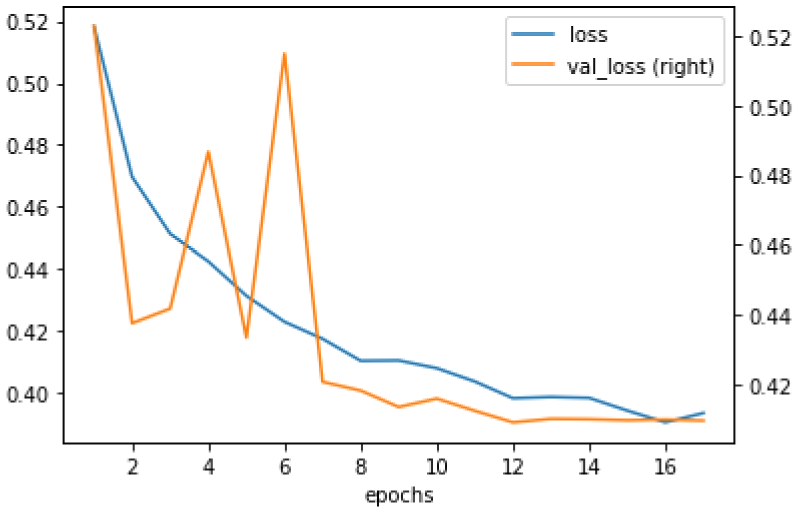
accuracy compared to Gradient boosting classifier.

4.4 graphs plot

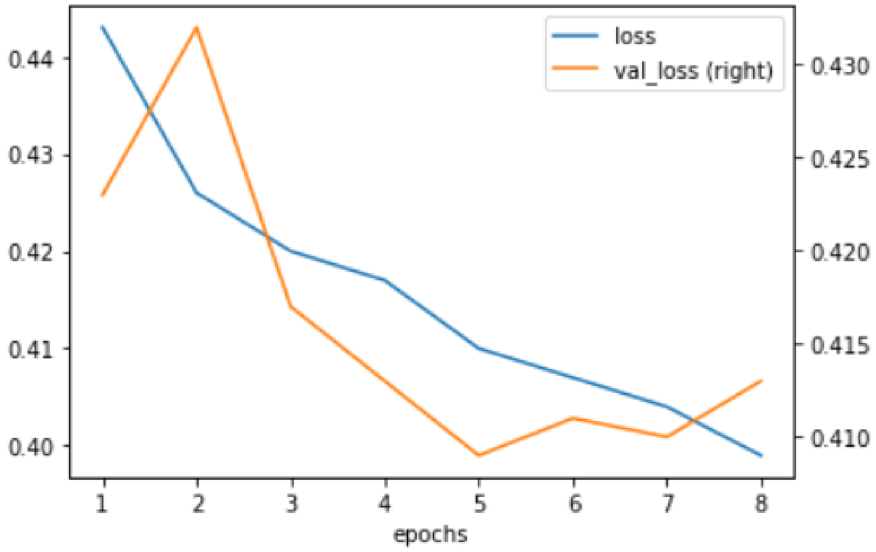
1.efficient net model



2.efficent net B7 model



3.resnet-50



5. Conclusion

This problem has a noisy data set and it is a classification problem. In this problem a model is built which can be able to detect the presence of a Gravitational wave. The final model is submitted on the challenge website, it achieved an accuracy of 0.8751. The final model had achieved 494th rank in the public leaderboard. Further improvement on the challenge would be ensembling and matched filtering, both of which can be explored in the future.

6.References

1.G2Net Gravitational Wave Detection | Kaggle – Challenge ,Data set links

2.(PDF) Machine Learning Based Analysis of Gravitational Waves (researchgate.net) - Machine Learning Based Analysis of Gravitational Waves

3.LIGO Scientific Collaboration - The science of LSC research – LIGO