Epileptic Seizure Prediction Using Intracranial Brain Records (EEG)

Submitted in partial fulfillment of the requirements of the degree of

"Bachelor of Engineering"

By

67. Pranav Karnani

68. Awais Mulla

69. Adarsh Pandey

Supervisor(s):

Prof. Amarja Ma'am

DEPARTMENT OF INFORMATION TECHNOLOGY

K.C. COLLEGE OF ENGINEERING & MANAGEMENT STUDIES & RESEARCH, THANE (E)

Certificate

This is to certify that the project entitled **"Epileptic Seizure Prediction Using Intracranial Brain Records (EEG) "** is a bonafide work of " **Pranav Karnani (67), Awais Mullah (68), Adarsh Pandey (69)"** submitted to the University of Mumbai in partial fulfillment of the requirement for the award of the degree of **"Bachelor of Engineering"** in **"Information Technology"**.

Name and sign

Supervisor/Guide

Name and sign

Co Supervisor/Guide

Name and sign

Head of Department

Name and sign

Principal

Project Report Approval for B.E.

This project report entitled (***Epileptic Seizure Prediction Using Intracranial Brain Records (EEG)***) by (***Pranav Karnani (67), Awais Mullah (68)***, ***Adarsh pandey (69)***) is approved for the degree of Bachelor of Engineering in **Information Technology.**

Examiners

1.---

2.---

Date:

Place:

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Name of student and Roll No. Signature

1. Pranav Karnani (67)

2. Awais Mullah (68)

3. Adarsh Pandey (69)

Date:

ABSTRACT

Epilepsy is the second most common brain disorder after migraine. Automatic detection of epileptic seizures can considerably improve the patients' quality of life. Current Electroencephalogram (EEG)-based seizure detection systems encounter many challenges in reallife situations. The EEGs are non-stationary signals and seizure patterns vary across patients and recording sessions. Moreover, EEG data are prone to numerous noise types that negatively affect the detection accuracy of epileptic seizures. To address these challenges, we introduce the use of a deep learning-based approach that automatically learns the discriminative EEG features of epileptic seizures. Specifically, to reveal the correlation between successive data samples, the timeseries EEG data are first segmented into a sequence of non-overlapping epochs. Second, Long Short-Term Memory (LSTM) network is used to learn the high-level representations of the normal and the seizure EEG patterns. Third, these representations are fed into Softmax function for training and classification. The results on a well-known benchmark clinical dataset demonstrate the superiority of the proposed approach over the existing state-of-the-art methods. Furthermore, our approach is shown to be robust in noisy and real-life conditions. Compared to current methods that are quite sensitive to noise, the proposed method maintains its high detection performance in the presence of common EEG artifacts (muscle activities and eye-blinking) as well as white noise.

Chapter 1
Introduction

1.1 Description

Deep Learning (LSTM) based Seizure prediction using Scalp based EEG Readings.

1.2 PROBLEM FORMULATION:

People suffering from epilepsy or also called seizure disorder, a disorder in which nerve cell activity in the brain is disturbed, causing seizures. These seizure do not have a fixed time they occur at random times causing great disturbance to emotional state and cognitive state of the mind. Their ability to work, social, economic situations comes to a sudden halt. Having the option to foresee epileptic seizures will incredibly improve the personal satisfaction of individuals with epilepsy by either giving them an admonition of an approaching seizure so they can move to security or enacting an embedded seizure control gadget that can turn away seizures through medication conveyance or electrical incitement of the cerebrum. How might we arrange the stage epileptic seizures utilizing electrical accounts of mind action?

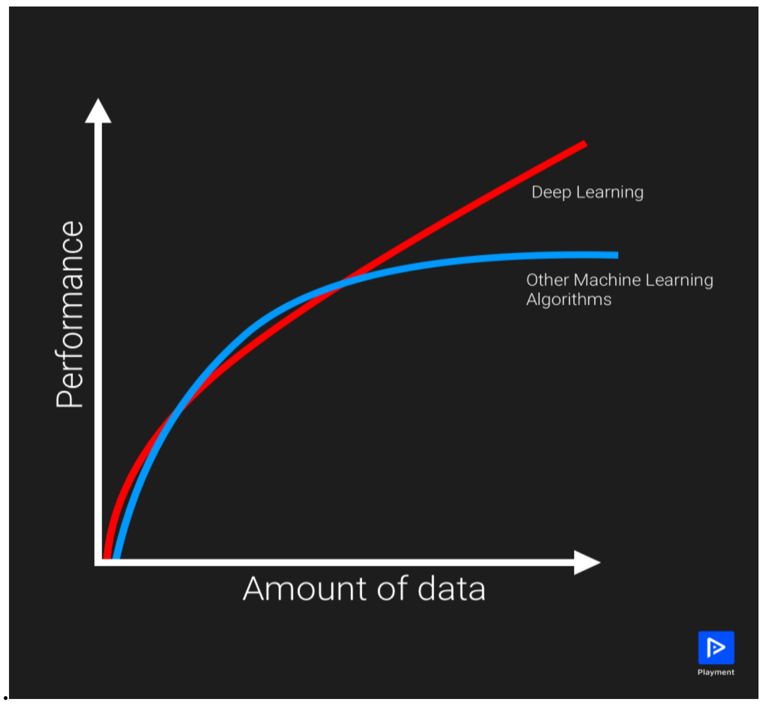
1.3 MOTIVATION:

This world consist of 70 million people having PWE i.e. epilepsy and nearly 12 million PWE patients reside in india; which amounts to about a sixth of the world. The general pervasiveness (3.5-12.1 per 1,600 populace) which has a frequency (0.3-0.9 per 1,400 populace for every year) information from ongoing examinations in India on all inclusive community are similar to the paces of high-income nations regardless of checked varieties in populace attributes & information systems. Having a differential conveyance of PWE among different socio-demographic & monetary gatherings with higher rates detailed for the male, country populace, and low financial status. A changing example in the age-explicit event of epilepsy with prevalence towards the more established age bunch is seen due to sociodemographic and epidemiological progress. Neuro-infections, neuro-cysticercosis , and neurotrauma alongside birth wounds have risen as significant hazard factors for optional PWE. In spite of its shifted etiology ,PWE are reasonable in outlook.
Programmed identification of epileptic seizures can impressively improve the patients' personal satisfaction. Current Electroencephalogram (EEG)- based seizure discovery frameworks experience numerous difficulties, all things considered, circumstances. The EEGs are non-fixed signs and seizure designs fluctuate across patients and recording meetings. Besides, EEG information are inclined to various clamor types that contrarily influence the recognition precision of epileptic seizures. To address these difficulties, we present the utilization of a profound learning-based methodology that naturally learns the discriminative EEG highlights of epileptic seizures. In particular, to uncover the connection between's progressive information tests, the time-arrangement EEG information are first sectioned into a succession of

nonoverlapping ages. Second, Long Short-Term Memory (LSTM) arrange is utilized to become familiar with the elevated level portrayals of the ordinary and the seizure EEG designs. Third, these portrayals are taken care of into Softmax function for training. The outcomes on a notable benchmark clinical dataset show the prevalence of the proposed approach over the current best in class strategies. Moreover, our methodology is demonstrated to be hearty in uproarious and genuine conditions. Contrasted with current techniques that are very touchy to clamor, the proposed strategy keeps up its high recognition execution within the sight of basic EEG ancient rarities (muscle exercises and eye-squinting).

1.4 Proposed Solution:

In contrast to increasingly customary strategies for machine learning methods, deep learning classifiers are prepared through component adapting as opposed to task-explicit calculations. This means the machine would understand the patterns in the pictures as it is given instead of requiring the human administrator to characterize the examples which the machines should look for in the picture. The component learning system is utilized each day by the way we show a kid to perceive distinctive objects. Deep learning is a sort of AI that impersonates the neuron of the neural systems present in the human cerebrum. PC Vision Deep learning models are prepared on a lot of pictures a.k.a preparing information, to explain an undertaking. These deep learning models are for the most part utilized in the field of Computer Vision which permits a PC to see and picture as a human would. This kind of system is like the organic sensory system, with every hub going about as a neuron inside a bigger network. Thus, these models are a class of fake neural systems. This models calculations adapt dynamically about the picture as it experiences each neural system layer. Prior layers figure out how to recognize low-level highlights like edges, and ensuing layers consolidate highlights from prior layers into an increasingly all-encompassing
And complete potryal

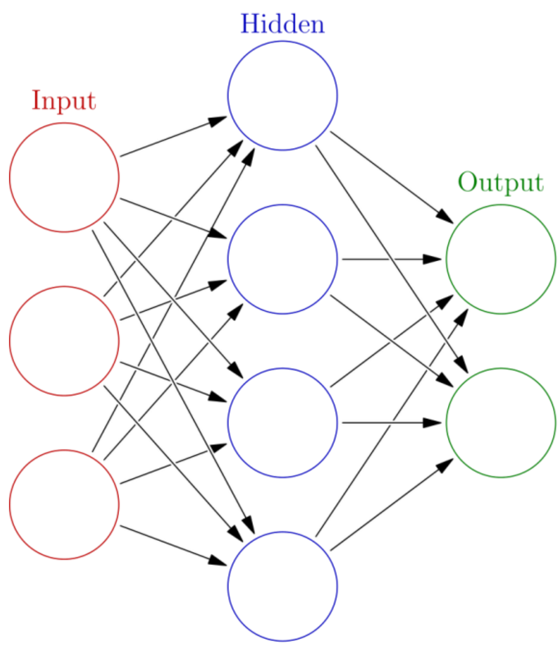


1.4.1 Deep Learning vs Machine Learning Algorithms.

Deep learning is a sort of AI that emulates the neuron of the neural systems which human mind has. Deep learning models are prepared considering lot of pictures for preparing information, to comprehend an assignment. These deep learning models are for the most part utilized in the sector of Computer Vision, which allows a PC to compare and envision as a real preson would.

These models could be visualized as a group of points wherein each of those points makes a decision based on the nodal input. The human biological neural system as has the same type of network, wherein each node has a neural within a network which is larger than itself.Thus, these models are a class of fake neural systems. Deep learning algorithms adapt continuously about the picture as it experiences each neural system layer. Early layers figure out how to distinguish low-level highlights like edges, and resulting layers consolidate highlights from prior classes into a progressively all-encompassing and thorough portrayal.

ANN is a system that simulates the neural networks present in living beings. The work in a similar way as to that of animals or more closely like humans. They aren't programmed for a specific task. We use examples to make them learn. The best use of ANN may be image recognition. Labelled pictures of an object such as dogs or cats are manually fed to the system. No information of the constituents of the picture are entered. The ANN takes the features into account while creating the model.

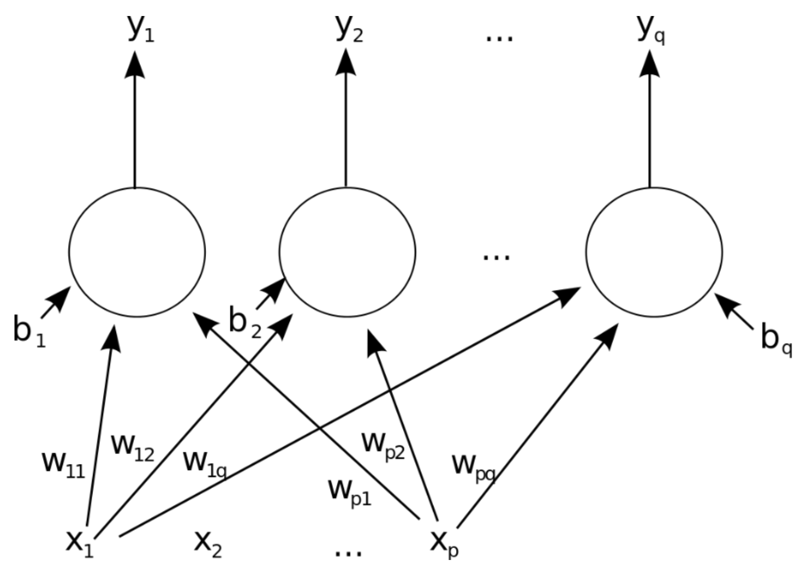


1.4.2Basic Model Of ANN

Artificial neurons are connected to each other. It mimics the human brain. Each neuron can transmit signal to the other neurons. A neuron receives a signal then processes it and sends this processed signal further. The input signal is a real number. A non-linear function computes the sum of inputs and passes it ahead. The connections are known as edges.

The learning is adjusted by the weight associated with the edges. The weight increases or decreases the strength of the signal at a connection. A threshold determines which signal goes across and which doesn't. There are multiple layers between the input and output layers. There can be zero or one or more number of layers. Different transformations happen to the signal at each layer providing meaningful changes to the system as whole. Right now ANN is only used to solve a specific type of problem. General intelligence is still not possible. Therefore ANNs can be used for a variety of things like image recognition, understanding speech or grammar and in this case medical diagnosis.

Each link between the neurons or nodes have influence over the other nodes. The whole system forms a directed or undirected graph based on what it is trying to achieve.



1.4.3Single Layer of ANN

PRINCIPLE

The basic working is based on the principle that a small change in input brings a large change in output. A given neuron can have multiple inputs or a single input. The same is true for outputs as well.

Propagation function computes the output by processing the inputs from the previous neuron.

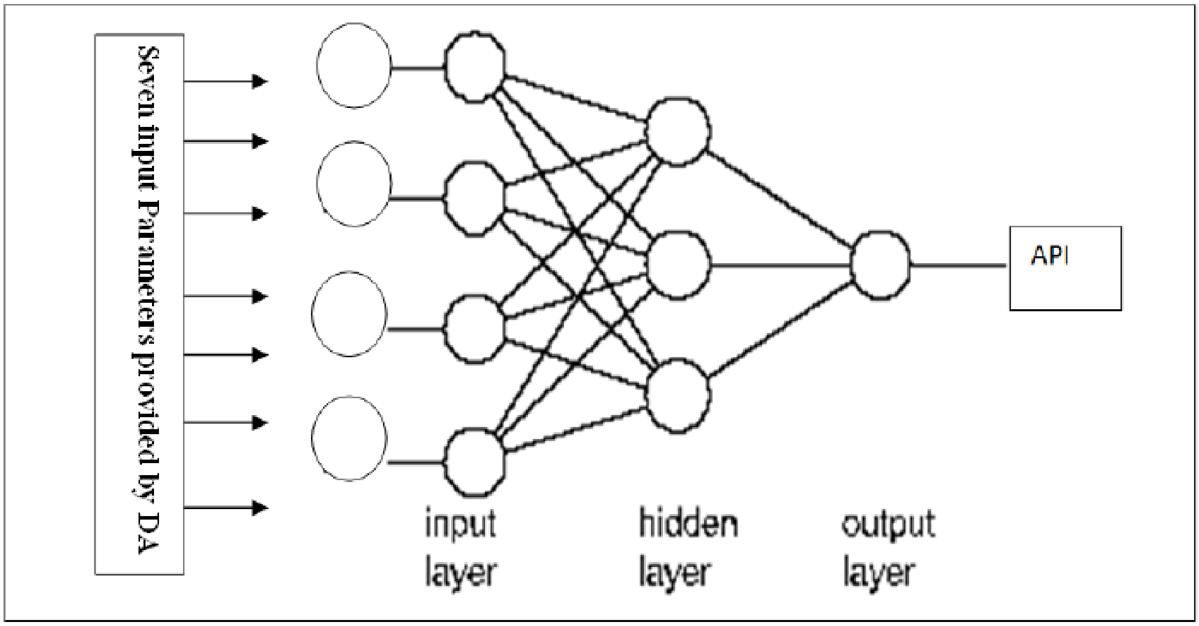
Deep Learning may use multiple layers as it is more advanced in computing results. Multiple connections are also possible between the layers. It could be a many-to-one connection or one-to-one.

LEARNING

The actual process of making a system which computes data correctly is this step. Sample data is used to adjust the weights of the network to improve accuracy. Errors are minimized. The process happens over and over again till the error cannot be reduced any further. The error rate cannot be reduced to zero.

We may redesign the network if the output does not prove to be accurate. Hundred percent accuracy is not possible in such a system but ninety plus is generally considered good.

Learning rate is the steps the model takes to adjust errors in each observation. A high learning rate gets the work done faster but the accuracy suffers.



1.4.4 Multilayer ANN

Feedforward ANN

The information flow is unidirectional. The information does not return to a neuron once it leaves it. It is used in Simple Pattern recognition. The inputs and outputs are fixed.

Feedback ANN

The same neuron may get a signal multiple times

Supervised Learning

Here the system comes up with guesses as to what the correct answer may be. The correct answer is given to the system. According to the difference between the answer and guess the network makes changes. The cost function here eliminates incorrect deductions. The most commonly used cost is mean-squared error. Pattern recognition and regression are done using this method. Handwriting and gesture recognition problems can be dealt with using this technique.

Unsupervised Learning

Here the output in not known therefore the correct answers to the dataset aren’t available. This could be used when searching for new patterns. The input data along with cost function are provided at the input.

Reinforcement Learning

It works on the principle of observing the environment. The best example of this method is playing video game. The environment could give you unpredictable response in every game but you start learning from each game. The goal is to get the lowest cost possible. The goal has to be achieved as early as possible. Initially everything is estimated but as the game progresses things start getting better.

Backpropagation

Backpropagation is used to modify the weights to compensate for the errors. Mathematically the derivative of the cost function is linked with a given state.

Applications

Used is quantum chemistry, playing games, face identification, signal classification, 3D construction sequence recognition, medical diagnosis, controlling car, trajectory prediction, finance, trading, data mining, social network filtering and spam filtering. It is also used in diagnosis of cancer.

It is also used in geoscience, coastal engineering, cybersecurity and geomorphology.

It can also be used to simulate properties of quantum systems although it is still in research stage.

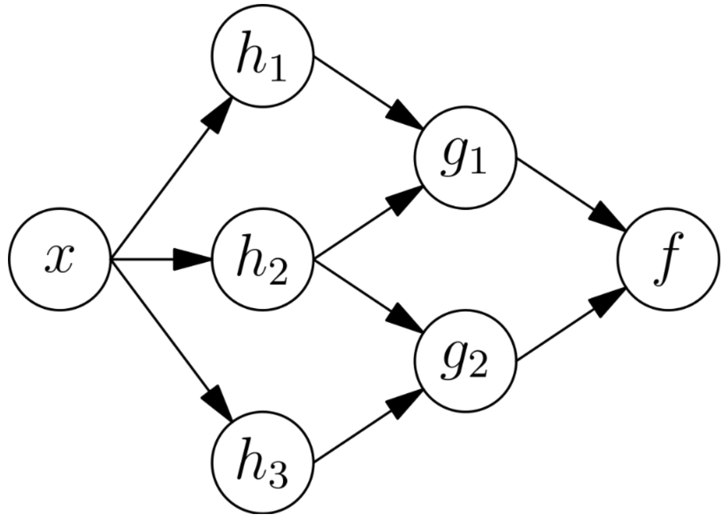
EEG, ECG analysis, designing prosthetics, speech classification, text to speech conversion, image compression, automated information services, braking vehicles, audio signal processing, anomaly detection, guidance system, weapons, electronics, manufacturing products etc.

Processing Power Required

Huge and viable neural systems require significant computer resources. While the brain has equipment custom fitted to the undertaking of preparing signals through a chart of neurons, recreating even a rearranged neuron on von Neumann design may expend immense measures of memory and capacity. Moreover, the fashioner frequently needs to transmit flags through a large number of these associations and their related neurons – which require colossal CPU force and time.

The resurgence of neural systems in the twenty-first century is to a great extent owing to propels in equipment: from 1991 to 2015, registering power, particularly as conveyed by GPUs has expanded around a million-overlay, making the standard backpropagation calculation achievable for preparing systems that are a few layers further than before. The utilization of quickening agents, for example, FPGAs and GPUs can diminish preparing times from months to days.

Neuromorphic building tends to the equipment trouble straightforwardly, by developing non-von-Neumann chips to legitimately actualize neural systems in hardware. Another kind of chip upgraded for neural system handling is known as a Tensor Processing Unit, or TPU.



1.4.5 Dependency graph

Mathematics

Applications where the system generalizes is all the inputs is seen as an overtrained model. This happens when the network capacity significantly exceeds the free parameters. We can use cross-validation and many other to check the presence and minimize error.

We can also use regularization where a larger probability is use over simpler models.

Supervised networks that use mean squared error cost function use statistics to determine the confidence of the trained model. Variance can also be found using MSE. This value can then be used to calculate the confidence interval of network output, assuming a normal distribution. A confidence analysis made this way is statistically valid as long as the output probability distribution stays the same and the network is not modified.

By allocating a softmax initiation work, a speculation of the strategic capacity, on the yield layer of the neural system (or a softmax segment in a segment based system) for all out objective factors, the yields can be deciphered as back probabilities. This is valuable in characterization as it gives a sureness measure on orders.

Advantages of using ANN
Learning what the ANN has learned is much easier to analyse than that of the biological neural network. Also more and more research involving the learning algorithms of the system are uncovered which help the data process faster. Local vs Non-Local Learning.

Advanced
The current system may be linked to the human mind in the future for us to get all the data at lightning speed.

Types of learning
Two modes are available. Stochastic and Batch. In stochastic learning, each input makes a weight adjustment. In batch learning weights are adjusted based on a batch of inputs, accumulating errors over the batch. Stochastic learning introduces "noise" into the process, using the local

gradient calculated from one data point; this reduces the chance of the network getting stuck in local minima. However, batch learning typically yields a faster, more stable descent to a local minimum, since each update is performed in the direction of the batch's average error. A common compromise is to use "mini-batches", small batches with samples in each batch selected stochastically from the entire data set.

Artificial neural systems are further classified in 3 classes.

They are:

- Multilayer Perceptrons (MLPs)
- Convolutional Neural Networks (CNNs)
- Recurrent Neural Networks (RNNs)

These 3 styles of systems give a great deal of adaptability & have demonstrated themselves over many years to be helpful and dependable in a huge scope of issues. Who likewise contains numerous sub-types for helping them practice them to the characteristics of various framings of expectation issues & diversified datasets.

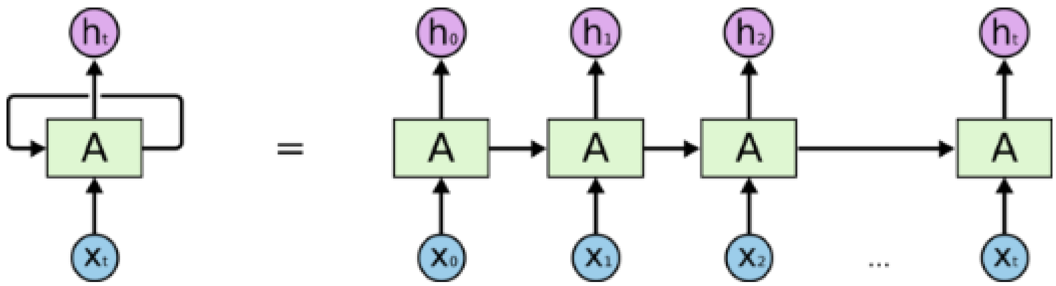
Its common from the vibes of the dataset that going with RNNs is the best choice.

We can use Recursive Neural Network for:

- Text data
- Speech data
- Classification prediction problems
- Regression prediction problems
- Generative models
- Time Series Data

1.4.1 What are RNN’s?

The idea at the back of Recursive Neural Networks is to use sequential facts. In a conventional neural network we assume that each one inputs (and outputs) are not dependent to every other. But for few duties that’s a poor idea. If someone desire to anticipate the further phrase in a sentence one should know which words came prior to it. Recursive Neural Networks are referred to as recurrent because they carry out the equal project for every element of a sequence, with the output being depended on the preceding calculations and also you already recognize that they have a “memory” which accounts the result approximately calculated till now.

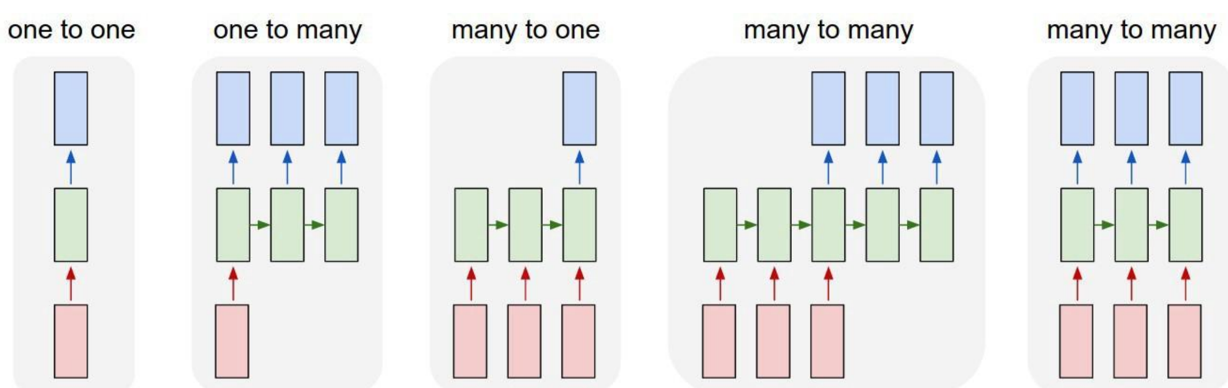


An unrolled recurrent neural network.

“Whenever there is a sequence of data and that temporal dynamics that c
more important than the spatial content of each individual frame.”

– Lex Fridman (MIT)
Various kinds of RNN’s:

The center explanation that recurrent network are all the more energizing is that they permit us to work over groupings of vectors: Sequences in the info, the yield, or in the most broad case both. A couple of models may make this increasingly concrete:



1.4.6 Types of RNN

Various kinds of Recurrent Neural Networks.
(1) Sequence yield (for example picture subtitling takes a picture and yielding few sentences consisting words).
(2) Sequence input (for example sentiment examination wherein a sentence given is delegated communicating +ve or -ve assumption).
(3) Sequence input-output(for example Machine Translation: a RNN peruses a sentence in English and afterwards yielding a sentence in Spanish).
(4) Synced sequence input and output(for example video characterization where we wish to mark each casing of the video). Notice that for each situation are no pre-indicated limitations on the lengths groupings in light of the fact that the intermittent change (green) is fixed and can be applied the same number of times as we like.

Every single rectangle in the images shown above shows us vectors and arrows representing functions. Red colour attributes input vectors while the blue ones are output vectors. The green contains the RNN's state.

One-to-One(O2O):
This is called as Plain-Vaniall Neural systems likewise. It manages constraint size of contribution to Constant size of Output wherein they are free of past data/yield.

Model: Image characterization.
One-to-Many(O2M):
This manages constant size of data as information that outputs grouping of information just as yield.
Example: The same happens in Image Captioning where it accepts picture as info and yields a sentence.

Many-to-One(M21):
It takes a series of data as information & yields a fixed size of yield.
Example: When a given sentence is classified as expressing +ve or -ve sentiment is called sentiment analysis.

Many-to-Many(M2M):
It accepts a Series of data considering it information & procedure it intermittently yields a Sequence of information.

Model: Machine Translation, where a RNN peruses a sentence in English and afterward yields a sentence in French.

Bidirectional Many-to-Many:

Consider that for each situation are no pre-indicated limitations on the lengths groupings in light of the fact that the repetitive change (green) is fixed and can be applied the same number of times as we like.

Model: video characterization wherein we want to characterize individual frame consisting in the video.

Data Acquisition and Preprocessing:
a. Intracranial Electroencephalograms i.e. 10-minute long recording of the brain, is collected from
different 15 epileptic patients by using 16 channels surgically inculcated at different position of electrodes segmented at 400 Hz.
b. Each recording signals stores either preictal or interictal state of epileptic seizure.
c. Then, EEG channels are converted into single signal called as surrogate channel in order to increase the signal-noise ratio by using averaging filter and laplacian filter.

Feature Extraction:
a. After preprocessing of signals, features are extracted from the noise-free channels to train and fit the neural network model.

b. Features are extracted in two levels:
i. At the first level, the energy, spectrum, length of curve, entropy, mean, variance are extracted between the positive (preictal) and negative (interictal) samples of signals.
ii. Then the extracted first level feature becomes the input to the second level, then the predicted features are calculated from each of the signal samples.
iii. The prediction indicator features are mean, median, slope, derivate, entropy, sum, variance standard deviation, maximum, slope, skewness, curve length, variance.

Long Short Term Memory Model:

a. Segmentation
b. i. The iEEG signal are further divided into smaller epochs or segments.
c. ii. These epochs are converted into stationary data for better performance and evaluation.
d. iii. Each of these epochs or segments are of specific lengths and performance indicator
e. features are extracted from epochs of specific length L.
f. b. Learning
g. i. The extracted features of signals are passed to LSTM layer which contains 120
h. neurons.
i. ii. The output of LSTM layer is sent to the Average Pooling layer activation
j. function, in order to consider the equal label prediction from all epochs.
k. c. Classification
l. i. The output of the average pooling layer is sent to the softmax layer.
m. ii. Softmax layer gives the probability distribution of each class label such as preictal or
n. interictal state.
o.
p. Average Pooling – It includes finding out the mean for each patch within the characteristic map. This way every 2×2 square within the feature map is down sized to the common content inside the square. Consider an example:- the result of the line detector convolutional filter within the preceding section was a 6×6 function map
q.
r. Softmax Classifier - Softmax Regression (equivalent words: Multinomial Logistic, Maximum Entropy Classifier, or just Multi-class Logistic Regression) is a speculation of calculated relapse that we can use for multi-class characterization (under the presumption that the classes are fundamentally unrelated). Conversely, we utilize the (norm) Logistic Regression model in double characterization assignments.

1.4.2 CNN

Convolution comes from the Latin convolvere, "to convolve" means to roll together. Convolution is a mathematical operation on two functions (f and g) to construct a third function that represents how the shape of one is modified by the other. It is the integral calculating how much two functions overlap as one passes over the other. Assume convolution as a way of combining two functions by multiplying them. 2012 was the first year that neural nets grew to prominence as Alex Krizhevsky used them to win that year's ImageNet competition (basically, the annual Olympics of computer vision), dropping the classification error record from 26% to 15%, an astounding improvement at the time.Ever since then, a host of companies have been using deep learning at the core of their services. Facebook uses neural nets for their automatic tagging algorithms, Google for their photo search, Amazon for their product recommendations, Pinterest for their home feed personalization, and Instagram for their search infrastructure.

Whenever we see an image, our brain looks for the features in the image to classify that image. We categorize things by recognizing the features. If we look to the right side of the image, we will see a person looking towards the right side while, if we look in the center, we will perceive that the person is looking towards us. Does our brain struggle a lot in identifying these different scenarios and is confused if the person is looking in right or towards us? This happens because our brain studies for the features in the image and then assume what to take.

No! That's because the features depicted are inadequate to assist the brain in classifying them. All the above images are addressed to understand that our brain functions on the features of the image it sees and then classifies it accordingly. In a similar manner, neural networks work. We can see in the image below, the neural network has successfully classified cheetah and bullet train but was unsuccessful in predicting hand glass. This is because of the unclear features in the image. In simple words, Neural Networks works exactly like a human mind.

Algorithm: -

1. Convolution

A convolution is a combined integration of two functions that shows you how one function

modifies the other.

$$
\begin{aligned}
(f * g)(t) &\stackrel{\text{def}}{=} \int_{-\infty}^{\infty} f(\tau) g(t-\tau)\, d\tau \\
&= \int_{-\infty}^{\infty} f(t-\tau) g(\tau)\, d\tau .
\end{aligned}
$$

There are three important items to say during this process: the input image, the feature detector, and therefore the feature map. The input image is the image being detected. The feature detector

may be a matrix, usually 3x3 (it could even be 7x7). A feature detector is additionally mentioned as a kernel or a filter.
Intuitively, the matrix representation of the input image is multiplied element-wise with the feature detector to supply a feature map, also referred to as a convolved feature or an activation map. The aim of this step is to scale back the dimensions of the image and make processing faster and easier. Some of the features of the image are lost during this step.
However, the most features of the image that are important in image detection are retained. These features are the ones that are unique to identifying that specific object. For example each animal has unique features that enable us to spot it. The way we prevent loss of image information is by having many feature maps. Each feature map detects the situation of certain features within the image.

Edge Detection Example
In the previous article, we saw that the first layers of a neural network detect edges from a picture . Deeper layers could be ready to detect the explanation for the objects and even more deeper layers might detect the explanation for complete objects (like a person’s face).
In this section, we'll specialise in how the sides are often detected from a picture . Suppose we are given the below image:

1.4.7 Edge detection example

As you can see, there are many vertical and horizontal edges in the image. The first thing to do is to detect these edges:

But how do we detect these edges? To illustrate this, let’s take a 6 X 6 grayscale image (i.e. only one channel):

3	0	1	2	7	4
1	5	8	9	3	1
2	7	2	5	1	3
0	1	3	1	7	8
4	2	1	6	2	8
2	4	5	2	3	9

Next, we convolve this 6 X 6 matrix with a 3 X 3 filter:

3	0	1	2	7	4
1	5	8	9	3	1
2	7	2	5	1	3
0	1	3	1	7	8
4	2	1	6	2	8
2	4	5	2	3	9

6 X 6 image

*

1	0	-1
1	0	-1
1	0	-1

3 X 3 filter

After the convolution, we will get a 4 X 4 image. The first element of the 4 X 4 matrix will be calculated as:

3^{1}	0^{0}	1^{-1}
1^{1}	5^{0}	8^{-1}
2^{1}	7^{0}	2^{-1}

So, we take the primary 3 X 3 matrix from the 6 X 6 image and multiply it with the filter. Now, the primary element of the 4 X 4 output are going to be the sum of the element-wise product of those values, i.e. 3*1 + 0 + 1*-1 + 1*1 + 5*0 + 8*-1 + 2*1 + 7*0 + 2*-1 = -5. To calculate the second element of the 4 X 4 output, we'll shift our filter one step towards the proper and again get the sum of the element-wise product:

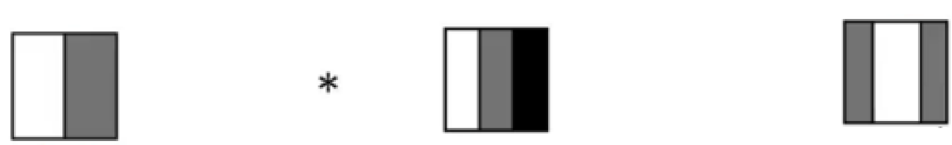
0^{1}	1^{0}	2^{-1}
5^{1}	8^{0}	9^{-1}
7^{1}	2^{0}	5^{-1}

So, convolving a 6 X 6 input with a 3 X 3 filter gave us an output of 4 X 4.

-5	-4	0	8
-10	-2	2	3
0	-2	-4	-7
-3	-2	-3	-16

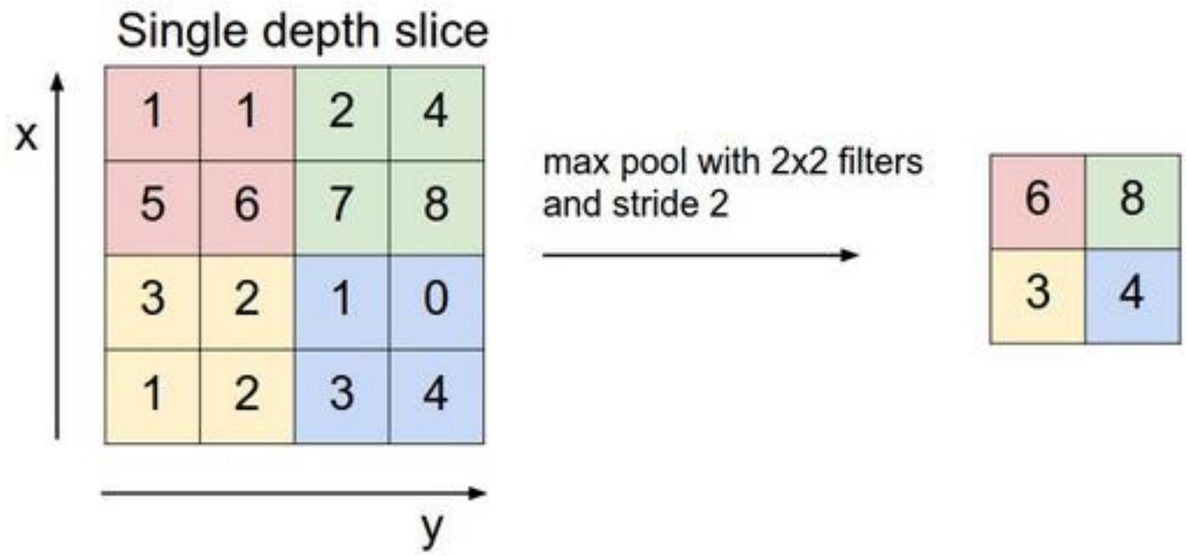
So, convolving a 6 X 6 input with a 3 X 3 filter gave us an output of 4 X 4. Consider one more example:

$$\begin{bmatrix} 10 & 10 & 10 & 0 & 0 & 0 \\ 10 & 10 & 10 & 0 & 0 & 0 \\ 10 & 10 & 10 & 0 & 0 & 0 \\ 10 & 10 & 10 & 0 & 0 & 0 \\ 10 & 10 & 10 & 0 & 0 & 0 \\ 10 & 10 & 10 & 0 & 0 & 0 \end{bmatrix} * \begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix} = \begin{bmatrix} 0 & 30 & 30 & 0 \\ 0 & 30 & 30 & 0 \\ 0 & 30 & 30 & 0 \\ 0 & 30 & 30 & 0 \end{bmatrix}$$



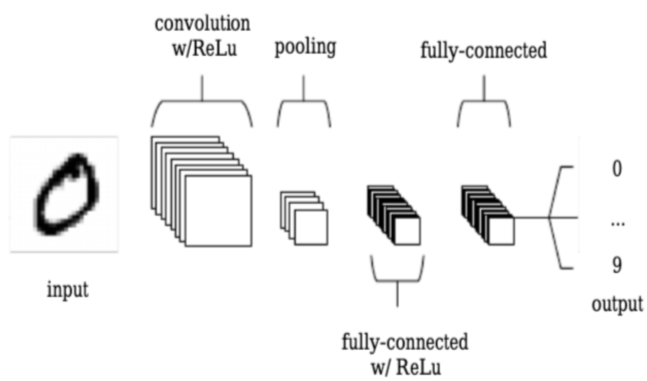
Max Pooling/Down Sampling with CNN

The next layer during a convolutional network has three names: max pooling, downsampling and subsampling. The activation maps are fed into a downsampling layer, and like convolutions, this method is applied one patch at a time. In this case, max pooling simply takes the most important value from one patch of a picture , places it during a new matrix next to the max values from other patches, and discards the remainder of the knowledge contained within the activation maps.



Only the locations on the image that showed the strongest correlation to every feature (the maximum value) are preserved, and people maximum values combine to make a lower-dimensional space.
Much information about lesser values is lost during this step, which has spurred research into alternative methods. But downsampling has the advantage, precisely because information is lost, of decreasing the quantity of storage and processing required.

CNNs are comprised of three types of layers. These are convolutional layers,pooling layers and fully-connected layers. When these layers are stacked, a CNN architecture has been formed. A simplified CNN architecture for MNIST classification is illustrated in Figure below



1.4.8 CNN

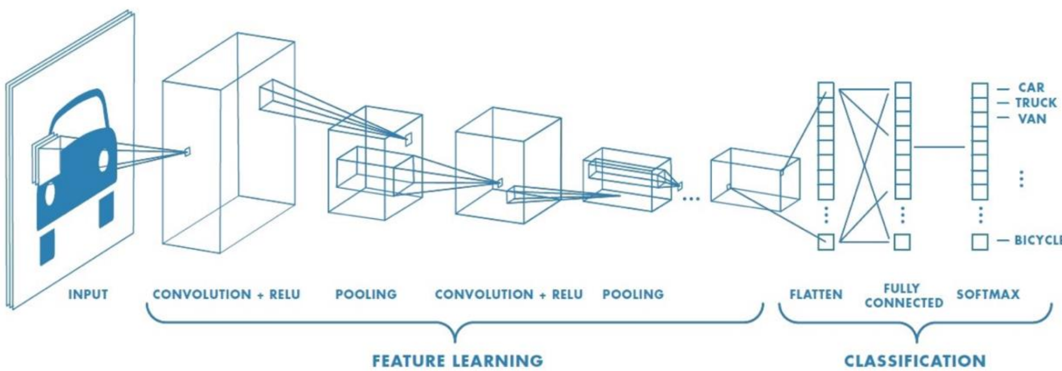
The basic functionality of the instance CNN above are often weakened into four key areas.

1. As found in other sorts of ANN, the input layer will hold the pixel values of the image.

2. The convolutional layer will determine the output of neurons of which are connected to local regions of the input through the calculation of the scalar product between their regions and the weights connected to the input volume. The rectified linear unit (commonly shortened to ReLu) aims to apply an 'elementwise' activation function such as sigmoid to the output of the activation produced by the previous layer.

3. The pooling layer will then simply perform down sampling along the spatial dimensionality of the given input, further lessening the number of parameters within that activation.

4. The fully-connected layers will then perform the same duties found in standard Artificial Neural Networks and decide to produce class scores from the activations, required for classification. It is also suggested that ReLu could also be used between these layers, on improving performance.



1.4.9 Feature Learning and Classification in CNN

As compared to RNN one should be in right mind that time dependency content could get nullified within the convolution and pooling filters, so CNN won't always be appropriate for statistic problem. The advantage though, is that CNN is comparatively easy to coach as compared to a RNN

1.5 Scope of the project:

Detecting the appearance of preictal state, the state before the seizure begins using preprocessing, feature extraction and neural network model.
These models approach high-level representation and efficiently distinguish between normal and epileptic actions of the brain with robustness and accuracy.

Chapter 2

REVIEW OF LITERATURE

Paper 1:
Algorithm: Random Forest, WPF used for feature extraction

WPFs

The electroencephalogram signal measures the mind wave designs. It has been demonstrated to be viable in following issues related telectrical action of the cerebrum, for example, analyzing with checking the seizure issue along with the rest issue. In the previous experiments, serious execution has been accomplished for epilepsy seizure grouping dependent on EEG. However, little consideration has been given to the preictal state order, also which is extremely useful for the seizure forecast and along these lines is essentially progressively significant in forestalling the coincidental losses brought about by the seizure. A precise forecast of the seizure time is testing. Rather, on the off chance that we convenient recognize and group the preictal express, the seizure can be anticipated beforehand. In this paper, we isolate the 1 hour EEG sign of the preictal state into 3 sequential and non-covered fragments, comparing it to the three preictal sub-states namely the Pre-A, Pre-B, and Pre-C states, separately. The particular time spans for the Pre-A, Pre-B, and Pre-C states are 0∼20, 20∼40, 40∼60 minutes before the seizure. A five-class characterization issue for the three preictal sub-states in addition to the seizure and the interictal states is then defined. At that point, the seizure is anticipated by arranging which expresses an electroencephalogram signal has a place with.

WPD

It is also called the optimal sub-band tree organizing, applies the wavelet change on the high pass filtering results along with the estimation attributes which are the low pass filtering results coming from the past level to accomplish an ideal portrayal of the signal. In the DWT, each level is determined by passing just the past wavelet guess attributes through discrete-time low and high pass quadrature reflect channels. For n levels of disintegration, the WPD produces 2n various arrangements of coefficients (or hubs) instead of (3n + 1) sets for the DWT. Be that as it may, because of the downsampling procedure the general number of coefficients is as yet the equivalent and there is no repetition. There were applicable examinations in signal handling and interchanges fields to address the choice of subband trees (symmetrical premise) of different sorts, for example normal, dyadic, unpredictable, as for execution measurements of interests including vitality compaction, subband connections, and others. The discrete wavelet change hypothesis offers an estimation to change discrete (examined) signals. Interestingly, the discrete subband change hypothesis gives an ideal portrayal of discrete signals. Wavelet bundles are utilized in preclinical determination.

Random Forest

Random forest, similar to its name infers, comprises of countless individual choice trees that work as an outfit. Every individual tree in the forest lets out a class expectation and the class with the most agreement turns into our model's prediction. The crucial idea driving arbitrary woods is a basic however amazing one — the shrewdness of groups. In data science talk, the explanation that the arbitrary timberland model works so well is: An enormous number of generally uncorrelated models (trees) working as an advisory group will beat any of the individual constituent models. The low connection between's models is the key. Much the same as how speculations with low relationships (like stocks and securities) meet up to shape a portfolio that is more prominent than the aggregate of its parts, uncorrelated models can create gathering expectations that are more precise than any of the individual forecasts. The purpose behind this superb impact is that the trees shield each other from their individual blunders (as long as they don't continue all fail a similar way). While a few trees might not be right, numerous different trees will be correct, so as a gathering the trees can move in the right course. So the requirements for random forests to perform well are: There should be some real sign in our highlights so models constructed utilizing those highlights show improvement over arbitrary guessing. The forecasts (and in this manner the blunders) made by the individual trees need to have low connections with one another

Results of Random Forest

The characterization rate for the most part increments with the quantity of trees and the best execution is accomplished at 1200 trees with a normal arrangement precision of 85.2%.

Paper 2:
Algorithm: Sparse Classifier with Window Based Feature Extraction

A Sparse Classified(SC) is being used where the order of the preictal and interictal signals. A sparse portrayal classifier by and large accept that a test can be composed as a straight blend of preparing tests and the weights given to every one of the preparation tests fluctuates relying upon which class the test has a place with. The relating coefficient network will generally have inadequate qualities and it will have non zero qualities just comparing to that preparation test which has a place with a similar unit as which the given test as has. Consider Vtest as a test, V as a unit of preparing units, Coefficient matrix can be represented by α. In basic terms, we also can compose the preparation test to be a direct mix of the tests as,

Vtest = αV (1)

The error value,

|| Vtest – αV|| = E (2)

This looks like a development trouble wherein within this assignment is to discover a coefficient matrix α which outputs the least errors(E). This is the outcome when it is miles conveyed as a LC(linear classifier). In case of linearly non-separable records, the linear sparse classifier would result in blunder. In such situations, a sparse kernel feature is used to map the records into better dimensions in such a way that in this better dimension space, the data could be mentioned as separable data linearly. The kernel characteristic is applied onto the set of schooling samples in addition to the test samples and then, it's far reached within the exact way as within the case of case which is linearly separable. A sparse classifier contains non-linear facts

Results

This algorithm classifies the test data with the sensitivity of 86%.

Paper 3:
Algorithm & Feature extraction: CNN Used for both

Convolutional Neural Network(CNN)

A CNN is a sort of engineered neural system utilized in picture notoriety and preparing that is uniquely intended to way pixel information.
CNNs are ground-breaking photo preparing, engineered insight (AI) that utilizes profound acing to play out each generative and enlightening undertakings, routinely utilizing machine vison that incorporates picture and video notoriety, related to recommender frameworks and natural language processing (NLP).
A neural system is an arrangement of equipment as well as programming program designed after the activity of neurons inside the human mind. Conventional neural systems aren't ideal for picture handling and must be taken care of pictures in diminished goals pieces. CNN have their "neurons" composed progressively like those of the frontal flap, the territory chargeable for preparing visual boosts in individuals and various creatures. The layers of neurons are organized in such a way as to cover the whole sight see battling off the piecemeal picture preparing issue of conventional neural. A Convolutional Neural Network utilizes a contraption like a multi-layer approach which has been calculated for abated handling prerequisites. The layers of a Convolutional Neural Network have the information layer, a yield layer and a shrouded layer which then contains various convolutional layers, pooling layers, totally associated layers and standardization layers. The evacuation of conditions and addition in execution for photograph preparing outcomes in a gadget that is far more prominent powerful, less complex to trains kept for picture handling and characteristic language handling

Feature Extraction in CNN:

Convolutional Neural Network have exhibited impressive accomplishment because of their capacity to display neighborhood conditions in the info & diminish the quantity of prepared criterion in a neural system over weight sharing. We influence two of these attributes on the wavelet-changed EEG. We utilize the convolutional channels to understand highlights that catch the transient fleeting conditions of the EEG, and simultaneously search for connections between close frequency bands. We will quickly portray a few highlights of the CNN depth, that we exploit here. Notwithstanding learning channel maps, CNN likewise include two other normal segments; max-pooling and dropout. The maximum pooling technique permits the system to learn highlights which are transiently constant. The system after which would be able to distinguish designs in the coefficients of the wavelet tensor without thinking about whether the example happens to start with or end of the sign cut. Dropout haphazardly sets the yield of units in the system to zero during preparing, keeping these from influencing the yield or the inclination of the misfortune work for an update step. This fills in as a regularization method to improve speculation by guaranteeing the system doesn't overfit by relying upon explicit concealed units. The Convolutional Neural

Network prepared had 6 convolutional layers after which were two thick layers.
The output layer comprised of three units with a soft-max enactment speaking to a probability dissemination over the three classes. We showed up at this design subsequent to trying different things with a wide range of structures, both shallower and more extensive. More profound models demonstrated hard to prepare on the accessible equipment, although shallower setups experienced bad precision. The conv layers are deformed consistently, diminishing the quantity of attributes in individual layer. Filter dimensions were set at 3x3 and 2x2 as analyses along with bigger channels indicated no advancement. Maxpooling layers were planted after each other convolutional layer and dropout included after each layer. Rather than the commonplace sigmoid nonlinear initiation, the newly created - linear layer is utilized for its non-immersing effects which do hinder the evaporating gradient issue

Results:

The outcomes on the EEG database of 199 chronicles show that (i) the preictal stage progress happens about 12 minutes before seizure beginning, (ii) forecast outcomes on the test set are assuring, along with an affectability of 90% plus a lower bogus expectation pace of 0.122 FP/h

Chapter 3

SYSTEM ANALYSIS

3.1. Functional Requirements

3.1.1 Business Rules

The users must be able to use the product under proper licensed time durations only.

3.1.2. Authentication

Only an authorized neurologist must get an access to the system, thus proper authentication measures must be applied**.**

3.1.3. Audit Tracking

Proper logging mechanisms for convenient report generations and auditing should be established by the product.

3.1.4. Ability to classify

The algorithm must be able to classify the patients with appropriate preictal/non-preictal states as accurately as possible.

3.1.5 Ability to Continuously Learn

The algorithm must be able to learn from the historic data and keep on improving its accuracy.

3.2. Non Functional Requirements

3.2.1. Performance

In our system, performance is given prime importance. Our system is based on real time event and the classification of data must be fast. The system should be able to handle multiple incoming data. Under these conditions, the performance should not hinder. Thus, the system should work under any environment and type of adversarial poisoning.

3.2.2. Accuracy

Accuracy is the most important part in our project. As the real time implementations of these problems are critical, the system should generate accurate outputs to measure accuracy, along with confusion matrix we will be using other performance measures that include precision recall.

3.2.3. Reliability

The system should be reliable so that the chances of getting an intruder into the system becomes less. Since every data in our system is unique, the number of false negative should be less and also handle concept drift in the data with increase in time.

3.2.4. Recoverability

If the system fails, the system should be immediately restarted and should be connected to the database. The database should keep log of classified data and if new data is entered then system should classify it as early as possible.

3.3. Specific Requirements

3.3.1. Hardware Requirements

16GB RAM (8GB works as well but not for the performance you may want and or expect)

Quad core Intel Core i7 Skylake or higher (Dual core is not the best for this kind of work, but manageable).

M.2 PCIe or regular PCIe SSD with at minimum 256GBs of storage, although having 512GB gives you the best performance. The performance of the system is directly proportional to the loading and saving of your application. (SATA III might hamper the system's performance)

Premium graphics cards, so things with GTX 980 or 980Ms would be the best for a laptop, and 1080s or 1070s would be the best for the desktop setup. (try not to sacrifice too much here. While a 980TI or a 970m may be cheaper, this as well is a critical part of the system, others you will notice a drastic drop in performance).

3.3.2. Software Requirements

Front End:- HTML, CSS, Bootstrap

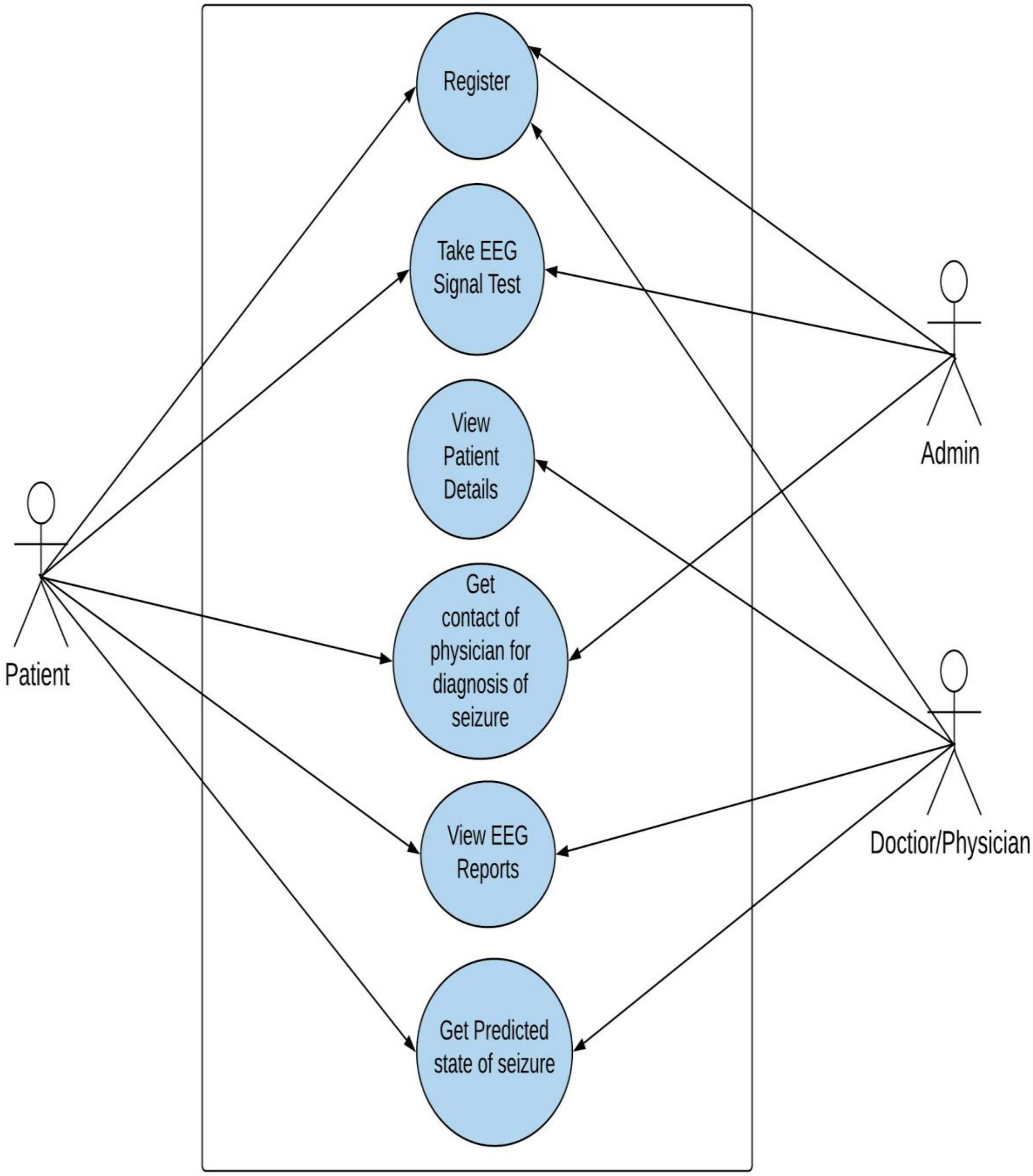
Flask

Programming Language:- Python

Keras - Tensorflow

PyEEG Module

3.4. Use-Case Diagrams and description

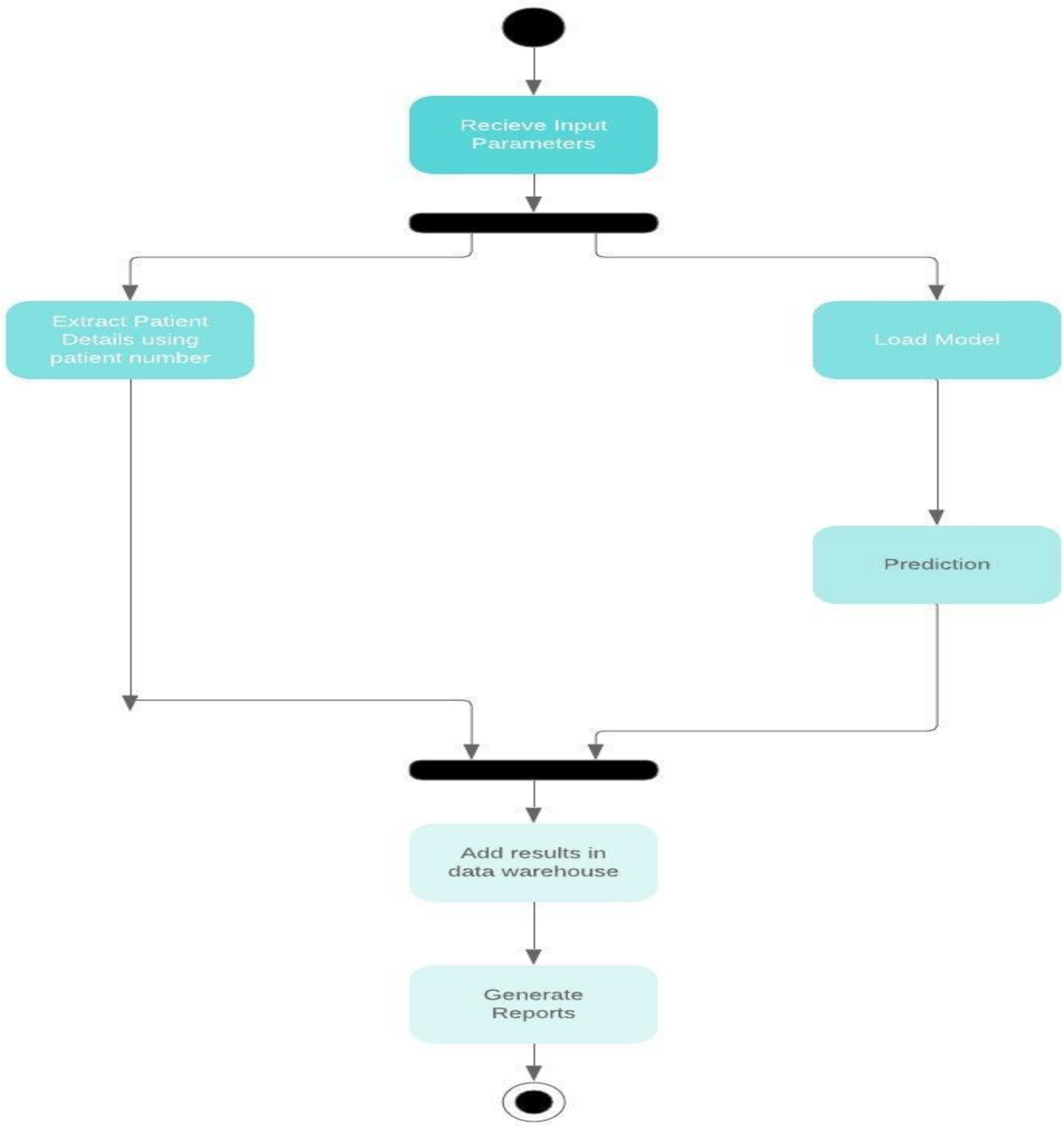


3.4 Use Case Diagram

Chapter 4

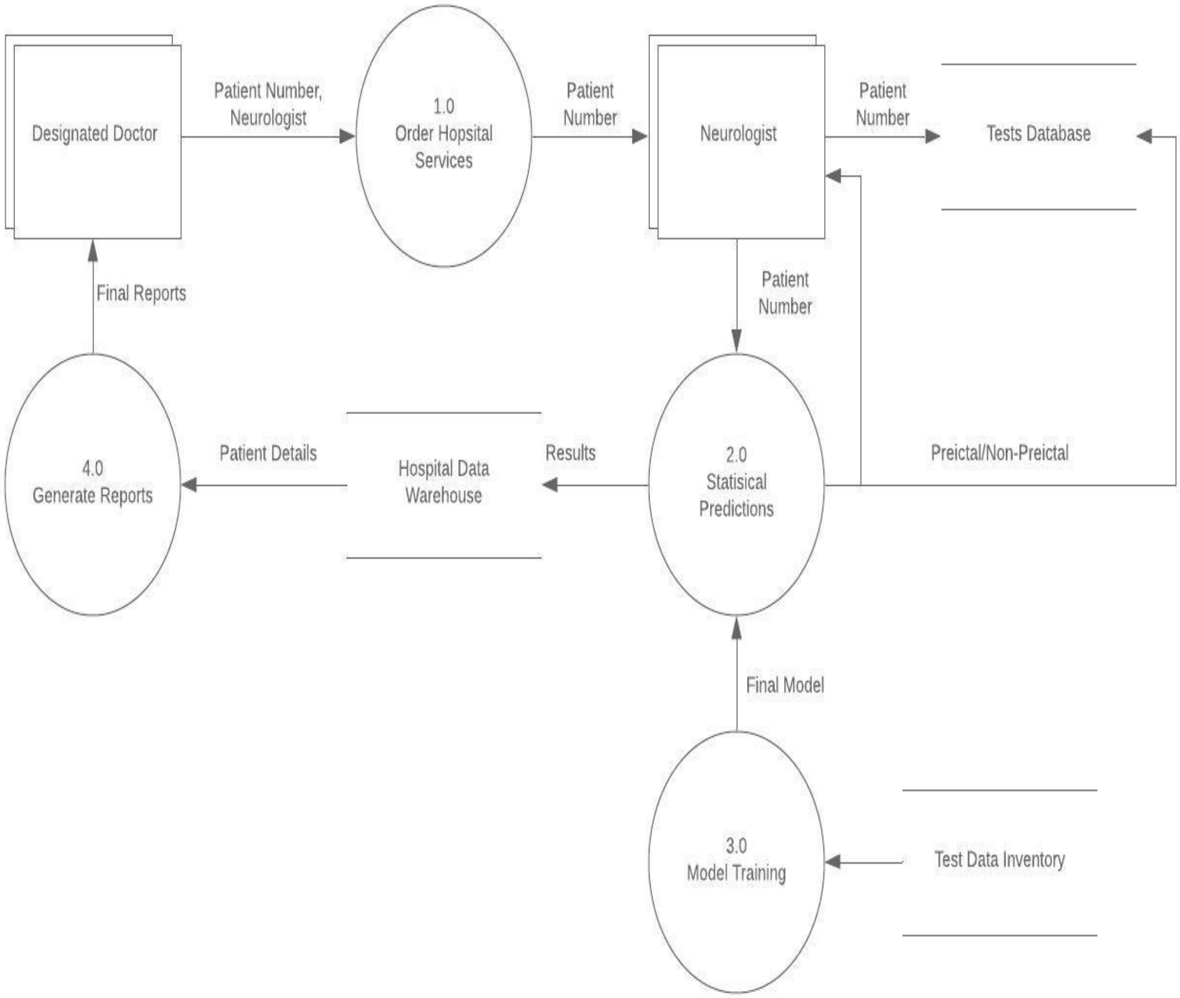
ANALYSIS MODELING

4.1 Activity Diagrams / Class Diagram



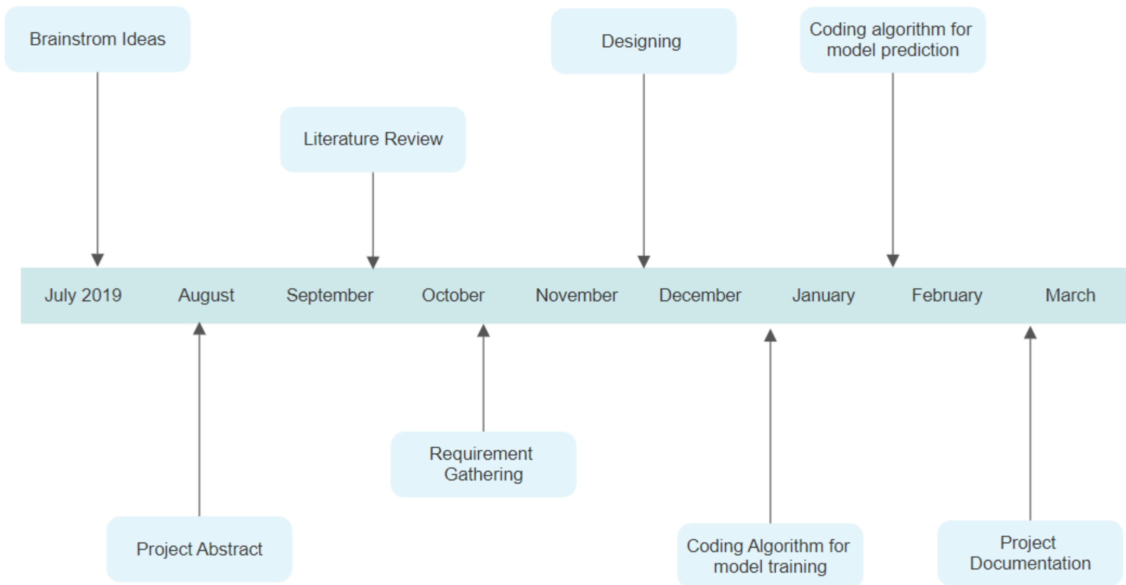
4.1 Activity Diagram/Class Diagram

4.2 Functional Modeling



4.2 Functional Modeling

4.3 Timeline Chart

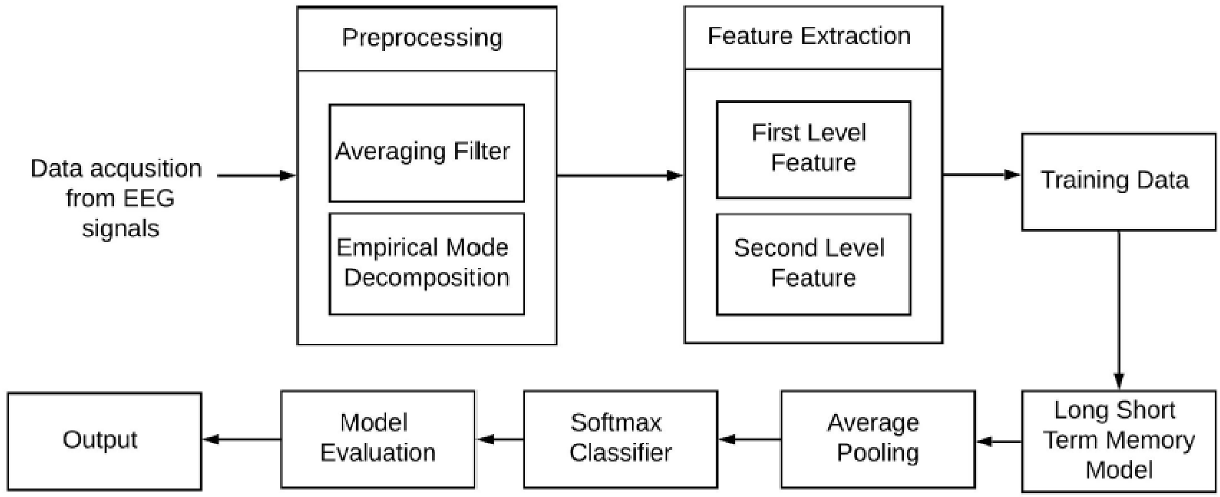


4.3 Timeline Chart

Chapter 5

DESIGN

5.1 Architectural Design



5.1 Architectural Design

Chapter 6

Implementation

6.1 Exploratory Data Analysis

The Dataset here corresponds to what looks as a set of 11500 patients. Each data point in the dataset for one particular patient corresponds to a time value which was recorded at the time. Each patient corresponds to 178 distinct values of point in time at the time of recording or curating of the dataset.

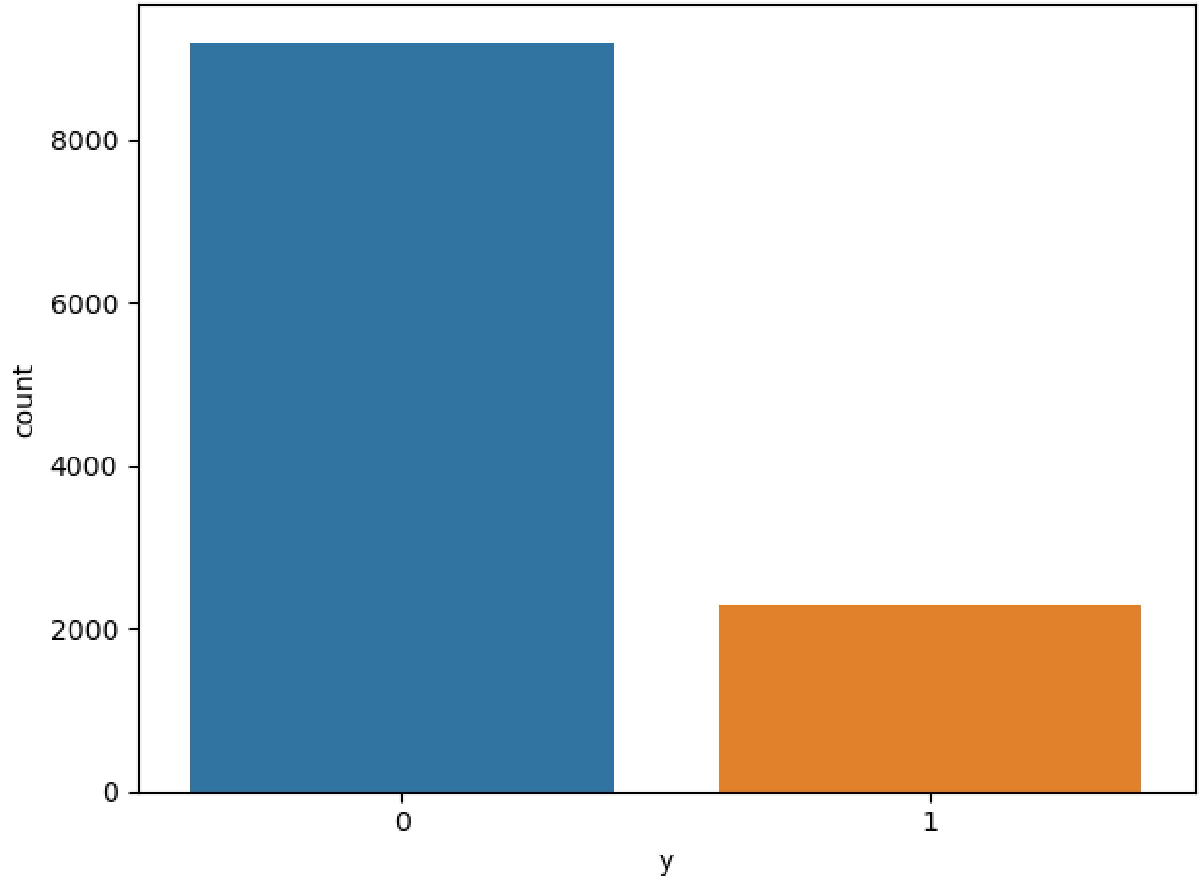
The last column in the dataset is the labelled prediction of one patient. It originally revolves around 5 distinct labels i.e. 1, 2, 3, 4, 5. But since we're making a binary classification of the dataset we will have to truncate it down to 2 labels namely 1.0.

```
F:\capstone\venv\Scripts\python.exe F:/capstone/EDA.py
      Unnamed   X1   X2   X3   X4   X5  ...  X174  X175  X176  X177  X178  y
0  X21.V1.791  135  190  229  223  192  ...  -103  -127  -116   -83   -51  4
1  X15.V1.924  386  382  356  331  320  ...   157   156   154   143   129  1
2     X8.V1.1  -32  -39  -47  -37  -32  ...   -12   -30   -35   -35   -36  5
3   X16.V1.60 -105 -101  -96  -92  -89  ...   -85   -77   -72   -69   -65  5
4   X20.V1.54   -9  -65  -98 -102  -78  ...   -41   -65   -83   -89   -73  5

[5 rows x 180 columns]
```

6.1.1 EDA

As we can see that we have 178 eeg features with 5 possible classes. The main aim of the project here is to classify the correctly whether an eeg data provided has a possibility of seizure or not. Thus this problem deals with the binary classification.



6.1.2 data distribution of data among the two classes

The following graph shows the data distribution of data among the two classes which would be used for binary classification later in the project.

6.1. Preprocessing of data

1. Null Value Analysis

```
X163      0
X164      0
X165      0
X166      0
X167      0
X168      0
X169      0
X170      0
X171      0
X172      0
X173      0
X174      0
X175      0
X176      0
X177      0
X178      0
y         0
Length: 180, dtype: int64
```

6.1.3 Null value Analysis

By the figure above it’s completely evident that there are 180 null values in our sample of 11500 tuples with 178 columns.

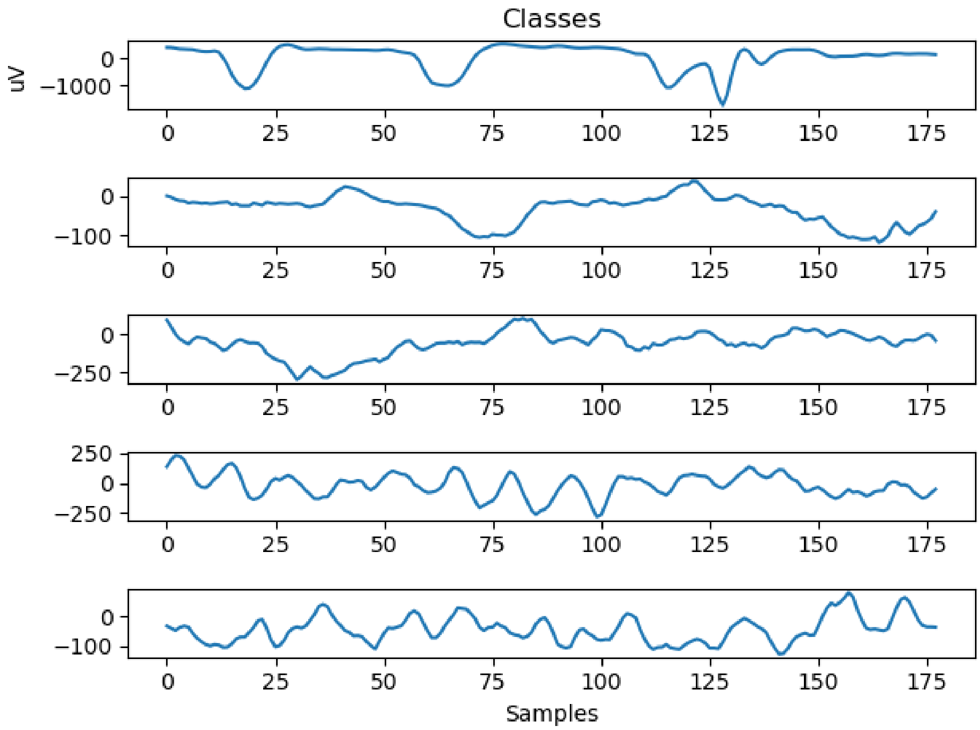
2. General Arithmetic Analysis

```
                 X1            X2  ...          X178             y
count  11500.000000  11500.000000  ...  11500.000000  11500.000000
mean     -11.581391    -10.911565  ...    -12.195652      0.200000
std      165.626284    166.059609  ...    164.852015      0.400017
min    -1839.000000  -1838.000000  ...  -1829.000000      0.000000
25%      -54.000000    -55.000000  ...    -55.000000      0.000000
50%       -8.000000     -8.000000  ...     -9.000000      0.000000
75%       34.000000     35.000000  ...     34.000000      0.000000
max     1726.000000   1713.000000  ...   1915.000000      1.000000

[8 rows x 179 columns]
```

6.1.4 General Arithmetic Analysis

3. Plotting outputs across inputs



6.1.5 Plotting Outputs across Inputs

Classes of data here are categorically placed across Y axis as samples of time within a given range.

6.2. CNN Model decisions

1. Gaussian Noise was added to improve generalization error which could be problem if not managed well. Adding noise for augmentation can prevent overfitting.
2. Batch Normalization and Dropout were used to decrease the overfitting of the model.
3. Adam optimizer was used instead of the classic stochastic gradient descent as it was especially designed for updating weights of deep neural networks. It is definitely more efficient.
4. The number of hidden convolution layers decrease gradually as the data passed across the network continually complexes. The number of neurons in the first layer is 24, the second has 16 and the third has 8.
5. Dense layers were used to make a robust fully densely connected neural network

6.3. LSTM Model Decisions

1. Gaussian Noise was added to improve generalization error which could be problem if not managed well. Adding noise for augmentation can prevent overfitting.
2. Batch Normalization and Dropout were used to decrease the overfitting of the model.
3. 60 neurons of lstm was used as it was 2/3rd of the number of input features.
4. Adam optimizer was used instead of the classic stochastic gradient descent as it was especially designed for updating weights of deep neural networks. It is definitely more efficient.
5. Dense layers were used to make a robust fully densely connected neural network.

Chapter 7

Testing

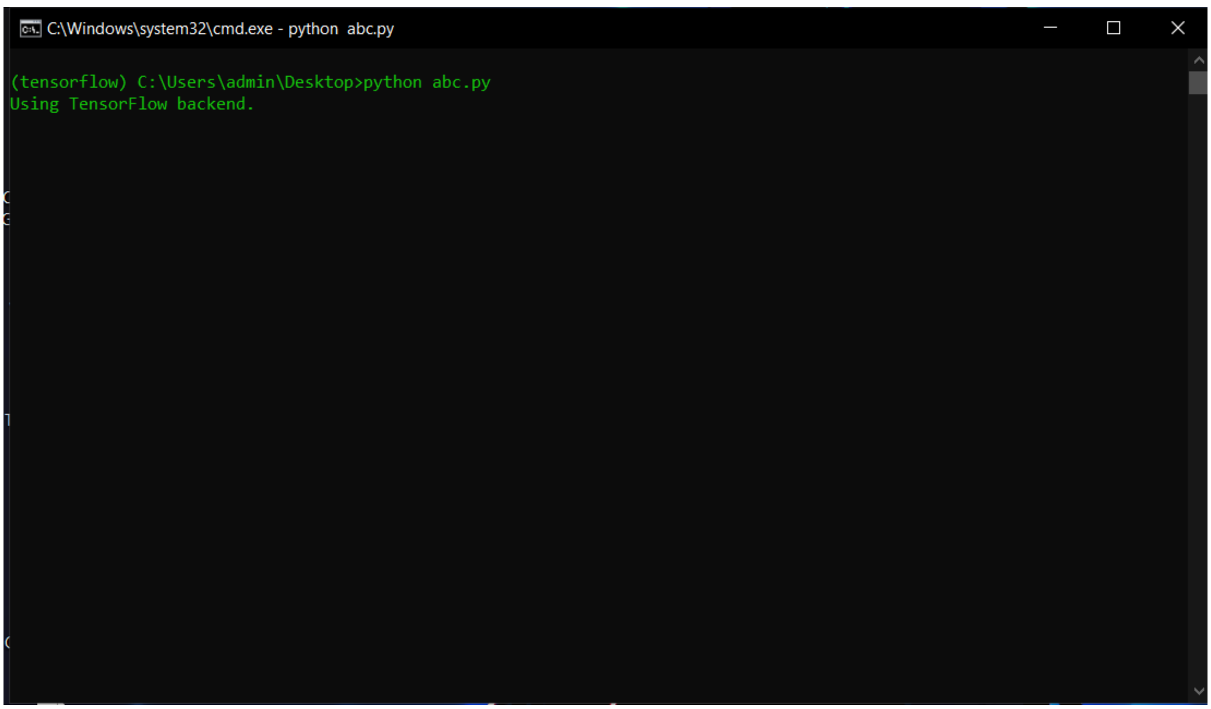
7.1 Test Cases

All the numbers between -150 to 150 were used to create a randomized patients EEG readings as a sample to feed the model for predicting whether the person with the given eeg is most likely to have a seizure or not.

After a rigorous trial and error we came to finalize this range as it gave the most diverse results which means the inputs are in moderation and not in the extreme ends of the input spectrum.

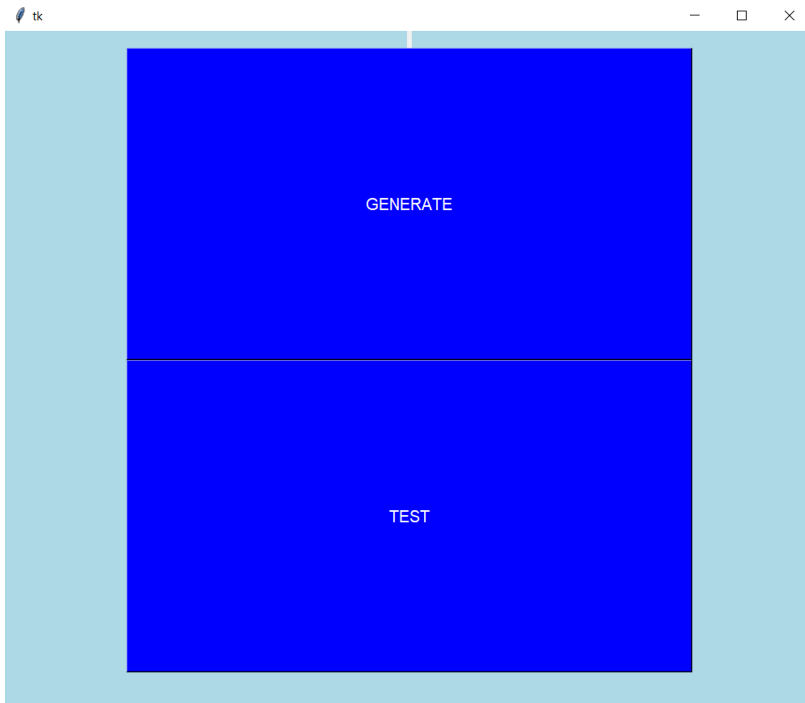
7.2. Black Box Testing

1. Open your terminal from the anaconda dialog box



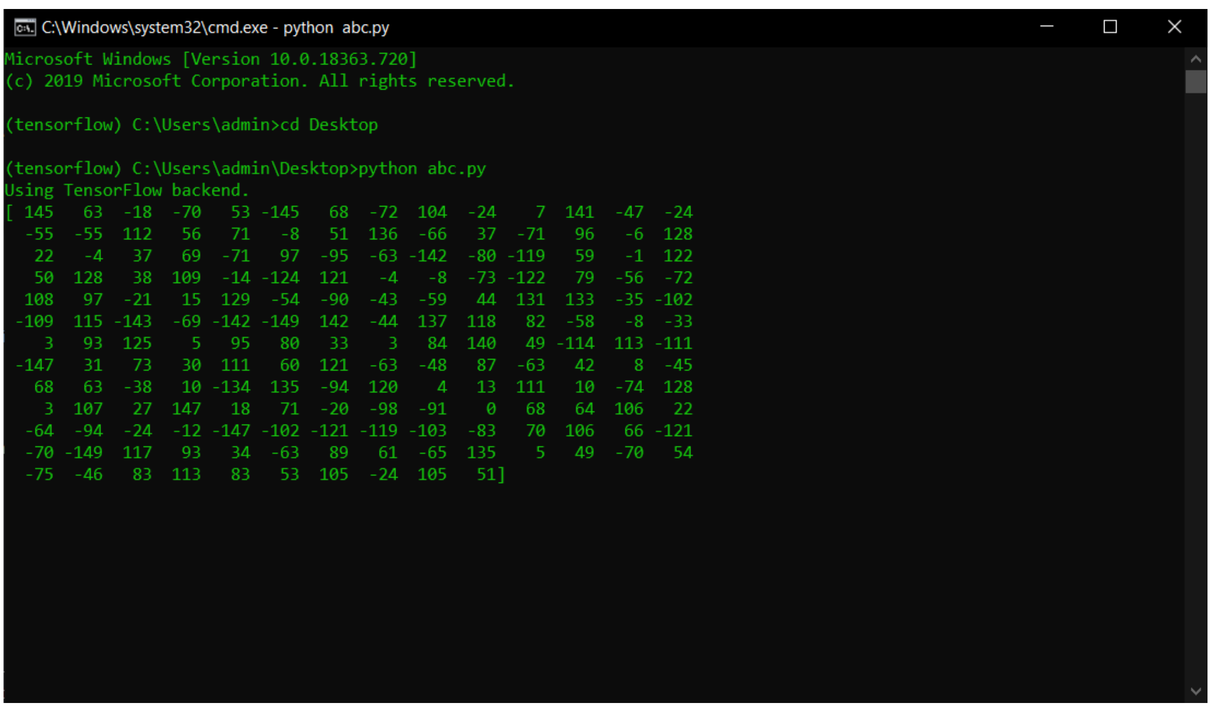
7.2.1 Open terminal

2. Generate a patient



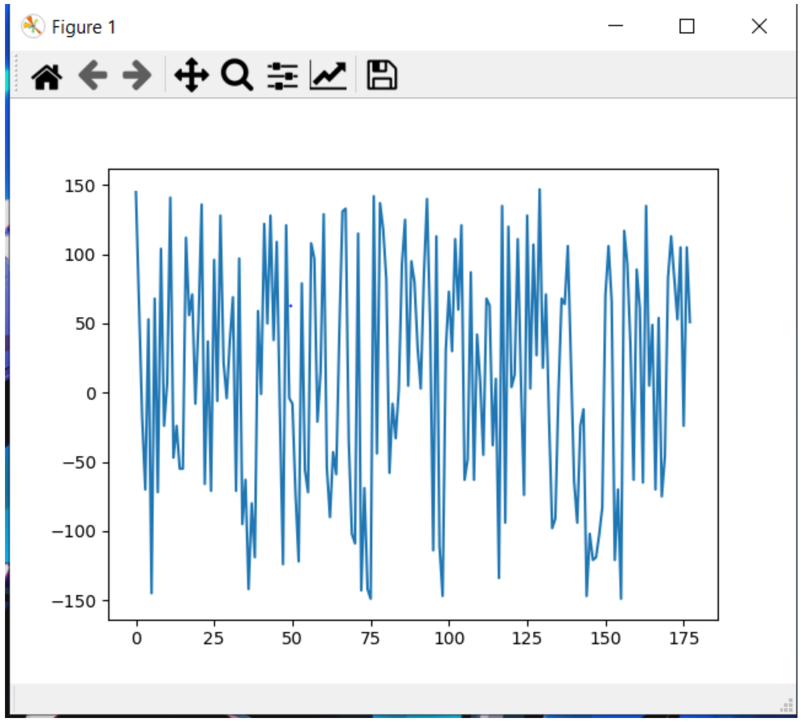
7.2.2 Generate a patient

3. You could see the matrix representation of the wave image



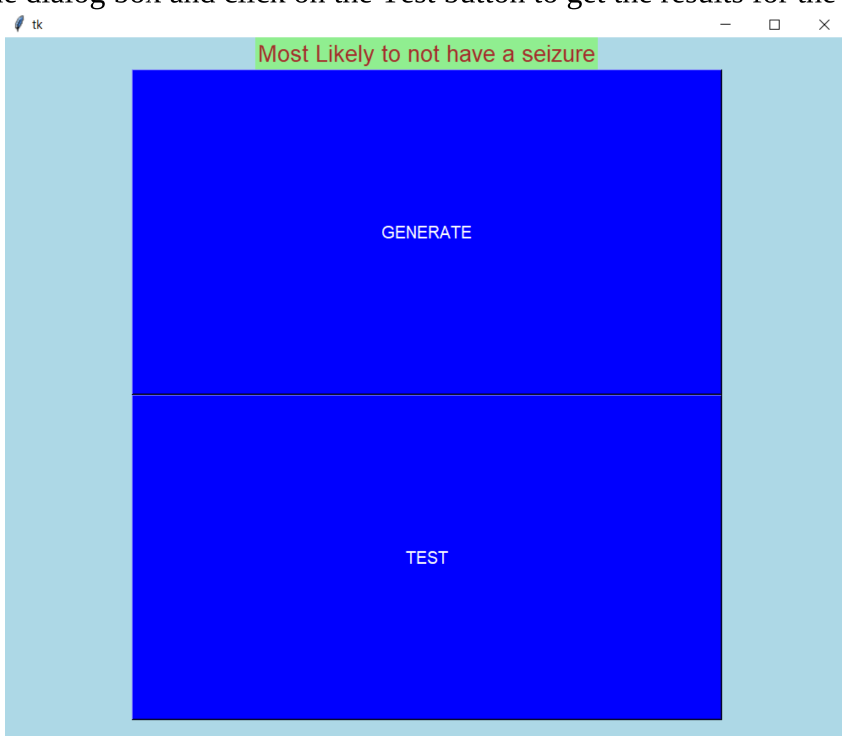
7.2.3 Matrix representation

4. A second window reopens which has the wave form depiction



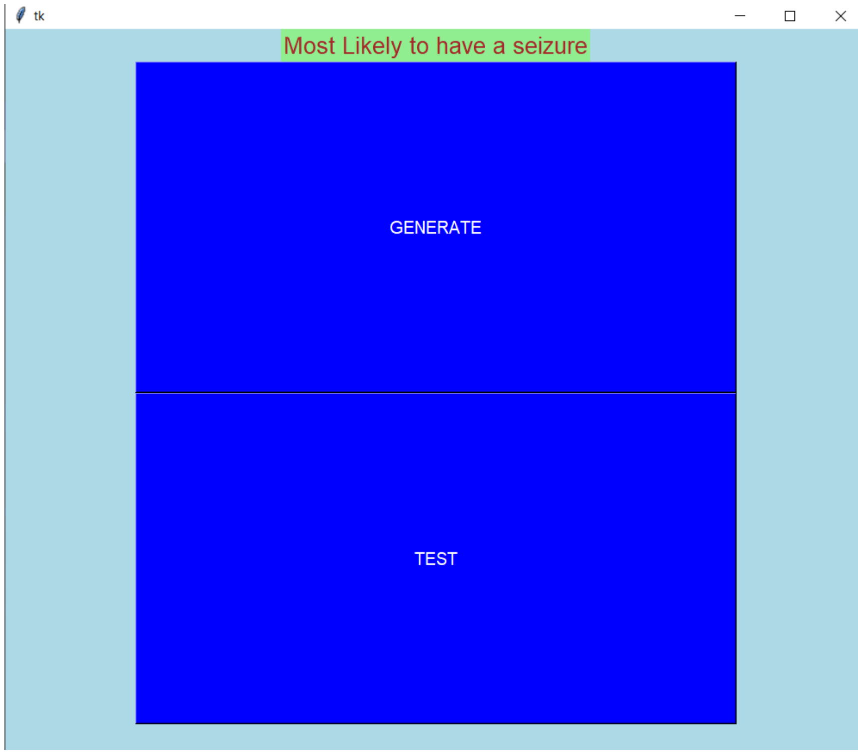
7.2.4 Graphical representation

5. Close the dialog box and click on the Test button to get the results for the given patient.



7.2.5 Results after clicking on test

OR



7.2.6 Results after clicking on test

Chapter 8

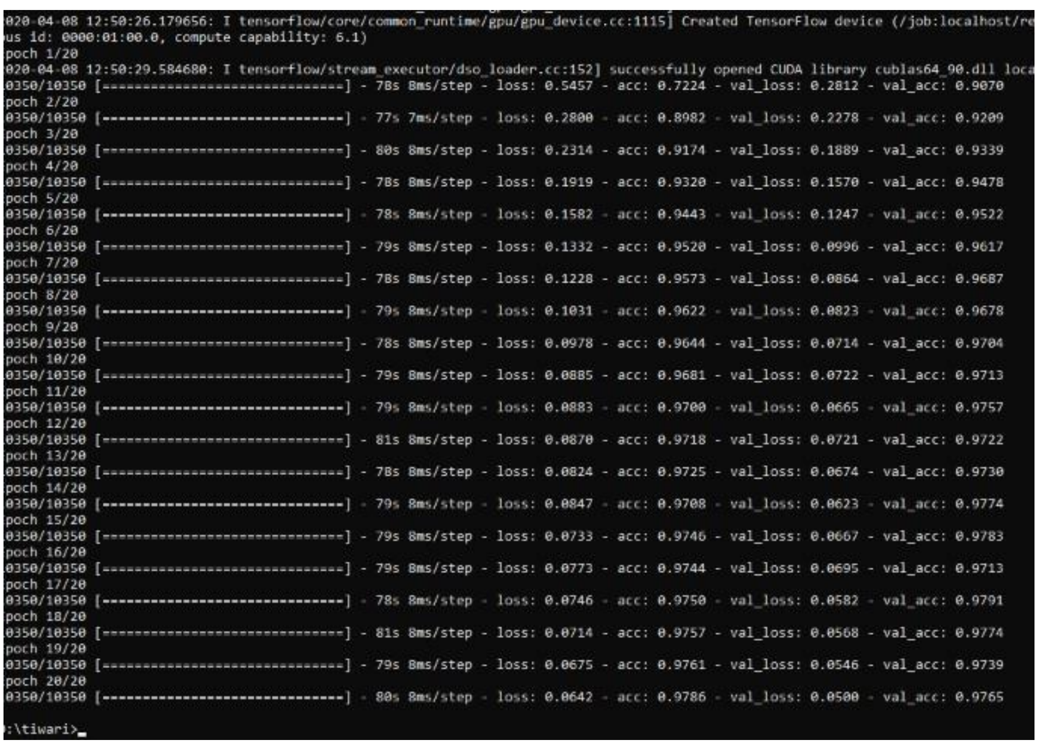
Results and Discussions

We were aiming for accuracies as good as possible. When we ran the training we were elated to see 90+ accuracies in both our approaches. It's worthwhile to state that both the accuracies were achieved in fewer epochs, which in turn means that our models in both the approaches i.e. CNN and LSTM were strong enough to fetch us these accuracies.

LSTM

As we can see, the lstm model fetched us an accuracy of 97% in just 20 epochs.

Loss while training	0.079
Accuracy	0.97
F1 Score	0.97
Recall	0.95
Precision	0.98

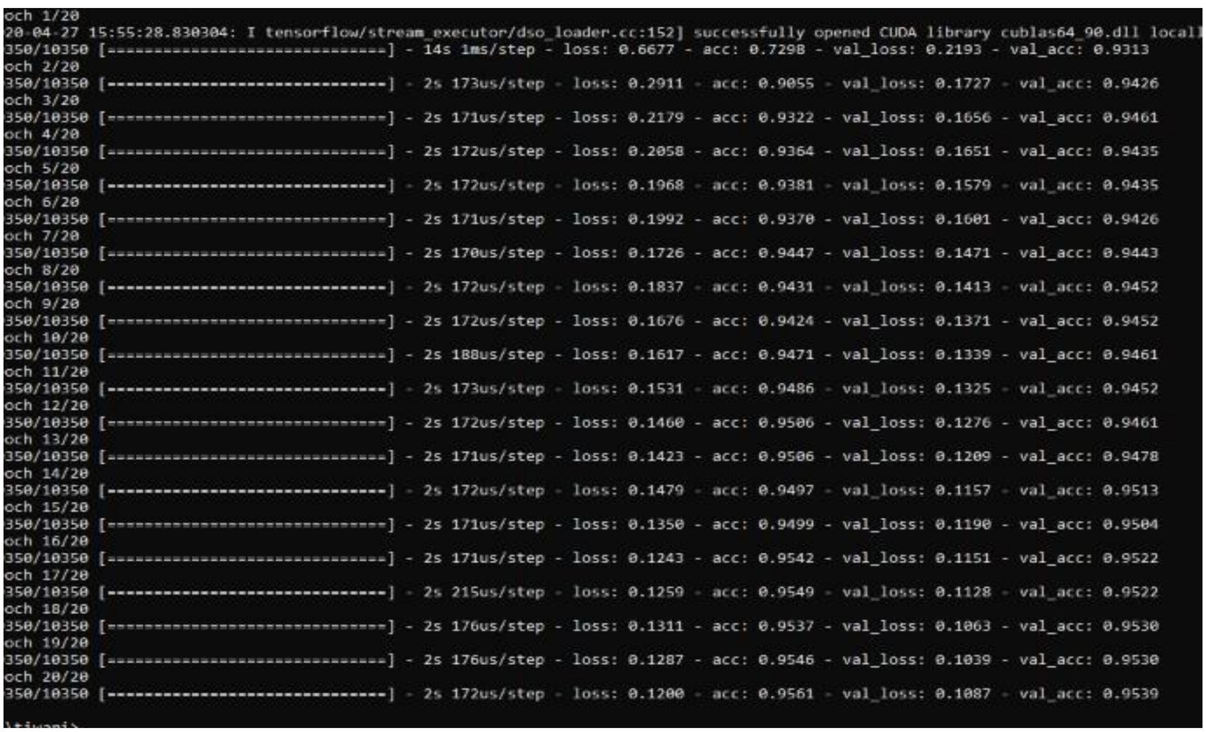


8.1 Accuracy for LSTM

Validation Loss	0.08
Accuracy	0.95
Precision	0.95
F1 Score	0.95
Recall	0.95

CNN

We can see that we achieved 95% accuracy with the given dataset and layer configurations and in 20 epochs.



```
och 1/20
20-04-27 15:55:28.830304: I tensorflow/stream_executor/dso_loader.cc:152] successfully opened CUDA library cublas64_90.dll locall
350/10350 [==============================] - 14s 1ms/step - loss: 0.6677 - acc: 0.7298 - val_loss: 0.2193 - val_acc: 0.9313
och 2/20
350/10350 [==============================] - 2s 173us/step - loss: 0.2911 - acc: 0.9055 - val_loss: 0.1727 - val_acc: 0.9426
och 3/20
350/10350 [==============================] - 2s 171us/step - loss: 0.2179 - acc: 0.9322 - val_loss: 0.1656 - val_acc: 0.9461
och 4/20
350/10350 [==============================] - 2s 172us/step - loss: 0.2058 - acc: 0.9364 - val_loss: 0.1651 - val_acc: 0.9435
och 5/20
350/10350 [==============================] - 2s 172us/step - loss: 0.1968 - acc: 0.9381 - val_loss: 0.1579 - val_acc: 0.9435
och 6/20
350/10350 [==============================] - 2s 171us/step - loss: 0.1992 - acc: 0.9370 - val_loss: 0.1601 - val_acc: 0.9426
och 7/20
350/10350 [==============================] - 2s 170us/step - loss: 0.1726 - acc: 0.9447 - val_loss: 0.1471 - val_acc: 0.9443
och 8/20
350/10350 [==============================] - 2s 172us/step - loss: 0.1837 - acc: 0.9431 - val_loss: 0.1413 - val_acc: 0.9452
och 9/20
350/10350 [==============================] - 2s 172us/step - loss: 0.1676 - acc: 0.9424 - val_loss: 0.1371 - val_acc: 0.9452
och 10/20
350/10350 [==============================] - 2s 188us/step - loss: 0.1617 - acc: 0.9471 - val_loss: 0.1339 - val_acc: 0.9461
och 11/20
350/10350 [==============================] - 2s 173us/step - loss: 0.1531 - acc: 0.9486 - val_loss: 0.1325 - val_acc: 0.9452
och 12/20
350/10350 [==============================] - 2s 172us/step - loss: 0.1460 - acc: 0.9506 - val_loss: 0.1276 - val_acc: 0.9461
och 13/20
350/10350 [==============================] - 2s 171us/step - loss: 0.1423 - acc: 0.9506 - val_loss: 0.1209 - val_acc: 0.9478
och 14/20
350/10350 [==============================] - 2s 172us/step - loss: 0.1479 - acc: 0.9497 - val_loss: 0.1157 - val_acc: 0.9513
och 15/20
350/10350 [==============================] - 2s 171us/step - loss: 0.1350 - acc: 0.9499 - val_loss: 0.1190 - val_acc: 0.9504
och 16/20
350/10350 [==============================] - 2s 171us/step - loss: 0.1243 - acc: 0.9542 - val_loss: 0.1151 - val_acc: 0.9522
och 17/20
350/10350 [==============================] - 2s 215us/step - loss: 0.1259 - acc: 0.9549 - val_loss: 0.1128 - val_acc: 0.9522
och 18/20
350/10350 [==============================] - 2s 176us/step - loss: 0.1311 - acc: 0.9537 - val_loss: 0.1063 - val_acc: 0.9530
och 19/20
350/10350 [==============================] - 2s 176us/step - loss: 0.1287 - acc: 0.9546 - val_loss: 0.1039 - val_acc: 0.9530
och 20/20
350/10350 [==============================] - 2s 172us/step - loss: 0.1200 - acc: 0.9561 - val_loss: 0.1087 - val_acc: 0.9539
```

8.12 Accuracy for CNN

Accuracy: - Accuracy is nothing but the ratio of true positives with the total size of the dataset. Accuracy is the most used factor in deep learning when we decide to optimize the model based on some metric

Accuracy = TP+TN/TP+FP+FN+TN

Precision:- Precision is the ratio of true positives with the ratio of total number of positives which include true positives and false positives. This metric gives us the true positives out of a total of positives out of which some could be faulty.

Precision = TP/TP+FP

Recall: - Recall is also known as sensitivity. Recall is the ratio of true positives out of all the given data. Recall here does the exactly opposite of what precision does. It just gives you the false positives + true positives. This is not the best measure to optimize your algorithm with.

Recall = TP/TP+FN

F1 Score: - F1 Score is the weighted average of Accuracy and Precision. F1 Score is not as easy to grasp as Precision and Accuracy are. But this could be an accurate metric to optimize your algorithm with because it takes into account both false positives and false negatives. Let alone it’s not the metric to completely rely on.

F1 Score = 2 x (Recall x Precision) / Recall + Precision

Chapter 9

Conclusions

We introduced a deep learning approach for the automatic detection of epileptic seizures using EEG signals. Compared to the state-of-the-art methods, this approach can learn the high-level representations, and can effectively discriminate between the normal and seizure EEG activities. Another advantage of this approach lies in its robustness against common EEG artifacts (e.g., muscle activities and eyeblinking) and white noise. The proposed approach has been examined on the Bonn EEG dataset and compared to several baseline methods. The experimental results demonstrate the effectiveness and superiority of the proposed method in detecting epileptic seizures. It achieves the superior detection accuracies under ideal and imperfect conditions.

APPENDIX

LIST OF ABBREVIATIONS

SR NO	Abbreviation	Expanded Form
I	EEG	Electroencephalogram
II	MLP	Multilayer Perceptrons
III	CNNs	Convolutional Neural Networks
IV	RNNs	Recurrent Neural Networks
V	LSTM	Long short-term memory
VI	CNN	Convolutional Neural Network
VII	NLP	Natural Language Processing

LITERATURE CITED

.

[1]M. D'Alessandro, R. Esteller, G. Vachtsevanos, A. Hinson, J. Echauz and B. Litt, "Epileptic seizure prediction using hybrid feature selection over multiple intracranial EEG electrode contacts: a report of four patients, in" IEEE Transactions on Biomedical Engineering, vol. 50, no. 5, pp. 603-615, May 2003.

[2] L. D. Iasemidis et al., "Adaptive epileptic seizure prediction system, in" IEEE Transactions on Biomedical Engineering,vol. 50, no. 5, pp. 616-627, May 2003.

[3] S. Li, W. Zhou, Q. Yuan and Y. Liu, "Seizure Prediction Using Spike Rate of Intracranial EEG, in " IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol.21, no. 6, pp. 880-886, Nov. 2013.

[4]N. Wang and M. R. Lyu, "Extracting and Selecting Distinctive EEG Features for Efficient Epileptic Seizure Prediction, in" IEEE Journal of Biomedical and Health Informatics,vol. 19, no. 5, pp. 1648-1659, Sept. 2015.

[5] **IEEE standard Book,** J. F. Curtis, (Ed.), *Processes and Disorders of Human Communication.* New York: Harper and Row, 1978.

[6] **Journal Paper,** J. Schroeterand M. M. Sondhi, "Techniques for estimating vocal-tract shapes from the speech signal,"*IEEE Trans. Speech Audio Process.*, vol. 2, no. 1, pp. 133–150, 1994.

[7] **Proceeding paper,** J. M. Pardo, "Vocal tract shape analysis for children," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 1982, pp. 763–766.

PLAGIARISM REPORT

SmallSEOTools

PLAGIARISM SCAN REPORT

Words	897	Date	May 03,2020
Characters	5980	Exclude Url	

0% Plagiarism	100% Unique	0 Plagiarized Sentences	43 Unique Sentences

Content Checked For Plagiarism

1.1 Description Deep Lear ning (LSTM) based Seizure prediction using Scalp based EEG Readings. 1.2 PROBLEM FORMULATION: People suffering from epilepsy or also called seizure disorder, a disorder in which nerve cell activity in the brain is disturbed, causing seizures. These seizure do not have a fixed time they occur at random times causing great disturbance to emotional state and cognitive state of the mind. Their ability to work, social, economic situations comes to a sudden halt. Having the option to foresee epileptic seizures will incredibly improve the personal satisfaction of individuals with epilepsy by either giving them an admonition of an approaching seizure so they can move to security or enacting an embedded seizure control gadget that can turn away seizures through medication conveyance or electrical incitement of the cerebrum. How might we arrange the stage epileptic seizures utilizing electrical accounts of mind action? 1.3 MOTIVATION: This world consist of 70 million people having PWE i.e. epilepsy and nearly 12 million PWE patients reside in india; which amounts to about a sixth of the world. The general pervasiveness (3.5-12.1 per 1,600 populace) which has a frequency (0.3-0.9 per 1,400 populace for every year) information from ongoing examinations in India on all inclusive community are similar to the paces of high-income nations regardless of checked varieties in populace attributes & information systems. Having a differential conveyance of PWE among different socio-demographic & monetary gatherings with higher rates detailed for the male, country populace, and low financial status. A changing example in the age-explicit event of epilepsy with prevalence towards the more established age bunch is seen due to sociodemographic and epidemiological progress. Neuro-infections, neuro-cysticercosis , and neurotrauma alongside birth wounds have risen as significant hazard factors for optional PWE. In spite of its shifted etiology ,PWE are reasonable in outlook. Programmed identification of epileptic seizures can impressively improve the patients' personal satisfaction. Current Electroencephalogram (EEG)- based seizure discovery frameworks experience numerous difficulties, all things considered, circumstances. The EEGs are non-fixed signs and seizure designs fluctuate across patients and recording meetings. Besides, EEG information are inclined to various clamor types that contrarily influence the recognition precision of epileptic seizures. To address these difficulties, we present the utilization of a profound learning-based methodology that naturally learns the discriminative EEG highlights of epileptic seizures. In particular, to uncover the connection between's progressive information tests, the time-arrangement EEG information are first sectioned into a succession of nonoverlapping ages. Second, Long Short-Term Memory (LSTM) arrange is utilized to become familiar with the elevated level portrayals of the ordinary and the seizure EEG designs. Third, these portrayals are taken care of into Softmax function for training. The outcomes on a notable benchmark clinical dataset show the prevalence of the proposed approach over the current best in class strategies. Moreover, our methodology is demonstrated to be hearty in uproarious and genuine conditions. Contrasted with current techniques that are very touchy to clamor, the proposed strategy keeps up its high recognition execution within the sight of basic EEG ancient rarities (muscle exercises and eye-squinting). 1 . 4 Proposed Solution: In contrast to increasingly customary strategies for machine learning methods, deep learning classifiers are prepared through component adapting as opposed to task-explicit calculations. This means the machine would understand the patterns in the pictures as it is given instead of requiring the human administrator to characterize the examples which the machines should look for in the picture. The component learning system is utilized each day by the way we show a kid to perceive distinctive objects. Deep learning is a sort of AI that impersonates the neuron of the neural systems present in the human cerebrum. PC Vision Deep learning models are prepared on a lot of pictures a.k.a preparing information, to explain an undertaking. These deep learning models are for the most part utilized in the field of Computer Vision which permits a PC to see and picture as a human would. This kind of system is like the organic sensory system, with every hub going about as a neuron inside a bigger network. Thus, these models are a class of fake neural systems. This models calculations adapt dynamically about the picture as it experiences each neural system layer. Prior layers figure out how to recognize low-level highlights like edges, and ensuing layers consolidate highlights from prior layers into an increasingly all-encompassing And complete potryal . 1.4.1 Deep Learning vs Machine Learning Algorithms. Deep learning is a sort of AI that emulates the neuron of the neural systems which human mind has. Deep learning models are prepared considering lot of pictures for preparing information, to comprehend an assignment. These deep learning models are for the most part utilized in the sector of Computer Vision, which allows a PC to compare and envision as a real preson would. These models could be visualized as a group of points wherein each of those points makes a decision based on the nodal input. The human biological neural system as has the same type of network, wherein each node has a neural within a network which is larger than itself.Thus, these models are a class of fake neural systems. Deep learning algorithms adapt continuously about the picture as it experiences each neural system layer. Early layers figure out how to distinguish low-level highlights like edges, and resulting layers consolidate highlights from prior classes into a progressively all-encompassing and thorough portrayal.

Sources	Similarity

SmallSEOTools

PLAGIARISM SCAN REPORT

Words	972	Date	May 03,2020
Characters	6218	Exclude Url	

0% Plagiarism	100% Unique	0 Plagiarized Sentences	49 Unique Sentences

Content Checked For Plagiarism

Artificial neural systems are further classified in 3 classes. They are: • Multilayer Perceptrons (MLPs) • Convolutional Neural Networks (CNNs) • Recurrent Neural Networks (RNNs) These 3 styles of systems give a great deal of adaptability & have demonstrated themselves over many years to be helpful and dependable in a huge scope of issues. Who likewise contains numerous sub-types for helping them practice them to the characteristics of various framings of expectation issues & diversified datasets. Its common from the vibes of the dataset that going with RNNs is the best choice. We can use Recursive Neural Network for: • Text data • Speech data • Classification prediction problems • Regression prediction problems • Generative models • Time Series Data What are RNN's? The idea at the back of Recursive Neural Networks is to use sequential facts. In a conventional neural network we assume that each one inputs (and outputs) are not dependent to every other. But for few duties that's a poor idea. If someone desire to anticipate the further phrase in a sentence one should know which words came prior to it. Recursive Neural Networks are referred to as recurrent because they carry out the equal project for every element of a sequence, with the output being depended on the preceding calculations and also you already recognize that they have a "memory" which accounts the result approximately calculated till now. "Whenever there is a sequence of data and that temporal dynamics that connects the data is more important than the spatial content of each individual frame." – Lex Fridman (MIT) Various kinds of RNN's: The center explanation that recurrent network are all the more energizing is that they permit us to work over groupings of vectors: Sequences in the info, the yield, or in the most broad case both. A couple of models may make this increasingly concrete: 1.4.2 Types of RNN Various kinds of Recurrent Neural Networks. (1) Sequence yield (for example picture subtitling takes a picture and yielding few sentences consisting words). (2) Sequence input (for example sentiment examination wherein a sentence given is delegated communicating +ve or -ve assumption). (3) Sequence input-output(for example Machine Translation: a RNN peruses a sentence in English and afterwards yielding a sentence in Spanish). (4) Synced sequence input and output(for example video characterization where we wish to mark each casing of the video). Notice that for each situation are no pre-indicated limitations on the lengths groupings in light of the fact that the intermittent change (green) is fixed and can be applied the same number of times as we like. Every single rectangle in the images shown above shows us vectors and arrows representing functions. Red colour attributes input vectors while the blue ones are output vectors. The green contains the RNN's state. One-to-One(O2O): This is called as Plain-Variall Neural systems likewise. It manages constraint size of contribution to Constant size of Output wherein they are free of past data/yield. Model: Image characterization. One-to-Many(O2M): This manages constant size of data as information that outputs grouping of information just as yield. Example: The same happens in Image Captioning where it accepts picture as info and yields a sentence. Many-to-One(M21): It takes a series of data as information & yields a fixed size of yield. Example: When a given sentence is classified as expressing +ve or -ve sentiment is called sentiment analysis. Many-to-Many(M2M): It accepts a Series of data considering it information & procedure it intermittently yields a Sequence of information. Model: Machine Translation, where a RNN peruses a sentence in English and afterward yields a sentence in French. Bidirectional Many-to-Many: Consider that for each situation are no pre-indicated limitations on the lengths groupings in light of the fact that the repetitive change (green) is fixed and can be applied the same number of times as we like. Model: video characterization wherein we want to characterize individual frame consisting in the video. Data Acquisition and Preprocessing: a. Intracranial

Electroencephalograms i.e. 10-minute long recording of the brain, is collected from different 15 epileptic patients by using 16 channels surgically inculcated at different position of electrodes segmented at 400 Hz. b. Each recording signals stores either preictal or interictal state of epileptic seizure. c. Then, EEG channels are converted into single signal called as surrogate channel in order to increase the signal-noise ratio by using averaging filter and laplacian filter. Feature Extraction: a. After preprocessing of signals, features are extracted from the noise-free channels to train and fit the neural network model. b. Features are extracted in two levels: i. At the first level, the energy, spectrum, length of curve, entropy, mean, variance are extracted between the positive (preictal) and negative (interictal) samples of signals. ii. Then the extracted first level feature becomes the input to the second level, then the predicted features are calculated from each of the signal samples. iii. The prediction indicator features are mean, median, slope, derivate, entropy, sum, variance standard deviation, maximum, slope, skewness, curve length, variance. Long Short Term Memory Model: a. Segmentation b. i. The iEEG signal are further divided into smaller epochs or segments. c. ii. These epochs are converted into stationary data for better performance and evaluation. d. iii. Each of these epochs or segments are of specific lengths and performance indicator e. features are extracted from epochs of specific length L. f. b. Learning g. i. The extracted features of signals are passed to LSTM layer which contains 120 h. neurons. i. ii. The output of LSTM layer is sent to the Average Pooling layer along with the activation j. function, in order to consider the equal label prediction from all epochs. k. c. Classification l. i. The output of the average pooling layer is sent to the softmax layer. m. ii. Softmax layer gives the probability distribution of each class label such as preictal or

Sources	Similarity

SmallSEOTools

PLAGIARISM SCAN REPORT

Words	877	Date	May 03,2020
Characters	5765	Exclude Url	

0% Plagiarism	100% Unique	0 Plagiarized Sentences	39 Unique Sentences

Content Checked For Plagiarism

n. interictal state. o. p. Average Pooling – It includes finding out the mean for each patch within the characteristic map. This way every 2×2 square within the feature map is down sized to the common content inside the square. Consider an example:- the result of the line detector convolutional filter within the preceding section was a 6×6 function map q. r. Softmax Classifier - Softmax Regression (equivalent words: Multinomial Logistic, Maximum Entropy Classifier, or just Multi-class Logistic Regression) is a speculation of calculated relapse that we can use for multi-class characterization (under the presumption that the classes are fundamentally unrelated). Conversely, we utilize the (norm) Logistic Regression model in double characterization assignments. 1.5 Scope of the project: Improved preprocessing of Intracranial Electroencephalograms (iEEG) signal to receive increased sensitivity of seizure. Detecting the appearance of preictal state, the state before the seizure begins using preprocessing, feature extraction and neural network model. This model approaches high-level representation and efficiently distinguish between normal and epileptic actions of the brain with robustness and accuracy. 2. REVIEW OF LITERATURE Paper 1: Algorithm: Random Forest, WPF used for feature extraction WPFs The electroencephalogram signal measures the mind wave designs. It has been demonstrated to be viable in following issues related telectrical action of the cerebrum, for example, analyzing with checking the seizure issue along with the rest issue. In the previous experiments, serious execution has been accomplished for epilepsy seizure grouping dependent on EEG. However, little consideration has been given to the preictal state order, also which is extremely useful for the seizure forecast and along these lines is essentially progressively significant in forestalling the coincidental losses brought about by the seizure. A precise forecast of the seizure time is testing. Rather, on the off chance that we convenient recognize and group the preictal express, the seizure can be anticipated beforehand. In this paper, we isolate the 1 hour EEG sign of the preictal state into 3 sequential and non-covered fragments, comparing it to the three preictal sub-states namely the Pre-A, Pre-B, and Pre-C states, separately. The particular time spans for the Pre-A, Pre-B, and Pre-C states are 0020, 20040, 40060 minutes before the seizure. A five-class characterization issue for the three preictal sub-states in addition to the seizure and the interictal states is then defined. At that point, the seizure is anticipated by arranging which expresses an electroencephalogram signal has a place with. WPD It is also called the optimal sub-band tree organizing, applies the wavelet change on the high pass filtering results along with the estimation attributes which are the low pass filtering results coming from the past level to accomplish an ideal portrayal of the signal. In the DWT, each level is determined by passing just the past wavelet guess attributes through discrete-time low and high pass quadrature reflect channels. For n levels of disintegration, the WPD produces 2n various arrangements of coefficients (or hubs) instead of (3n + 1) sets for the DWT. Be that as it may, because of the downsampling procedure the general number of coefficients is as yet the equivalent and there is no repetition. There were applicable examinations in signal handling and interchanges fields to address the choice of subband trees (symmetrical premise) of different sorts, for example normal, dyadic, unpredictable, as for execution measurements of interests including vitality compaction, subband connections, and others. The discrete wavelet change hypothesis offers an estimation to change discrete (examined) signals. Interestingly, the discrete subband change hypothesis gives an ideal portrayal of discrete signals. Wavelet bundles are utilized in preclinical determination. Random Forest Random forest, similar to its name infers, comprises of countless individual choice trees that work as an outfit. Every individual tree in the forest lets out a class expectation and the class with the most agreement turns into our model's prediction. The crucial idea driving arbitrary woods is a basic however amazing one — the shrewdness of groups. In data science talk, the explanation that the arbitrary timberland model works so well is: An enormous number of generally uncorrelated models (trees) working as an advisory group will beat any of the individual constituent models. The low connection between's models is the key. Much the same as how speculations with low relationships (like stocks and securities) meet up to shape a portfolio that is more prominent than the aggregate of its parts, uncorrelated models can create gathering expectations that are more precise than any of the individual forecasts. The purpose behind this superb impact is that the trees shield each other from their individual blunders (as long as they don't continue all fail a similar way). While a few trees might not be right, numerous different trees will be correct, so as a gathering the trees can move in the right course. So the requirements for random forests to perform well are: There should be some real sign in our highlights so models constructed utilizing those highlights show improvement over arbitrary guessing. The forecasts (and in this manner the blunders) made by the individual trees need to have low connections with one another Results of Random Forest The characterization rate for the most part increments with the quantity of trees and the best execution is accomplished at 1200 trees with a normal arrangement precision of 85.2%.

Sources	Similarity

SmallSEOTools

PLAGIARISM SCAN REPORT

Words	942	Date	May 03,2020
Characters	6084	Exclude Url	

0% Plagiarism	100% Unique	0 Plagiarized Sentences	41 Unique Sentences

Content Checked For Plagiarism

Paper 2: Algorithm: Sparse Classifier with Window Based Feature Extraction A Sparse Classified(SC) is being used where the order of the preictal and interictal signals. A sparse portrayal classifier by and large accept that a test can be composed as a straight blend of preparing tests and the weights given to every one of the preparation tests fluctuates relying upon which class the test has a place with. The relating coefficient network will generally have inadequate qualities and it will have non zero qualities just comparing to that preparation test which has a place with a similar unit as which the given test as has. Consider Vtest as a test, V as a unit of preparing units. Coefficient matrix can be represented by α. In basic terms, we also can compose the preparation test to be a direct mix of the tests as, $V_{test} = \alpha V$ (1) The error value, $|| V_{test} - \alpha V || = E$ (2) This looks like a development trouble wherein within this assignment is to discover a coefficient matrix α which outputs the least errors(E). This is the outcome when it is miles conveyed as a LC(linear classifier). In case of linearly non-separable records, the linear sparse classifier would result in blunder. In such situations, a sparse kernel feature is used to map the records into better dimensions in such a way that in this better dimension space, the data could be mentioned as separable data linearly. The kernel characteristic is applied onto the set of schooling samples in addition to the test samples and then, it's far reached within the exact way as within the case of case which is linearly separable. A sparse classifier contains non-linear facts Results This algorithm classifies the test data with the sensitivity of 86%. Paper 3: Algorithm & Feature extraction: CNN Used for both Convolutional Neural Network(CNN) A CNN is a sort of engineered neural system utilized in picture notoriety and preparing that is uniquely intended to way pixel information. CNNs are ground-breaking photo preparing, engineered insight (AI) that utilizes profound acing to play out each generative and enlightening undertakings, routinely utilizing machine vison that incorporates picture and video notoriety, related to recommender frameworks and natural language processing (NLP). A neural system is an arrangement of equipment as well as programming program designed after the activity of neurons inside the human mind. Conventional neural systems aren't ideal for picture handling and must be taken care of pictures in diminished goals pieces. CNN have their "neurons" composed progressively like those of the frontal flap, the territory chargeable for preparing visual boosts in individuals and various creatures. The layers of neurons are organized in such a way as to cover the whole sight see battling off the piecemeal picture preparing issue of conventional neural. A Convolutional Neural Network utilizes a contraption like a multi-layer approach which has been calculated for abated handling prerequisites. The layers of a Convolutional Neural Network have the information layer, a yield layer and a shrouded layer which then contains various convolutional layers, pooling layers, totally associated layers and standardization layers. The evacuation of conditions and addition in execution for photograph preparing outcomes in a gadget that is far more prominent powerful, less complex to trains kept for picture handling and characteristic language handling Feature Extraction in CNN: Convolutional Neural Network have exhibited impressive accomplishment because of their capacity to display neighborhood conditions in the info & diminish the quantity of prepared criterion in a neural system over weight sharing. We influence two of these attributes on the wavelet-changed EEG. We utilize the convolutional channels to understand highlights that catch the transient fleeting conditions of the EEG, and simultaneously search for connections between close frequency bands. We will quickly portray a few highlights of the CNN depth, that we exploit here. Notwithstanding learning channel maps, CNN likewise include two other normal segments: max-pooling and dropout. The maximum pooling technique permits the system to learn highlights which are transiently constant. The system after which would be able to distinguish designs in the coefficients of the wavelet tensor without thinking about whether the example happens to start with or end of the sign cut. Dropout haphazardly sets the yield of units in the system to zero during preparing, keeping these from influencing the yield or the inclination of the misfortune work for an update step. This fills in as a regularization method to improve speculation by guaranteeing the system doesn't overfit by relying upon explicit concealed units. The Convolutional Neural Network prepared had 6 convolutional layers after which were two thick layers. The output layer comprised of three units with a soft-max enactment speaking to a probability dissemination over the three classes. We showed up at this design subsequent to trying different things with a wide range of structures, both shallower and more extensive. More profound models demonstrated hard to prepare on the accessible equipment, although shallower setups experienced bad precision. The conv layers are deformed consistently, diminishing the quantity of attributes in individual layer. Filter dimensions were set at 3x3 and 2x2 as analyses along with bigger channels indicated no advancement. Maxpooling layers were planted after each other convolutional layer and dropout included after each layer. Rather than the commonplace sigmoid nonlinear initiation, the newly created - linear layer is utilized for its non-immersing effects which do hinder the evaporating gradient issue Results: The outcomes on the EEG database of 199 chronicles show that (i) the preictal stage progress happens about 12 minutes before seizure beginning, (ii) forecast outcomes on the test set are assuring, along with an affectability of 90% plus a lower bogus expectation pace of 0.122 FP/h

Sources	Similarity

SmallSEOTools

PLAGIARISM SCAN REPORT

Words	845	Date	May 03,2020
Characters	5112	Exclude Url	

0% Plagiarism	100% Unique	0 Plagiarized Sentences	35 Unique Sentences

Content Checked For Plagiarism

3. SYSTEM ANALYSIS 3.1. Functional Requirements 3.1.1 Business Rules The users must be able to use the product under proper licensed time durations only. 3.1.2. Authentication Only an authorized neurologist must get an access to the system, thus proper authentication measures must be applied. 3.1.3. Audit Tracking Proper logging mechanisms for convenient report generations and auditing should be established by the product. 3.1.4. Ability to classify The algorithm must be able to classify the patients with appropriate preictal/non-preictal states as accurately as possible. 3.1.5 Ability to Continuously Learn The algorithm must be able to learn from the historic data and keep on improving its accuracy. 3.2. Non Functional Requirements 3.2.1. Performance In our system, performance is given prime importance. Our system is based on real time event and the classification of data must be fast. The system should be able to handle multiple incoming data. Under these conditions, the performance should not hinder. Thus, the system should work under any environment and type of adversarial poisoning. 3.2.2. Accuracy Accuracy is the most important part in our project. As the real time implementations of these problems are critical, the system should generate accurate outputs to measure accuracy, along with confusion matrix we will be using other performance measures that include precision recall. 3.2.3. Reliability The system should be reliable so that the chances of getting an intruder into the system becomes less. Since every data in our system is unique, the number of false negative should be less and also handle concept drift in the data with increase in time. 3.2.4. Recoverability If the system fails, the system should be immediately restarted and should be connected to the database. The database should keep log of classified data and if new data is entered then system should classify it as early as possible. 3.3. Specific Requirements 3.3.1. Hardware Requirements 16GB RAM (8GB works as well but not for the performance you may want and or expect) Quad core Intel Core i7 Skylake or higher (Dual core is not the best for this kind of work, but manageable). M.2 PCIe or regular PCIe SSD with at minimum 256GBs of storage, although having 512GB gives you the best performance. The performance of the system is directly proportional to the loading and saving of your application. (SATA III might hamper the system's performance) Premium graphics cards, so things with GTX 980 or 980Ms would be the best for a laptop, and 1080s or 1070s would be the best for the desktop setup. (try not to sacrifice too much here. While a 980TI or a 970m may be cheaper, this as well is a critical part of the system, others you will notice a drastic drop in performance). 3.3.2. Software Requirements Front End:- HTML, CSS, Bootstrap Flask Programming Language:- Python Keras - Tensorflow PyEEG Module 3.4. Use-Case Diagrams and description 3.4 Use Case Diagram 4. ANALYSIS MODELING 4.1 Activity Diagrams / Class Diagram 4.1 Activity Diagram/Class Diagram 4.2 Functional Modeling 4.2 Functional Modeling 4.3 Timeline Chart 4.3 Timeline Chart 5. DESIGN 5.1 Architectural Design 5.1 Architectural Design PRATIKS PART NOT EDITED EDA (Beginning of implementation) The Dataset here corresponds to what looks as a set of 11500 patients. Each data point in the dataset for one particular patient corresponds to a time value which was recorded at the time. Each patient corresponds to 178 distinct values of point in time at the time of recording or curating of the dataset. The last column in the dataset is the labelled prediction of one patient. It originally revolves around 5 distinct labels i.e. 1, 2, 3, 4, 5. But since we're making a binary classification of the dataset we will have to truncate it down to 2 labels namely 1.0. As we can see that we have 178 eeg features with 5 possible classes. The main aim of the project here is to classify the correctly whether an eeg data provided has a possibility of seizure or not. Thus this problem deals with the binary classification. The following graph shows the data distribution of data among the two classes which would be used for binary classification later in the project. Preprocessing of data 1. Null Value Analysis By the figure above it's completely evident that there are 180 null values in our sample of 11500 tuples with 178 columns. 2. General Arithmetic Analysis 3. Plotting outputs across inputs Classes of data here are categorically placed across Y axis as samples of time within a given range. Results & Analysis (Last) We were aiming for accuracies as good as possible. When we ran the training we were elated to see 90+ accuracies in both our approaches. It's worthwhile to state that both the accuracies were achieved in fewer epochs, which in turn means that our models in both the approaches i.e. CNN and LSTM were strong enough to fetch us these accuracies. LSTM As we can see, the lstm model fetched us an accuracy of 97% in just 20 epochs. Loss while training 0.079 Accuracy 0.97 F1 Score 0.97 Recall 0.95 Precision 0.98

Sources	Similarity

SmallSEOTools

PLAGIARISM SCAN REPORT

Words	536	Date	May 03,2020
Characters	3152	Exclude Url	

3% Plagiarism	97% Unique	1 Plagiarized Sentences	29 Unique Sentences

Content Checked For Plagiarism

CNN We can see that we achieved 95% accuracy with the given dataset and layer configurations and in 20 epochs. Validation Loss 0.08 Accuracy 0.95 Precision 0.95 F1 Score 0.95 Recall 0.95 Accuracy: - Accuracy is nothing but the ratio of true positives with the total size of the dataset. Accuracy is the most used factor in deep learning when we decide to optimize the model based on some metric Accuracy = TP+TN/TP+FP+FN+TN Precision:- Precision is the ratio of true positives with the ratio of total number of positives which include true positives and false positives. This metric gives us the true positives out of a total of positives out of which some could be faulty. Precision = TP/TP+FP Recall: - Recall is also known as sensitivity. Recall is the ratio of true positives out of all the given data. Recall here does the exactly opposite of what precision does. It just gives you the false positives + true positives. This is not the best measure to optimize your algorithm with. Recall = TP/TP+FN F1 Score: - F1 Score is the weighted average of Accuracy and Precision. F1 Score is not as easy to grasp as Precision and Accuracy are. But this could be an accurate metric to optimize your algorithm with because it takes into account both false positives and false negatives. Let alone it's not the metric to completely rely on. F1 Score = 2 x (Recall x Precision) / Recall + Precision Black Box Testing of the GUI (Add this under testing section) 1. Open your terminal from the anaconda dialog box 2. Generate a patient 3. You could see the matrix representation of the wave image 4. A second window reopens which has the wave form depiction 5. Close the dialog box and click on the Test button to get the results for the given patient. OR Implementation 1. CNN Model decisions 1.1. Gaussian Noise was added to improve generalization error which could be problem if not managed well. Adding noise for augmentation can prevent overfitting. 1.2. Batch Normalization and Dropout were used to decrease the overfitting of the model. 1.3. Adam optimizer was used instead of the classic stochastic gradient descent as it was especially designed for updating weights of deep neural networks. It is definitely more efficient. 1.4. The number of hidden convolution layers decrease gradually as the data passed across the network continually complexes. The number of neurons in the first layer is 24, the second has 16 and the third has 8. 1.5. Dense layers were used to make a robust fully densely connected neural network 2. LSTM Model Decisions 2.1. Gaussian Noise was added to improve generalization error which could be problem if not managed well. Adding noise for augmentation can prevent overfitting. 2.2. Batch Normalization and Dropout were used to decrease the overfitting of the model. 2.3. 60 neurons of lstm was used as it was 2/3rd of the number of input features. 2.4. Adam optimizer was used instead of the classic stochastic gradient descent as it was especially designed for updating weights of deep neural networks. It is definitely more efficient. 2.5. Dense layers were used to make a robust fully densely connected neural network.

Sources	Similarity
MachineLearning/DecisionTree at master · HariniGB/MachineLearning... score is 0.982 or 98%. note: defining some of the attributes like max_depth, max_leaf_nodes, min_impurity_split, and min_samples_leaf can help prevent over-fitting the model to the training data.f1 score = 2 x (recall x precision) / (recall + precision). https://github.com/HariniGB/MachineLearning/tree/master/DecisionTree	5%

PLAGIARISM SCAN REPORT

Words	954	Date	May 03,2020
Characters	5378	Exclude Url	

6% Plagiarism	94% Unique	3 Plagiarized Sentences	46 Unique Sentences

Content Checked For Plagiarism

Convolution comes from the Latin word, "to convolve" means to roll together. Convolution may be a mathematical process on two functions (f and g) to construct a 3rd function that represents how the form of 1 is modified by the opposite . it's the integral calculating what proportion two functions overlap together passes over the opposite . Assume convolution as how of mixing two functions by multiplying them. 2012 was the primary year that neural nets grew to prominence as Alex Krizhevsky used them to win that year's ImageNet competition (basically, the annual Olympics of computer vision), dropping the classification error record from 26% to fifteen , an astounding improvement at the time.Ever since then, a number of companies are using deep learning at the core of their services. Facebook uses neural nets for his or her automatic tagging algorithms, Google for his or her photo search, Amazon for his or her product recommendations, Pinterest for his or her home feed personalization, and Instagram for his or her search infrastructure. Whenever we see a picture , our brain looks for the features within the image to classify that image. We categorize things by recognizing the features. If we glance to the proper side of the image, we'll see an individual looking towards the proper side while, if we glance within the center, we'll perceive that the person is looking towards us. Does our brain struggle tons in identifying these different scenarios and is confused if the person is looking in right or towards us? This happens because our brain studies for the features within the image then assume what to require . No! That's because the features depicted are inadequate to help the brain in classifying them. All the above images are addressed to know that our brain functions on the features of the image it sees then classifies it accordingly. during a similar manner, neural networks work. we will see within the image below, the neural network has successfully classified cheetah and bullet but was unsuccessful in predicting hand glass. this is often due to the unclear features within the image. In simple words, Neural Networks works exactly sort of a human mind. Algorithm: - 1. Convolution 1. Convolution A convolution may be a combined integration of two functions that shows you ways one function modifies the opposite . There are three important items to mention during this process: the input image, the feature detector, and thus the feature map. The input image is that the image being detected. The feature detector could also be a matrix, usually 3x3 (it could even be 7x7). A feature detector is additionally mentioned as a kernel or a filter. the row-column representation of the input image is multiplied element-wise with the feature detector to provide a feature map, also mentioned as a convolved feature or an activation map. The point of this step is to minimize the size of the image and make processing quicker and hassle free. a number of the features of the image are lost during this step. However, the foremost features of the image that are important in image detection are retained. These features are those that are unique to identifying that specific object. for instance each animal has unique features that enable us to identify it. The way we prevent loss of image information is by having many feature maps. Each feature map detects things of certain features within the image. Edge Detection Example Edge Detection Example In the previous article, we saw that the first layers of a neural network detect edges from a picture . Deeper layers could be ready to detect the explanation for the objects and even more deeper layers might detect the explanation for complete objects (like a person's face). In this section, we'll specialise in how the sides are often detected from a picture . Suppose we are given the below image: As you can see, there are many edges in the image. The first thing to do is to find these edges: But how do we detect these edges? To demonstrate this, let's take a 6 X 6 grayscale image (i.e. only one channel): Next, we convolute this 6 X 6 matrix with a 3 X 3 filter: After the convolve, we will get a 4 X 4 image. The first entry of the 4 X 4 matrix will be calculated as: So, we take the primary 3 X 3 matrix from the 6 X 6 image and multiply it with the filter. Now, the primary element of the 4 X 4 output are going to be the sum of the element-wise product of those values, i.e. 3*1 + 0 + 1*-1 + 1*1 + 5*0 + 8*-1 + 2*1 + 7*0 + 2*-1 = -5. To calculate the second element of the 4 X 4 output, we'll shift our filter one step towards the proper and again get the sum of the element-wise product: So, convolving a 6 X 6 input with a 3 X 3 filter gave us an output of 4 X 4. So, convoluting a 6 X 6 input with a 3 X 3 filter gave us an O/P of 4 X 4. Consider one more example: Max Pooling/Down Sampling with CNN The next layer during a convolutional network has three names: max pooling, downsampling and subsampling. The maps are fed into a downsampling layer, and like convolutions, this method is applied one patch at a time. In this case, max pooling simply takes the most important value from one patch of a picture , places it during a new matrix next to the max values from other patches, and discards the remainder of the knowledge contained within the activation maps.

Sources	Similarity
A Beginner's Guide To Understanding Convolutional Neural Networks... 2012 was the first year that neural nets grew to prominence as Alex Krizhevsky used them to win that year's ImageNet competition (basically, the annual Olympics of computer vision), dropping the classification error record from 26% to 15%, an astounding improvement at the time.Ever since... https://www.sites.google.com/site/nttrungmtwiki/home/it/deep-learning/convolution-neural-network/a-beginner-s-guide-to-understanding-convolutional-neural-networks	10%
Convolutional Neural Networks: An Intro Tutorial - Heartbeat the way we prevent loss of image information is by having many feature maps. each feature map detects the location of certain features in the image.we are going to preprocess the images using keras to prevent overfitting. this processing is known as image augmentation. the keras utility we... https://heartbeat.fritz.ai/a-beginners-guide-to-convolutional-neural-networks-cnn-cf26c5ee17ed	4%
A Comprehensive Tutorial on Convolutional Neural Networks (CNNs) so, convolving a 6 x 6 input with a 3 x 3 filter gave us an output of 4 x 4. consider one more exampleone layer of a convolutional network. once we get an output after convolving over the entire image using a filter, we add a bias term to those outputs and finally apply an activation function... https://www.analyticsvidhya.com/blog/2018/12/guide-convolutional-neural-network-cnn/	3%

SmallSEOTools

PLAGIARISM SCAN REPORT

Words	848	Date	May 03,2020
Characters	5254	Exclude Url	

2% Plagiarism	98% Unique	1 Plagiarized Sentences	50 Unique Sentences

Content Checked For Plagiarism

ANN is a system that simulates the neural networks present in living beings. The work in a similar way as to that of animals or more closely like humans. They aren't programmed for a specific task. We use examples to make them learn. The best use of ANN may be image recognition. Labelled pictures of an object such as dogs or cats are manually fed to the system. No information of the constituents of the picture are entered. The ANN takes the features into account while creating the model. Basic Model Of ANN Artificial neurons are connected to each other. It mimics the human brain. Each neuron can transmit signal to the other neurons. A neuron receives a signal then processes it and sends this processed signal further. The input signal is a real number. A non-linear function computes the sum of inputs and passes it ahead. The connections are known as edges. The learning is adjusted by the weight associated with the edges. The weight increases or decreases the strength of the signal at a connection. A threshold determines which signal goes across and which doesn't. There are multiple layers between the input and output layers. There can be zero or one or more number of layers. Different transformations happen to the signal at each layer providing meaningful changes to the system as whole. Right now ANN is only used to solve a specific type of problem. General intelligence is still not possible. Therefore ANNs can be used for a variety of things like image recognition, understanding speech or grammar and in this case medical diagnosis. Each link between the neurons or nodes have influence over the other nodes. The whole system forms a directed or undirected graph based on what it is trying to achieve. Single Layer of ANN PRINCIPLE The basic working is based on the principle that a small change in input brings a large change in output. A given neuron can have multiple inputs or a single input. The same is true for outputs as well. Propagation function computes the output by processing the inputs from the previous neuron. Deep Learning may use multiple layers as it is more advanced in computing results. Multiple connections are also possible between the layers. It could be a many-to-one connection or one-to-one. LEARNING The actual process of making a system which computes data correctly is this step. Sample data is used to adjust the weights of the network to improve accuracy. Errors are minimized. The process happens over and over again till the error cannot be reduced any further. The error rate cannot be reduced to zero. We may redesign the network if the output does not prove to be accurate. Hundred percent accuracy is not possible in such a system but ninety plus is generally considered good. Learning rate is the steps the model takes to adjust errors in each observation. A high learning rate gets the work done faster but the accuracy suffers. Multilayer ANN Feedforward ANN The information flow is unidirectional. The information does not return to a neuron once it leaves it. It is used in Simple Pattern recognition. The inputs and outputs are fixed. Feedback ANN The same neuron may get a signal multiple times Supervised Learning Here the system comes up with guesses as to what the correct answer may be. The correct answer is given to the system. According to the difference between the answer and guess the network makes changes. The cost function here eliminates incorrect deductions. The most commonly used cost is mean-squared error. Pattern recognition and regression are done using this method. Handwriting and gesture recognition problems can be dealt with using this technique. Unsupervised Learning Here the output in not known therefore the correct answers to the dataset aren't available. This could be used when searching for new patterns. The input data along with cost function are provided at the input. Reinforcement Learning It works on the principle of observing the environment. The best example of this method is playing video game. The environment could give you unpredictable response in every game but you start learning from each game. The goal is to get the lowest cost possible.

The goal has to be achieved as early as possible. Initially everything is estimated but as the game progresses things start getting better. Backpropagation Backpropagation is used to modify the weights to compensate for the errors. Mathematically the derivative of the cost function is linked with a given state. Applications Used is quantum chemistry, playing games, face identification, signal classification, 3D construction sequence recognition, medical diagnosis, controlling car, trajectory prediction, finance, trading, data mining, social network filtering and spam filtering. It is also used in diagnosis of cancer. It is also used in geoscience, coastal engineering, cybersecurity and geomorphology. It can also be used to simulate properties of quantum systems although it is still in research stage. EEG, ECG analysis, designing prosthetics, speech classification, text to speech conversion, image compression, automated information services, braking vehicles, audio signal processing, anomaly detection, guidance system, weapons, electronics, manufacturing products etc.

Sources	Similarity
Artificial neural network - Wikipedia the weight increases or decreases the strength of the signal at a connection. neurons may have a threshold such that a signal is sent only if the aggregate signal crosses that threshold.each link has a weight, which determines the strength of one node's influence on another.[40]. https://en.wikipedia.org/wiki/Artificial_neural_network	5%

SmallSEOTools

PLAGIARISM SCAN REPORT

Words	654	Date	May 03,2020
Characters	4191	Exclude Url	

0%	100%	0	32
Plagiarism	Unique	Plagiarized Sentences	Unique Sentences

Content Checked For Plagiarism

Huge and viable neural systems require significant computer resources. While the brain has equipment custom fitted to the undertaking of preparing signals through a chart of neurons, recreating even a rearranged neuron on von Neumann design may expend immense measures of memory and capacity. Moreover, the fashioner frequently needs to transmit flags through a large number of these associations and their related neurons – which require colossal CPU force and time. The resurgence of neural systems in the twenty-first century is to a great extent owing to propels in equipment: from 1991 to 2015, registering power, particularly as conveyed by GPUs has expanded around a million-overlay, making the standard backpropagation calculation achievable for preparing systems that are a few layers further than before. The utilization of quickening agents, for example, FPGAs and GPUs can diminish preparing times from months to days. Neuromorphic building tends to the equipment trouble straightforwardly, by developing non-von-Neumann chips to legitimately actualize neural systems in hardware. Another kind of chip upgraded for neural system handling is known as a Tensor Processing Unit, or TPU. Dependency graph Mathematics Applications where the system generalizes is all the inputs is seen as an overtrained model. This happens when the network capacity significantly exceeds the free parameters. We can use cross-validation and many other to check the presence and minimize error. We can also use regularization where a larger probability is use over simpler models. Supervised networks that use mean squared error cost function use statistics to determine the confidence of the trained model. Variance can also be found using MSE. This attribute can then be used to calculate the confidence interval of network output, assuming a normal distribution. A confidence analysis made this way is valid as long as the output probability distribution stays the same and the network is not modified. By allocating a softmax initiation work, a speculation of the strategic capacity, on the yield layer of the neural system (or a softmax segment in a segment based system) for all out objective factors, the yields can be deciphered as back probabilities. This is valuable in characterization as it gives a sureness measure on orders. Advantages of using ANN Learning what the ANN has learned is much easier to analyse than that of the biological neural network. Also more and more research involving the learning algorithms of the system are uncovered which help the data process faster. Local vs Non-Local Learning. Advanced The current system may be linked to the human mind in the future for us to get all the data at lightning speed. Types of learning Two modes are available. Stochastic and Batch. In stochastic learning, each input makes a weight adjustment. In batch learning weights are adjusted supported a batch of inputs, accumulating errors over the batch. Stochastic learning introduces "noise" into the method , using the local gradient calculated from one data point; this reduces the prospect of the network getting stuck in local minima. However, batch learning typically yields a faster, more stable descent to an area minimum, since each update is performed within the direction of the batch's average error. A common compromise is to use "mini-batches", small batches with samples in each batch selected stochastically from the whole data set. Chapter 9 Results and Discussions This shall form the penultimate chapter of the report and shall include a radical evaluation of the investigation administered and convey out the contributions from the study. The discussion shall logically cause inferences and conclusions also as scope for possible further future work. Chapter 10 Conclusions This will be the ultimate chapter of the report. a quick report of the work administered shall form the primary a part of the Chapter. Results derived from the analysis exhibited in the Results and Discussions Chapter shall be presented and clearly calculated, each point stated separately. Scope for future work should be stated lucidly within the last a part of the chapter.

Sources	Similarity