DETECTION OF INTRACRANIAL BLEEDING USING AN EFFECTIVE NEURAL NETWORK

A PROJECT REPORT

Submitted by

KARTHIKEYAN K	**311117106030**
SANJAY ROY J	**311117106047**
SURIYA KUMAR S	**311117106051**
TERESSA ALPHONSA DOMINIC	**311117106053**

in partial fulfilment for the award of the degree of

BACHELOR OF ENGINEERING

in

ELECTRONICS AND COMMUNICATION ENGINEERING

LOYOLA–ICAM COLLEGE OF ENGINEERING & TECHNOLOGY,

CHENNAI -600034

ANNA UNIVERSITY: CHENNAI 600 025

April 2021

ANNA UNIVERSITY: CHENNAI 600 025

BONAFIDE CERTIFICATE

Certified that this project report **"DETECTION OF INTRACRANIAL HAEMORRHAGE USING AN EFECTIVE NEURAL NETWORK"** is the bonafide work of**," KARTHIKEYAN K , SANJAY ROY J, SURIYA KUMAR S and TERESSA ALPHONSA DOMINIC"** who carried out the project work under my supervision.

SIGNATURE	**SIGNATURE**
Dr. D. EGFIN NIRMALA, Ph.D.	**Dr. K. KUNARAJ, Ph.D.**
HEAD OF THE DEPARTMENT	**SUPERVISOR**
Professor	Associate Professor
Department of ECE,	Department of ECE
LICET,	LICET,
Loyola Campus,	Loyola Campus,
Nungambakkam,	Nungambakkam,
Chennai-600 034	Chennai-600 034

Submitted for the viva-voce examination held on______.

INTERNAL EXAMINER **EXTERNAL EXAMINER**

ACKNOWLEDGEMENT

We are grateful to our Almighty God for His blessings to complete our project successfully. We would also like to thank our parents for their continuous support and encouragement in this regard. We express our sincere gratitude to Rev.Dr.Maria Wenisch SJ, Director, LICET.

We extend our gratitude to Dr.L.Antony Michael Raj, Principal, for his support towards our project. We wish to convey our sincere thanks to Rev.Dr.D.Caleb Chanthi Raj, Dean of Students, for his support and inspiration towards this project.

We express our sincere appreciation to Dr.D.Egfin Nirmala, Professor and Head, Department of Electronics and Communication Engineering for her constant care, motivation and reassurance for being our source of inspiration throughout this project and during the course of this semester.

We are deeply indebted to our supervisor Dr.K.Kunaraj, Associate Professor, for his able guidance and whole-hearted cooperation which enabled us to finish this project successfully. We convey our special thanks to our Class Advisor, Mrs.R.Vidhya, Assistant Professor, for providing us required resources and timely support for completing this project.

ABSTRACT

Intracranial haemorrhage is a serious medical emergency because the accumulation of blood inside the skull may cause intracranial pressure to rise, potentially crushing delicate brain tissue or limiting its blood supply. Brain herniation occurs when portions of the brain are squeezed past structures in the skull due to severe increases in intracranial pressure (ICP). In the United States, intracranial haemorrhages account for about 10% of strokes, making stroke the fifth-leading cause of death. Hence, the location and type of any haemorrhage present must be determined before the patient can be treated. The detection of the bleeding inside the brain is often time consuming and the diagnosis of the haemorrhage is delayed often.

In order to overcome this difficulty, we propose certain methods, based on supervised learning algorithms for the early detection of intracranial haemorrhage. The first method we implemented uses Alexnet which is a pre-trained neural network and it is further trained and a classification model is created to classify the CT images of haemorrhages. The trained model using the CT images of the brain are then used to classify the input images into different types of intracranial haemorrhage. In the second method we use the CT images of brain to the train the neural network using a modified CNN model instead of transfer learning. The goal is to evolve a suitable model to detect and classify intracranial haemorrhages with utmost accuracy.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

SERIAL NO.	ABBREVIATION	MEANING
1	ANN	Artificial Neural Network
2	CNN	Convolutional Neural Network
3	CT	Computed Tomography
4	DA	Data Analysis
5	FCNN	Fully Connected Neural Network
6	GAN	Generative Adversarial Network
7	LSTM	Long Short-Term Memory
8	MRI	Magnetic Resonance Imaging
9	ReLU	Rectified Linear Unit
10	RGB	Red, Blue, Green
11	SVM	Support Vector Machine

CHAPTER 1

INTRODUCTION

1.1 Intracranial haemorrhage

Intracranial haemorrhage (ICH) refers to acute bleeding inside your skull or brain. It’s a life-threatening emergency. Intracranial bleeding occurs when a blood vessel within the skull is ruptured or leaks. It can result from physical trauma (as occurs in head injury) or non-traumatic causes (as occurs in haemorrhagic stroke) such as a ruptured aneurysm. Anticoagulant therapy, as well as disorders with blood clotting can heighten the risk that an intracranial haemorrhage will occur. More than half of all cases of intracranial haemorrhage is the result of hypertension.

There are five types of ICH [18]:

1. Epidural hematoma
2. Subdural hematoma
3. Subarachnoid haemorrhage
4. Intra-ventricular haemorrhage
5. Intra-parenchymal haemorrhage

1.1.1 Epidural hematoma

Hematoma is a collection of blood, in a clot or ball, outside of a blood vessel. An epidural hematoma occurs when blood accumulates between your skull and the outermost covering of your brain.[18] It typically follows a head injury, and usually with a skull fracture. High-pressure bleeding is a prominent feature. If you have an epidural hematoma, you may briefly lose consciousness and then regain consciousness.

1.1.2 Intraventricular haemorrhage

Intraventricular haemorrhage (IVH), also known as intraventricular bleeding, is a bleeding into the brain's ventricular system, where the cerebrospinal fluid is produced and circulates through towards the subarachnoid space. It can result from physical trauma or from haemorrhagic stroke.

1.1.3 Subdural hematoma

Subdural Hematoma is a collection of blood on the surface of your brain. It’s typically the result of your head moving rapidly forward and stopping, such as in a car accident. However, it could also suggest abuse in children. This is the same type of movement a child experiences when being shaken. A subdural hematoma is more common than other ICHs in older people and people with history of heavy alcohol use.

1.1.4 Subarachnoid haemorrhage

Subarachnoid haemorrhage is when there’s bleeding between the brain and the thin tissues that cover the brain. These tissues are called meninges. The most common cause is trauma, but it can also be caused by rupture of a major blood vessel in the brain, such as from an aneurysm. A sudden, sharp headache usually comes before a subarachnoid haemorrhage. Typical symptoms also include loss of consciousness and vomiting.

1.1.5 Intraparenchymal haemorrhage

Intraparenchymal haemorrhage (IPH) is one form of intracranial bleeding in which there is a bleed inside the brain. Intraparenchymal haemorrhage accounts for approx. 8-13% of all strokes and results from a wide spectrum of disorders.

1.2 Motivation

Intracerebral haemorrhage (ICH) is a life-threatening condition. In the United States, the overall prevalence of spontaneous ICH is 24.6 per 100,000 person-years, with 40,000 to 67,000 cases per year. Only 20% of survivors are expected to have complete functional recovery at 6 months, according to the 30-day death rate, which ranges from 35% to 52%. Approximately half of all deaths occur within the first 24 hours, emphasising the crucial importance of prompt and effective treatment in the ER (ED). ICH is twice as frequent in low-to-middle income countries compared to high-income countries. In the United States, several studies have shown that the incidence of ICH is greater in African Americans and Hispanics than in whites.

1.3 Objective

Intracranial bleeding requires often rapid and intensive medical care and the diagnosis is often time consuming. The diagnosis can be done by using neural networks that can detect the bleeding in the CT scan and predict the type of the hemorrhage thereby reducing the time needed for the detection of intracranial hemorrhages in normal cases.

1.4 Approach

The data is first grouped into five groups, totaling around 32,000 images. We perform image pre-processing techniques such as image resizing and RGB to grayscale conversion with the available images. The image may become stretched during resizing, so we normalize it to provide proper feed to the model. Following the completion of the image pre-processing techniques, the CNN model is created and trained with values. Over 80% of the total data is fed into the training model, and the model is capable of detecting bleeding and classifying it into its types. It can be examined by running the model through the remaining 20% of images.[17]

1.5 Result

Detection of Intracranial bleed at their early stage increases the survival rate of patients. The main reason for many dying of this intracranial bleeding is because doctors cannot diagnose this at an early stage. The proposed system will help doctors to find the bleed at early stage.

CHAPTER 2

LITERATURE SURVEY

A significant number of studies have been conducted in the past on the use of artificial intelligence to detect intracranial haemorrhages. Research has been described as papers published in e-journals and international conferences that use deep learning techniques to detect intracranial haemorrhages for the purposes of this study. The wealth of contributions on characterising skin lesions using artificial intelligence is revealed in brief reviews of the following research works.

1.The paper [1] gives a detailed description of intracranial haemorrhage and its subtypes are detected using a three-dimensional joint and recurrent neural network. A joint CNN-RNN classification framework was proposed with flexibility to train when subject level or slice level labels are available.

2. In the case of epilepsy [2], a device employs a convolutional long short-term memory (C-LSTM) neural network to improve detection of seizures and tumours. Other models have a lower recognition rate than this strategy (up to 98.80 percent). With an average accuracy of 93.38 percent, the proposed C-LSTM is the most accurate.

3. The paper [3] goes into great detail about The Deep Wave let Autoencoder-Deep Neural Network image classifier was used to analyse a brain image dataset. The Deep Wavelet Autoencoder-Deep Neural Network classifier's output was compared to that of other existing classifiers such as Autoencoder-Deep Neural Network and Deep Neural Network, and it was found that the proposed method in this paper produces better results.

4. Using two publicly available datasets, the paper [4] presents a comprehensive overview of a Deep Learning model based on a convolutional neural network that is proposed to identify different brain tumour types. With the best overall accuracy of 96.13 percent, the proposed network structure achieves a significant efficiency. The findings show that the model can be used to study brain tumours.

5.This paper [5] provides a comprehensive description of a 2D-CNN to 3D-CNN system to extract better variability in modes. The loss function also enhances the interference in brain tumour detection in other areas. This enhances the detection accuracy.

6.In this paper [6], two model variants are proposed, one of which is the 3D VGGNet architectures, Resnet, and the proposed method enhances classification performance significantly. The best F1 scores for binary classification were 0.96 (normal vs SAH), 0.93 (normal vs IPH), 0.98 (normal vs SDH), and 0.99 (normal vs IVH), while the average F1 score for four-class classification was 0.77.

7. A correlation learning mechanism (CLM) for deep neural network architectures is proposed in this paper [7], which combines convolutional neural network (CNN) and classic architecture. The support neural network aids CNN in deciding which filers are best for pooling and convolution layers. As a result, the main neural classifier learns faster and is more effective, achieving 96 percent accuracy, 95 percent precision, and 95 percent recall.

8. Automatic brain tumour classification is a difficult task due to the significant spatial and structural heterogeneity of the brain tumour surrounding area. The use of Convolutional Neural Networks (CNN) classification for automatic brain tumour detection is suggested in this paper [8]. Small kernels are used to design the deeper architecture. The neuron's weight is given as small. As compared to all other state-

of-the-art approaches, experimental findings show that CNN archives have a rate of 97.5 percent accuracy with low complexity.

9. In the proposed work [9], the self-organized map NN was used to train the extracted characteristics from the DWT merge wavelet, and the output morphological features and filter factors were then trained by NN, followed by a two-phase testing method. The planned NN classification scheme categorised the brain tumour in a binary qualified method, which offers a better presentation than the standard classification strategy.

10. The proposed diagnostic strategies [10] are divided into three parts. To begin, a method for segmenting the tumour is proposed, followed by an optimization of the data set and the use of Gray level Co-occurrence Matrix (GLCM) for texture analysis, and finally, a method for classifying the tumour into benign or malignant using a SVM classifier. The accuracy of the classifier is obtained using three kernel methods: radial basis function, linear kernel, and polynomial kernel.

11.The aim of this paper [11] is to create a model for detecting and classifying brain tumours, specifically whether the tumour is cancerous or non-cancerous, using the SVM algorithm. Many people have previously detected using ANN, which is based on Empirical Risk Minimization. To identify the images, we used the SVM algorithm, which works on structural risk minimization. For tumour extraction, the SVM algorithm is applied to medical images, and a Simulink model is created for the tumour classification function.

12.The most powerful and fundamental image processing technique is segmentation, which is used to remove suspicious regions from a given image. MRI images are used to detect brain tumours at advanced stages. The aim of this study [12] is to use a quantitative approach to measure brain tumour loss in MRI human head scans. For finding the centroid value to segment the brain tissue, this approach proposes Particle Swarm Optimization (PSO).

13.The authors proposed [13] a two-step GAN-based (Progressive Growing of GANs & Multimodal Un-supervised Image-to-image Translation) DA that generates and refines brain Magnetic Resonance (MR) images with/without tumours separately, significantly outperform the classic DA alone, in tumour detection.

14. This study [14] based on convolutional neural network, combined with MRI detection technology to construct a model adapted to brain tumour feature detection. The main function of this research model is to segment and recognize MRI brain tumours and use convolutional layer to perform convolution operation to improve recognition efficiency and rate and combine artificially selected features with machine learning features.

15.In this paper [15], an improved orthogonal gamma distribution-based machine-learning approach is used to analyse the under- segments and over-segments of brain tumour regions to automatically detect abnormalities in the ROI.

CHAPTER 3

METHODOLOGY

3.1 Artificial Intelligence in medicine

The technology and automation in medicine, regardless of whether or not we understand it, is already incredible. Medical records are digitalised, appointments can be scheduled online, patients can check in to health centres or clinics using their phones or computers. As technology has become more prominent in all facets of life, it has also quietly altered how we pursue medical treatment.

Increased AI use in medicine has the potential to minimise manual activities, free up physician time, and improve efficiency and productivity while also encouraging us to shift towards more "precision medicine."

It is the ability to gain information, to process and deliver a clearly defined result to end users that characterises AI technology from traditional healthcare technologies. AI does so through algorithms for machine learning. These algorithms can identify behavioural patterns and create their own logic. AI algorithms must be tested repeatedly in order to reduce the margin of error. In two ways, AI algorithms act differently than humans:

- ❖ Algorithms are literal: if you set a goal, the algorithm can't adjust itself and only understand what it has been told explicitly
- ❖ Algorithms are black boxes: algorithms can predict extremely precisely, but not the cause or the why.

It takes years of medical training to correctly diagnose diseases. Diagnosis is also often a time-consuming and arduous process. In many areas, Expert demand far surpasses the supply available. This strain doctors. Often, patient diagnosis is delayed. Machine learning, especially recent deep learning algorithms made enormous improvements in the automated diagnosis of diseases. More accessible and cheaper. The replacement of doctors by AI is unlikely.AI systems are used instead to identify the type of intracranial bleeding for the Expert, which allows the doctor to concentrate on interpreting the signals.

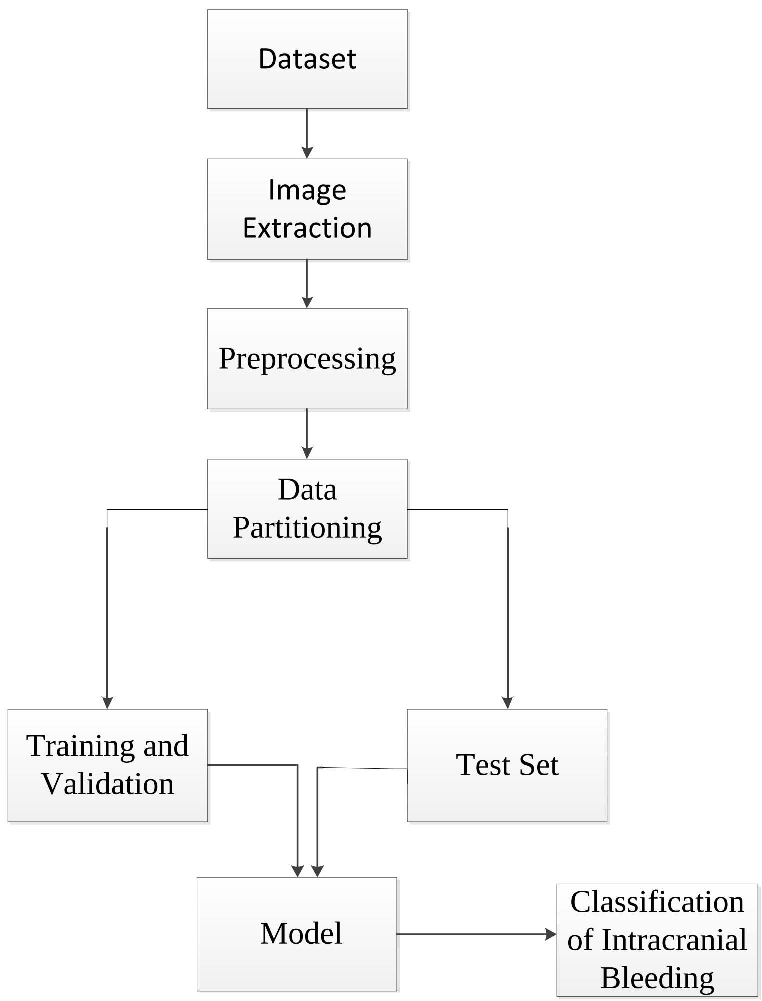

Fig 3.1 Block diagram of intracranial Bleeding Classification System

The figure 3.1 shows the block diagram of our proposed model. The dataset obtained from Kaggle[26]. is segregated into five types of intracranial bleeding. This segregated dataset is then pre-processed to identify the features which can be easily interpreted by the algorithm. The pre-processed images are then divided into train set and validation set. The train set is used to train the model (Alexnet or CNN). The test set is used to validate the results obtained after training of the model. The steps involved in the proposed system is discussed in detail below.

3.2 Image Pre-processing in deep learning:

There are a number of pre-processing steps we might wish to carry out before using this in any Deep Learning project. The paragraphs below list some of the most common.[19]

Uniform aspect ratio: One of the first steps is to ensure that the images have the same size and aspect ratio. Most of the neural network models assume a square shape input image, which means that each image needs to be checked if it is a square or not, and cropped appropriately. Cropping can be done to select a square part of the image, as shown. While cropping, we usually care about the part in the center.

3.2.1 Image Scaling

Once the images have some predetermined aspect ratio, then the images are scaled appropriately. There are a wide variety of up-scaling and down-scaling techniques and we usually use a library function to do this for us.

Mean, Standard Deviation of input data: Sometimes it's useful to look at the 'mean image' obtained by taking the mean values for each pixel across all training examples. Observing this could gave us insight into some underlying structure in the images. We may choose to augment our data with perturbed images if we don't want our input data to have a innate structure. Higher variance values show up whiter, so we see that the pictures vary a lot at the boundaries compared to the centre.

3.2.2 Normalizing image inputs:

Data normalization is an important step which ensures that each input parameter (pixel, in this case) has a similar data distribution. This makes convergence faster while training the network. Data normalization is done by subtracting the mean from each pixel and then dividing the result by the standard deviation. The distribution of such data would resemble a Gaussian curve centered at zero. For image inputs we need the pixel numbers to be positive, so we might choose to scale the normalized data in the range [0,1] or [0, 255].

3.2.3 Dimensionality reduction:

We could choose to collapse the RGB channels into a single gray-scale channel. There are often considerations to reduce other dimensions, when the neural network performance is allowed to be invariant to that dimension, or to make the training problem more tractable.

3.2.4 Data augmentation:

Another common pre-processing technique involves augmenting the existing data-set with perturbed versions of the existing images. Scaling, rotations and other affine transformations are typical. This is done to expose the neural network to a wide variety of variations. This makes it less likely that the neural network recognizes unwanted characteristics in the data-set.

There are multiple types of augmentations possible. The basic ones transform the original image using one of the following types of transformations:

1. Linear transformations
2. Affine transformations

3.2.5 Morphological Transformations

The term morphological transformation refers to any modification involving the shape and form of the images. These are very often used in image analysis tasks. Although they are used with all types of images, they are especially powerful for images that are not natural (come from a source other than a picture of the real world). The typical transformations are erosion, dilation, opening, and closing. Let’s now look at some code to implement these morphological transformations.

3.2.5.1 Thresholding

One of the simpler operations where we take all the pixels whose intensities are above a certain threshold and convert them to ones; the pixels having value less than the threshold are converted to zero. This results in a binary image.

3.2.5.2 Erosion, Dilation, Opening & Closing

Erosion shrinks bright regions and enlarges dark regions. Dilation on the other hand is exact opposite side — it shrinks dark regions and enlarges the bright regions.[20]

Opening is erosion followed by dilation. Opening can remove small bright spots (i.e. “salt”) and connect small dark cracks. This tends to “open” up (dark) gaps between (bright) features.

Closing is dilation followed by erosion. Closing can remove small dark spots (i.e. “pepper”) and connect small bright cracks. This tends to “close” up (dark) gaps between (bright) features.

The basic idea is to have a circular disk of a certain size move around the image and apply these transformations using it.

3.3 Dataset

We downloaded the dataset for the various types of intracranial bleeding from Kaggle [26]. The images we downloaded were in png format. We had approximately 32,000 images totally belonging to five different of intracranial bleeding. These images were later segregated into five types as shown below:

Table 3.1 Distribution of the dataset of intracranial hemorrhages.

Image Subclass	**No of images**
Epidural	5000
Intraventricular	8000
Intraparenchymal	6000
Subdural	8000
Subarachnoid	5000
Total	32000(approx.)

3.4 Weight Balancing

Weight balancing balances our data by altering the *weight* that each training example carries when computing the loss. Normally, each example and class in our loss function will carry equal weight i.e 1.0. But sometimes we might want certain classes or certain training examples to hold more weight if they are more important. We can give weight to the classes simply by multiplying the loss of each example by a certain factor depending on their class. We use focal loss method to balance the weighting of our training model.

in our dataset, we will naturally have some training examples that are easier to classify than others. During training, these examples will be classified with 99% accuracy, while other more challenging ones may still exhibit poor performance. The problem is that those easily classified training examples are still contributing to the loss.

Instead of giving equal weighting to all training examples, focal loss down-weights the well-classified examples. This has the net effect of putting more training emphasis on that data that is hard to classify. In a practical setting where we have a data imbalance, our majority class will quickly become well-classified since we have much more data for it. Thus, in order to ensure that we also achieve high accuracy on our minority class, we can use the focal loss to give those minority class examples more relative weight during training. The focal loss can easily be implemented in Keras as a custom loss function.

3.5 Image Analysis

In digital image processing, the histogram is used for graphical representation of a digital image. A graph is a plot by the number of pixels for each tonal value. Nowadays, image histogram is present in digital cameras. Photographers use them to see the distribution of tones captured.

In a graph, the horizontal axis of the graph is used to represent tonal variations whereas the vertical axis is used to represent the number of pixels in that particular pixel. Black and dark areas are represented in the left side of the horizontal axis, medium grey colour is represented in the middle, and the vertical axis represents the size of the area.

3.5.1 Applications of Histograms

1. In digital image processing, histograms are used for simple calculations in software.
2. It is used to analyse an image. Properties of an image can be predicted by the detailed study of the histogram.
3. The brightness of the image can be adjusted by having the details of its histogram.
4. The contrast of the image can be adjusted according to the need by having details of the x-axis of a histogram.
5. It is used for image equalization. Gray level intensities are expanded along the x-axis to produce a high contrast image.
6. Histograms are used in thresholding as it improves the appearance of the image.
7. If we have input and output histogram of an image, we can determine which type of transformation is applied in the algorithm.

3.5.2 Histogram Processing Techniques

3.5.2.1 Histogram Sliding

In Histogram sliding, the complete histogram is shifted towards rightwards or leftwards. When a histogram is shifted towards the right or left, clear changes are seen in the brightness of the image. The brightness of the image is defined by the intensity of light which is emitted by a particular light source.[21]

3.5.2.2 Histogram Stretching

In histogram stretching, contrast of an image is increased. The contrast of an image is defined between the maximum and minimum value of pixel intensity.

If we want to increase the contrast of an image, histogram of that image will be fully stretched and covered the dynamic range of the histogram.

From histogram of an image, we can check that the image has low or high contrast [21].

3.5.2.3 Histogram Equalization

Histogram equalization is used for equalizing all the pixel values of an image. Transformation is done in such a way that uniform flattened histogram is produced.

Histogram equalization increases the dynamic range of pixel values and makes an equal count of pixels at each level which produces a flat histogram with high contrast image.

While stretching histogram, the shape of histogram remains the same whereas in Histogram equalization, the shape of histogram changes and it generates only one image.

3.5.3 Image Analysis on our model

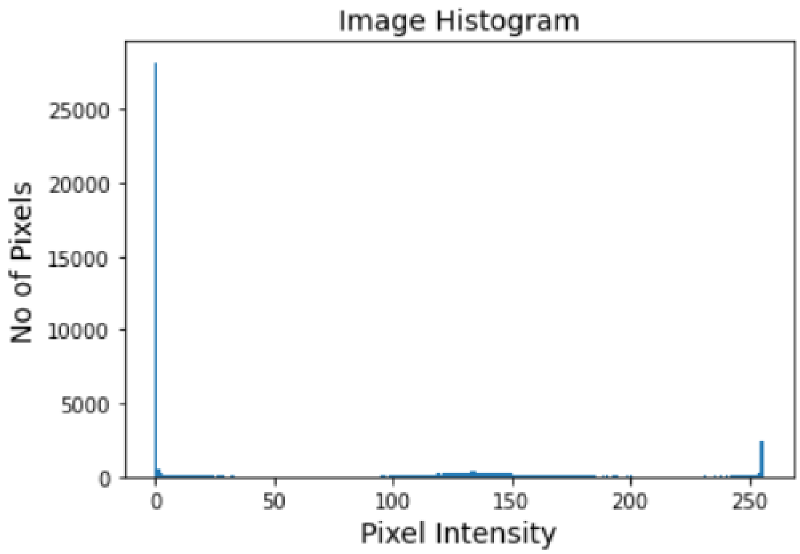

Fig 3.2 Histogram of pixel spread analysis in image dataset

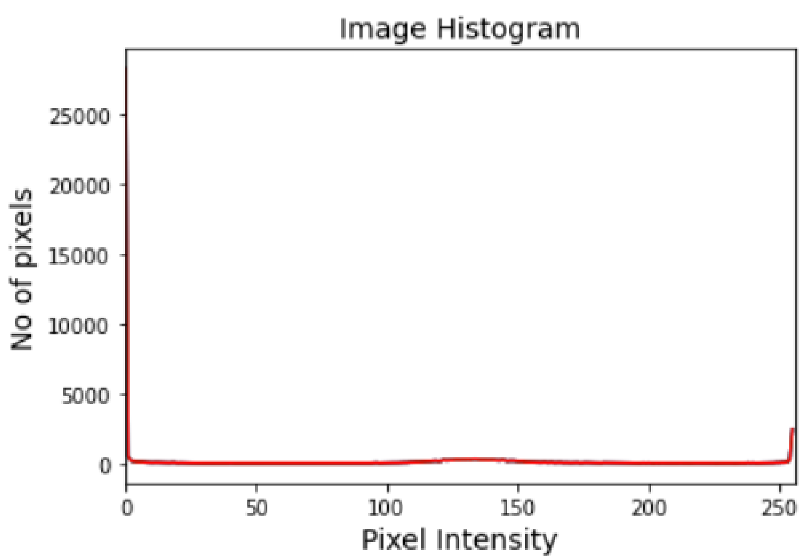

Fig 3.3 Histogram of pixel spread analysis in image dataset (BGR)

CHAPTER 4
DEEP LEARNING

Deep Learning is an artificial intelligence function that imitates the workings of the human brain in processing data and creating patterns for use in decision making. Deep learning is a subset of machine learning that has networks capable of learning unsupervised from data that is unstructured or unlabelled.[22]

Deep learning utilizes a hierarchical level of artificial neural networks to carry out the process of machine learning. The artificial neural networks are built like the human brain, with neuron nodes connected together like a web. While traditional programs build analysis with data in a linear way, the hierarchical function of deep learning systems enables machines to process data with a nonlinear approach.

In Deep Learning networks, each layer of nodes trains on a distinct set of features based on the previous layer's output. The further you advance into the neural net, the more complex the features your nodes can recognize, since they aggregate and recombine features from the previous layer. Deep Learning networks are distinguished from the single-hidden-layer neural networks by their depth, i.e. the number of node layers through which data must pass in a multistep process of pattern recognition [22].

Deep Learning then can be defined as neural networks with a large number of parameters and layers in one of four fundamental network architectures:

- Convolutional Neural Network
- Recurrent Neural Network
- Recursive Neural Network
- Pretrained Network

4.1 Convolutional Neural Network

Convolutional Neural Networks (CNN) is a type of deep neural network architecture designed for specific tasks like image classification. CNNs were inspired by the organization of neurons in the visual cortex of the animal brain. A CNN is composed of an input layer. For basic image processing, this input is a two-dimensional array of neurons which correspond to the pixels of an image. It also contains an output layer which is a one-dimensional set of output neurons. CNN uses a combination of sparsely connected convolutional layers, which perform image processing on their inputs. In addition, they contain down sampling layers called pooling layers to further reduce the number of neurons necessary in subsequent layers of the network. And finally, CNNs contain one or more fully connected layers to connect the pooling layer to the output layer. CNNs work well 22 for a variety of tasks including image recognition, image processing, image segmentation, video analysis, and natural language processing.

4.2 Recurrent Neural Network

The Recurrent Neural Network (RNN) can operate effectively on sequences of data with variable input length. This means that RNN uses knowledge of its previous state as an input for its current prediction, and this process can be repeated for an arbitrary number of steps allowing the network to propagate information via its hidden state through time. This feature makes RNNs very effective for working with sequences of data that occur over time. RNNs work well for applications that involve a sequence of data that change over time. These applications include natural language processing, speech recognition, language translation, image captioning and conversation modelling.

4.3 Recursive Neural Network

A recursive neural network is a kind of deep neural network created by applying the same set of weights recursively over a structured input, to produce a structured prediction over variable-size input structures or a scalar prediction on it by traversing a given structure in topological order. Recursive Neural Networks have been successful in learning sequence and tree structures in natural language processing, mainly phrase and sentence continuous representations based on word embedding.

4.4 Pretrained Network

A pretrained network is a model that was trained on a large benchmark dataset to solve a problem similar to the one that we want to solve. Accordingly, due to the computational cost of training such models, it is common practice to 23 import and use such pre-trained networks (e.g. Alexnet, squeezenet, googlenet). The process of importing and using a pre-trained network is referred to as transfer learning. Transfer learning is a popular method in computer vision because it allows us to build accurate models in a timesaving way. With transfer learning, instead of starting the learning process from scratch, it can be started from patterns that have been learned when solving a different problem.

CHAPTER 5

CNN MODEL

5.1 BUILDING A CNN MODEL

A CNN typically has three layers: a convolutional layer, a pooling layer, and a fully connected layer. Figure 5.1 shows an image of the architecture of CNN.

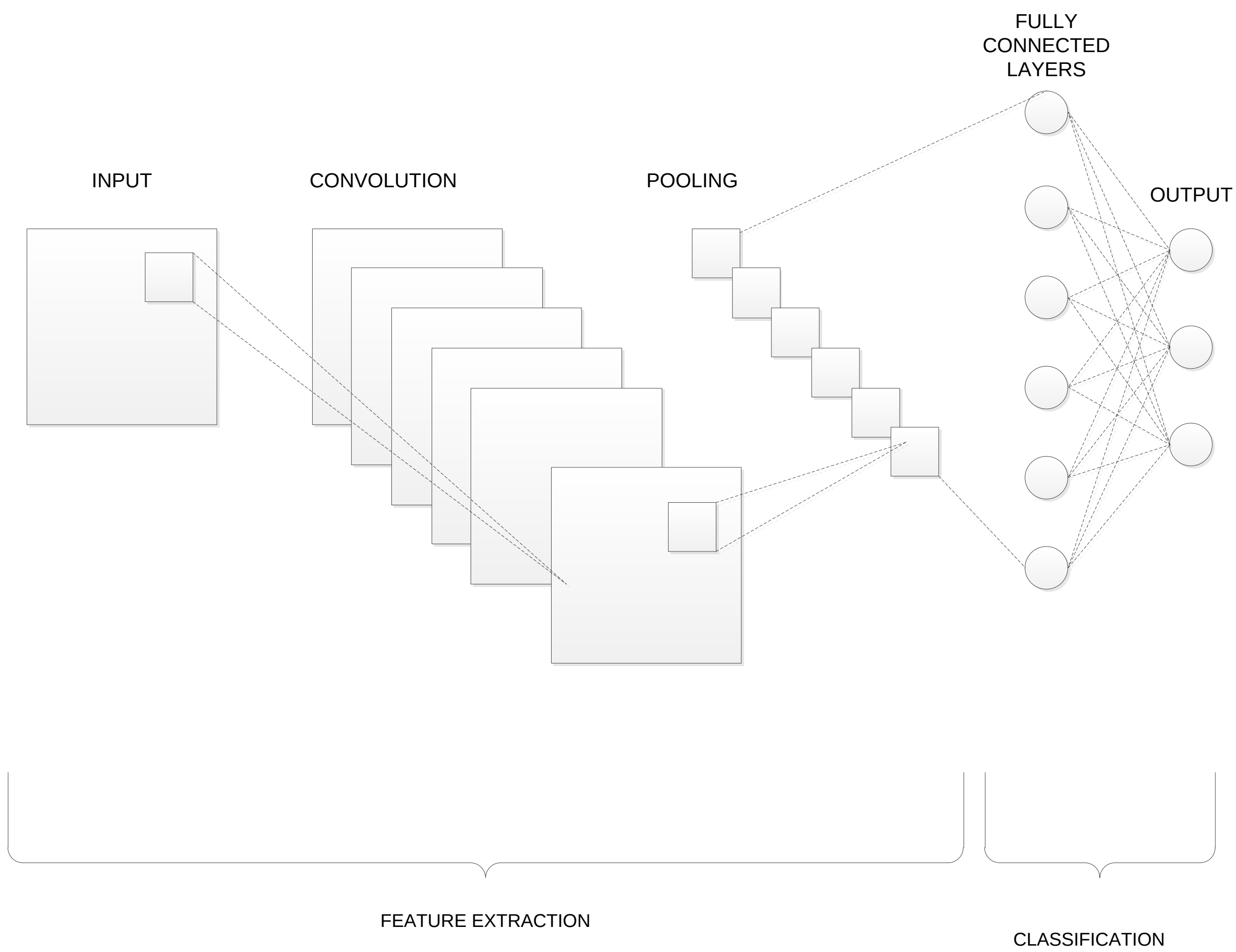

Fig 5.1 Architecture of Convolutional Neural Network (CNN)

5.1.1Convolution Layer

The core component of the CNN is the convolution layer. It carries the main computing load of the network.

This layer performs a dot product between two matrices, where one matrix is the kernel-set of learnable parameters, the other matrix the receptive part. The kernel is smaller spatially than a picture but more profound. That makes the kernel height and width spatially small if the image is made up of 3 (RGB) channels, however, the depth can be extended to all 3 channels.

The kernel slides across the image's height and width during the forward transfer, creating an image representation of that receptive area. The image is represented two-dimensionally as an activation map, which gives the kernel reaction at each image spatial position. The kernel's sliding size is known as a stride.

5.1.1.1Motivation behind Convolution

Convolution takes advantage of three main concepts of computer vision: sparse interaction, parameter sharing, and equivariant representation.

Matrix multiplication by a matrix of parameters representing the relationship between the input and output unit is used in trivial neural network layers. This implies that each output unit has a relationship with each input unit. Convolution neural networks, on the other hand, have sparse interaction. This is done by having the kernel smaller than the input; for example, an image can have millions or thousands of pixels, but by processing it with the kernel, we can detect relevant information in tens or hundreds of pixels. This means that we need to store fewer parameters, which decreases the model's memory requirements while also improving its statistical performance.[23]

If computing one feature at a spatial point (x1, y1) is useful then it should also be useful at some other spatial point say (x2, y2). It means that for a single two-dimensional slice i.e., for creating one activation map, neurons are constrained to use the same set of weights. In a traditional neural network, each element of the weight matrix is used once and then never revisited, while convolution network has shared parameters i.e., for getting output, weights applied to one input are the same as the weight applied elsewhere.

Due to parameter sharing, the layers of convolution neural network will have a property of equivariance to translation. It says that if we changed the input in a way, the output will also get changed in the same way.

5.1.2 Pooling Layer

The pooling layer uses a summary statistic of neighbouring outputs to replace the network's output at specific locations. This reduces the representation's spatial scale, which reduces the amount of computation and weights necessary. Every slice of the representation is individually processed during the pooling operation.[23]

The rectangular neighbourhood average, the L2 norm of the rectangular neighbourhood, and a weighted average based on the distance from the central pixel are all pooling functions. The most common method, however, is max pooling, which reports the neighbourhood's maximum performance. Figure 5.2 shows an image of max pooling.

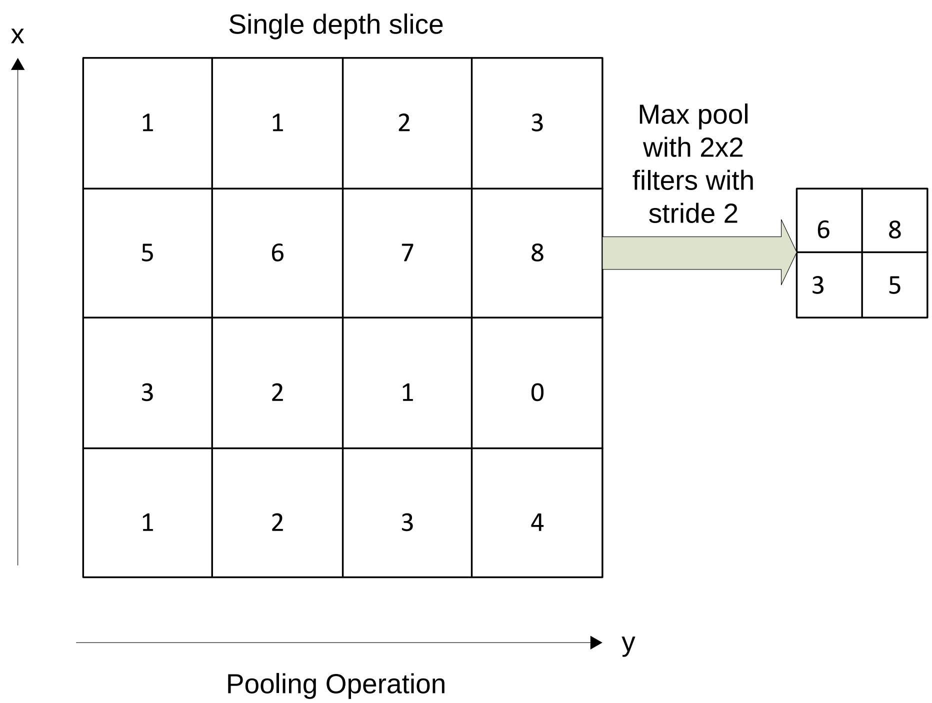

Fig 5.2 Pooling Operation

Pooling provides some translation invariance in all situations, ensuring that an entity can be identified regardless of where it appears on the frame.

5.1.3 Fully Connected Layer

Neurons in this layer have full connectivity with all neurons in the preceding and succeeding layer as seen in regular FCNN. This is why it can be computed as usual by a matrix multiplication followed by a bias effect.

5.1.4 Non-Linearity Layers

Since convolution is a linear operation and images are far from linear, non-linearity layers are often placed directly after the convolutional layer to introduce non-linearity to the activation map.

5.1.4.1. Sigmoid

The sigmoid non-linearity has the mathematical form $\sigma(k) = \frac{1}{1+e^{-}k}$ .It takes a real-valued number and “squashes” it into a range between 0 and 1. However, a very undesirable property of sigmoid is that when the activation is at either tail, the gradient becomes almost zero. If the local gradient becomes very small, then in backpropagation it will effectively “kill” the gradient. Also, if the data coming into the neuron is always positive, then the output of sigmoid will be either all positives or all negatives, resulting in a zig-zag dynamic of gradient updates for weight.

5.1.4.2. Tanh

Tanh squashes a real-valued number to the range [-1, 1]. Like sigmoid, the activation saturates, but unlike the sigmoid neurons its output is zero cantered.

5.1.4.3. ReLU

The Rectified Linear Unit (ReLU) has become very popular in the last few years. It computes the function $f(\kappa)=\max(0,\kappa)$. In other words, the activation is simply threshold at zero.

In comparison to sigmoid and tanh, ReLU is more reliable and accelerates the convergence by six times.

Unfortunately, a con is that ReLU can be fragile during training. A large gradient flowing through it can update it in such a way that the neuron will never get further updated. However, we can work with this by setting a proper learning rate.

CHAPTER 6
TRANSFER LEARNING

Transfer learning is the reuse of a pre-trained model on a new problem. It's currently very popular in deep learning because it can train deep neural networks with comparatively little data. This is very useful in the data science field since most real-world problems typically do not have millions of labelled data points to train such complex models. In transfer learning, a machine exploits the knowledge gained from a previous task to improve generalization about another. For example, in training a classifier to predict whether an image contains food, you could use the knowledge it gained during training to recognize drinks.[25]

The general idea is to adapt what a model has learned from a task with a large amount of labelled training data to a new task with less data. Because of the massive amount of computing power available, transfer learning is often used in computer vision and natural language processing tasks like sentiment analysis.

6.1 Working of transfer learning

In computer vision, for example, neural networks usually try to detect edges in the earlier layers, shapes in the middle layer and some task-specific features in the later layers. In transfer learning, the early and middle layers are used and we only retrain the latter layers. It helps leverage the labelled data of the task it was initially trained on.

For example, think of a model trained for recognizing a backpack on an image, which will be used to identify sunglasses. In the earlier layers, the model has learned to recognize objects, because of that we will only retrain the latter layers so it will learn what separates sunglasses from other objects.[25]

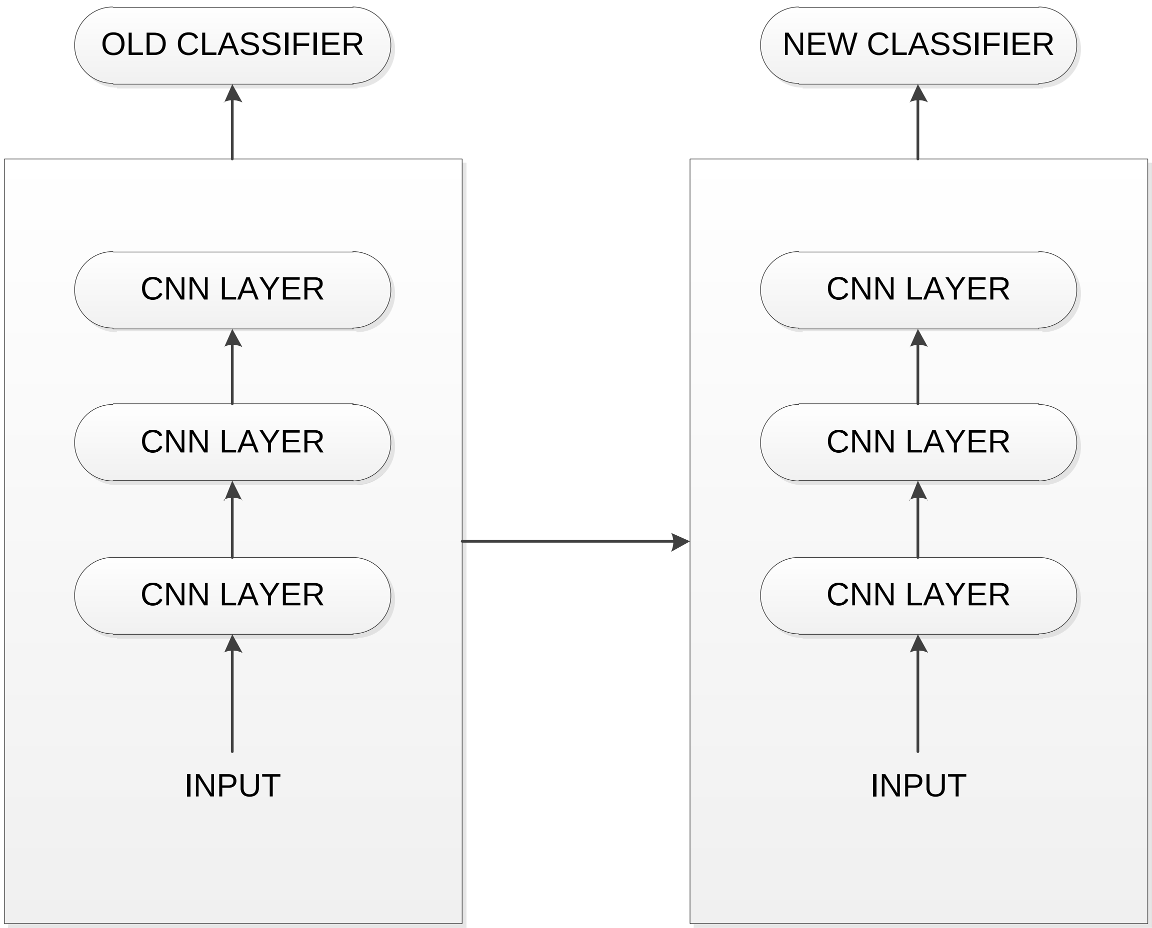

Fig 6.1 Transfer Learning

Figure 6.1 shows an image of transfer learning .In transfer learning, we try to transfer as much knowledge as possible from the previous task the model was trained on to the new task at hand. This knowledge can be in various forms depending on the problem and the data. For example, it could be how models are composed, which allows us to more easily identify novel objects.

6.1.1 Advantages of transfer learning

Transfer learning has a variety of benefits, the most important of which are reduced training time, enhanced neural network efficiency (in most cases), and the absence of a large amount of data.

To train a neural network from scratch, a lot of data is usually needed, but access to that data isn't always possible — this is where transfer learning comes in handy. Since the model has already been pre-trained, a solid machine learning model can be developed with very little training data using transfer learning. This is particularly useful in natural language processing, where large labelled datasets need a lot of expert knowledge. Additionally, training time is reduced because training a deep neural network from scratch on a complex task can take days or even weeks.

6.2 Approaches

6.2.1. Training a model to reuse it.

When we have a problem A and not enough data, one way around this is to find a related task B with an abundance of data. Train the deep neural network on task B and use the model as a starting point for solving task A. Whether you'll need to use the whole model or only a few layers depend heavily on the problem you're trying to solve.

If we have the same input in both tasks, possibly reusing the model and making predictions for your new input is an option. Alternatively, changing and retraining different task-specific layers and the output layer is a method to explore.

6.2.2 Using a pre trained model

The second approach is to use an already pre-trained model. There are a lot of these models out there, so make sure to do a little research. How many layers to reuse and how many to retrain depends on the problem.

Keras, for example, provides nine pre-trained models that can be used for transfer learning, prediction, feature extraction and fine-tuning. You can find these models, and also some brief tutorials on how to use them, here. There are also many research institutions that release trained models. This type of transfer learning is most commonly used throughout deep learning.

6.2.3 Feature Extraction

Another approach is to use deep learning to discover the best representation of your problem, which means finding the most important features. This approach is also known as representation learning, and can often result in a much better performance than can be obtained with hand-designed representation.

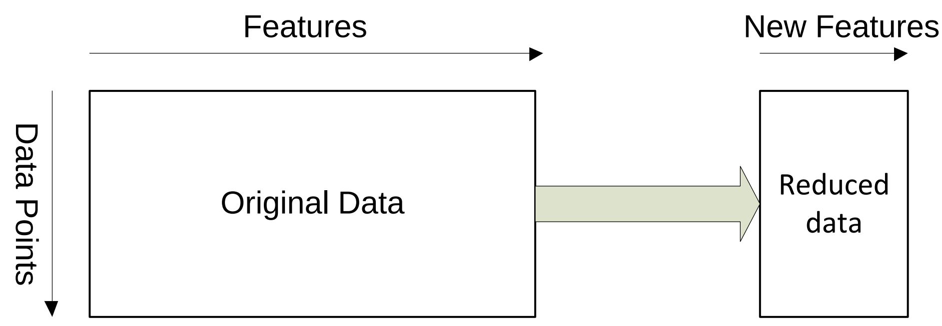

Fig 6.2 Feature Extraction

In machine learning, features are usually manually hand-crafted by researchers and domain experts. Fortunately, deep learning can extract features automatically. The figure 6.2 shows feature extraction using deep learning. Neural networks have the ability to learn which features are really important and which ones aren't. A representation learning algorithm can discover a good combination of features within a very short timeframe, even for complex tasks which would otherwise require a lot of human effort.

The learned representation can then be used for other problems as well. Simply use the first layers to spot the right representation of features, but don't use the output of the network because it is too task-specific. Instead, feed data into your network and use one of the intermediate layers as the output layer. This layer can then be interpreted as a representation of the raw data.

This approach is mostly used in computer vision because it can reduce the size of your dataset, which decreases computation time and makes it more suitable for traditional algorithms, as well.

6.3 Alexnet Architecture

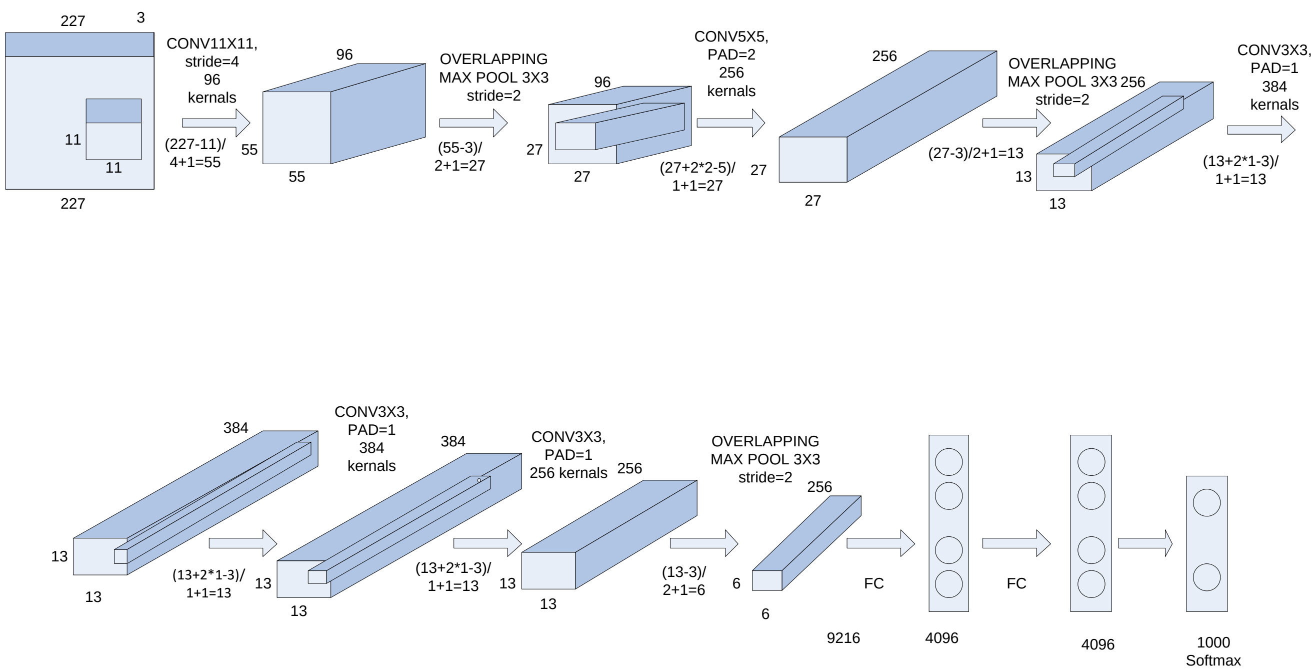

Fig 6.3 Architecture of Alexnet

The figure 6.3 shows the architecture of Alexnet. AlexNet contains total of eight layers. AlexNet architecture consists of 5 convolutional layers, 3 max-pooling layers, 2 normalization layers, 2 fully connected layers, and 1 soft max layer. Each convolutional layer consists of convolutional filters and a nonlinear activation function ReLU. For the first two convolutional layers, each convolutional layers is followed by a Overlapping Max Pooling layer. The pooling layers are used to perform max pooling. Third, fourth and fifth convolution layers are directly connected with each other. The fifth convolutional layer is followed by Overlapping Max Pooling Layer, which is then connected to fully connected layers. The fully connected layers have 4096 neurons each and the second fully connected layer is feed into a softmax classifier having 1000 classes. Input size is fixed due to the presence of fully connected layers. The input size is mentioned at most of the places as 224x224x3 but due to some padding which happens it works out to be 227x227x3. AlexNet overall has 60 million parameters.[24]

6.3.1 Overlapping Pooling:

Max Pooling layers help to down-sample an input representation (image, hidden-layer output matrix, etc.), reducing its dimensionality and allowing for assumptions to be made about features contained in the sub-regions binned. Max Pooling helps to reduce overfitting. Basically, it uses a max operation to pool sets of features, leaving us with a smaller number of them. Max Pooling and Overlapping is same except except the adjacent windows over which the max is computed overlap each other.

Overlapping Max Pool layers are similar to Max Pool layers except the adjacent windows over which the max is calculated overlaps each other. The authors of AlexNet used pooling windows, sized 3×3 with a stride of 2 between the adjacent windows. Due to this overlapping nature of Max Pool, the top-1 error rate was reduced by 0.4% and top-5 error rate was reduced by 0.3% respectively. If you compare this to using a non-overlapping pooling windows of size 2×2 with a stride of 2, that would give the same output dimensions.

The overlapping pooling reduces the top-1 and top-5 error rates by 0.4% and 0.3% compared to non-overlapping pooling, thus finding it very difficult to overfit.

6.3.2 ReLU Nonlinearity

Using ReLU non-linearity, AlexNet shows us that deep CNN's can be trained much faster with the help of saturating activation functions such as Tanh or Sigmoid. The figure shown below shows us that with the help of ReLUs(solid curve), AlexNet can achieve a 25% training error rate.

CHAPTER 7

TRAINING THE MODEL

7.1 Training using Alexnet

Different characteristics of pre-trained networks are important to consider when selecting a network to apply to a particular problem. The accuracy, speed, and size of the network are the most important characteristics. Choosing a network is normally a balance between these variables.

Table 7.1 No. of convolutional layers in various pre trained networks

Name of Deep Neural Network	**No. of convolutional layers**
Alexnet	8
VGG16	16
VGG19	19
Googlenet	22
Resnet50	50
Resnet101	101
Resnet152	152

A network with a short training and prediction time and an appropriate degree of accuracy would be a good choice. Alexnet is a network of only eight convolutional layers, making it simple to retrain for the problem at hand and providing better accuracy.

Alexnet prediction is based on 64 million parameters. Since using the only framework of the network will be similar to building the network from scratch and will consume a lot of time, we used transfer learning to train Alexnet.

The data sets for the various intracranial bleeding can be found in the kaggle archive. The dataset is then resized to 227x227 pixels, which is the normal size for alexnet input. After that, the dataset is split into two parts: 80 percent for training and 20% for validation. Alexnet, the pre-trained network, has been loaded. The layers are rearranged to identify the five different types of intracranial bleeding. The updated network was trained over 100 epochs, each with 30 iterations. The network had a 75.10 percent accuracy on average. This network has been preserved and will be used for research purposes.

7.1.1 RESULTS

The workspace should be loaded with the pretrained network first. According to the problem at hand, the network's final layers (fully connected layer and classification layer) must be replaced. The network must be conditioned for a certain number of epochs after the final layers are replaced. Within a single epoch, a certain number of iterations take place. The weights assigned to each convolutional layer are calibrated for each iteration in order to minimise prediction error.

The qualified network's accuracy is then predicted by counting the number of accurate predictions it makes. The network is then saved and tested and the accuracy can be further improved by retraining the saved network.

7.1.1.1 TRAIN SET

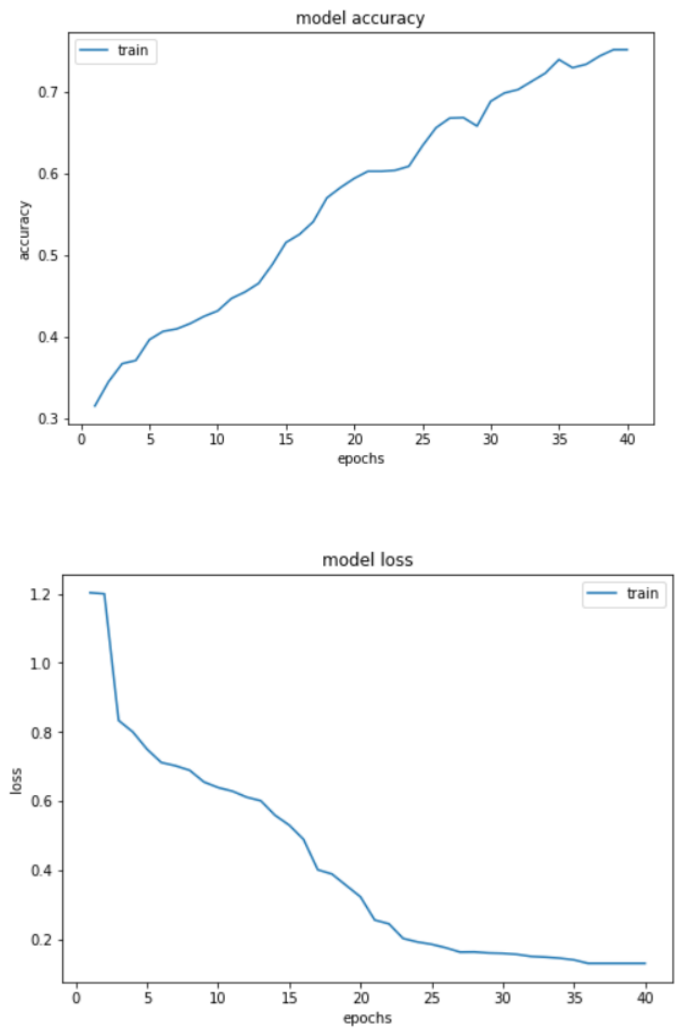

Fig 7.1 Train accuracy and loss of Alexnet Model

7.1.1.2 CONFUSION MATRIX

A confusion matrix, also known as an error matrix, is a table structure that enables the output of an algorithm to be visualised. The real class is represented by the columns of the 36 uncertainty matrix, while the projected class is represented by the rows. Correctly classified observations are represented by diagonal cells, whereas incorrectly classified observations are represented by off-diagonal cells.

Fig 7.3 Test accuracy of Alexnet Model

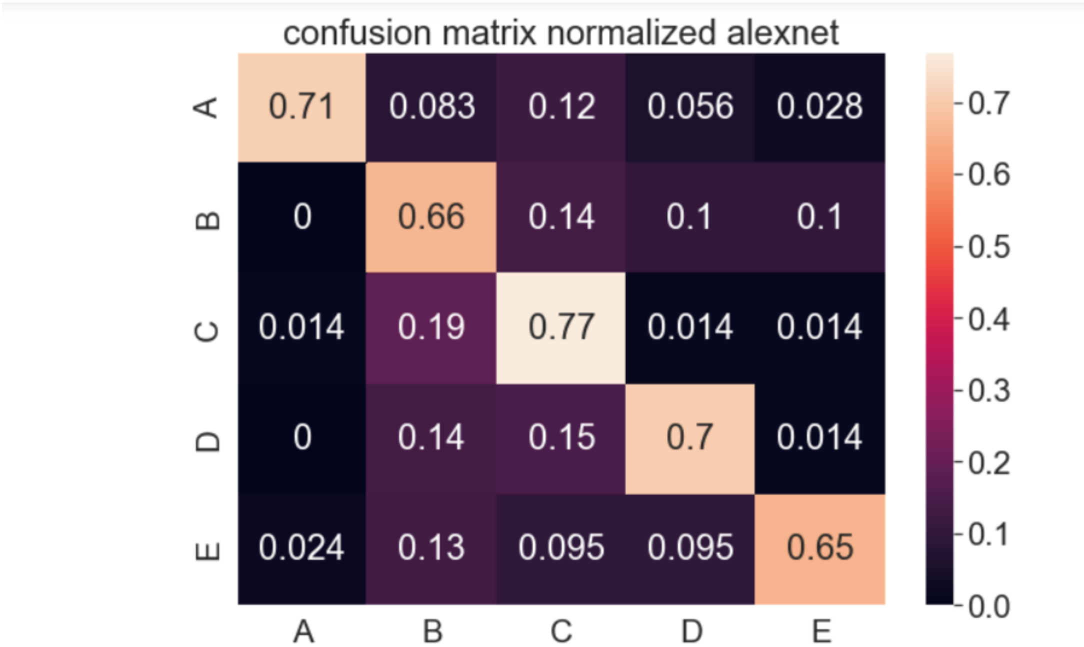

Fig. 7.2 Confusion matrix for alexnet

7.1.1.3 TEST SET

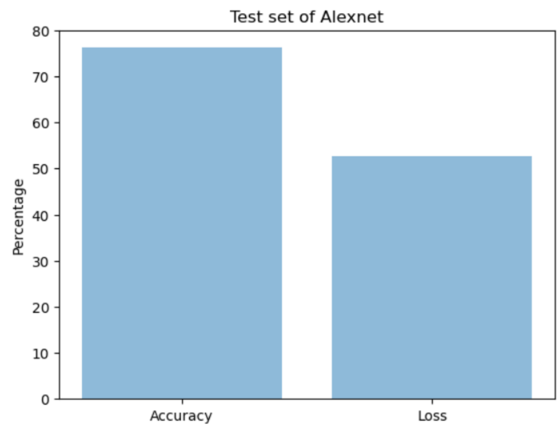

Fig 7.3 Test accuracy and loss of Alexnet Model

7.1.1.4 Output

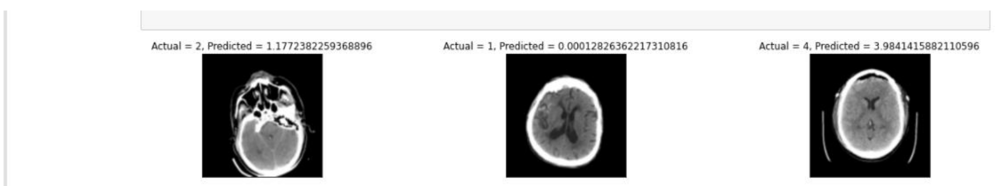

Fig 7.4 Output from testing a Alexnet Model

7.2 Training using CNN

The training of CNN model is similar to training of Alexnet. The CNN network should be pre-loaded into the workspace. The number of dense layers and pooling layers has been increased. The increased number of pooling layers and dense layers increases the network's performance by extracting more features from the dataset. The network also learns faster with the updated CNN network. After the final layers are replaced, the network must be conditioned for a certain number of epochs. There are a certain number of iterations in a single epoch. To reduce prediction error, the weights assigned to each convolutional layer are adjusted for each iteration.

The dataset is divided into two sections: 80% for training and 20% for validation. The CNN network has been updated and loaded. To distinguish the five different forms of intracranial bleeding, the layers are increased. Over 20 epochs, each with 35 iterations, the updated network was trained. On average, the network was 82.00 percent effective.

7.2.1 RESULTS

The accuracy of the eligible network is then estimated by counting the number of correct predictions it makes. The network is then saved and checked, with the accuracy of the saved network being improved further by retraining it.

7.2.1.1 TRAIN SET

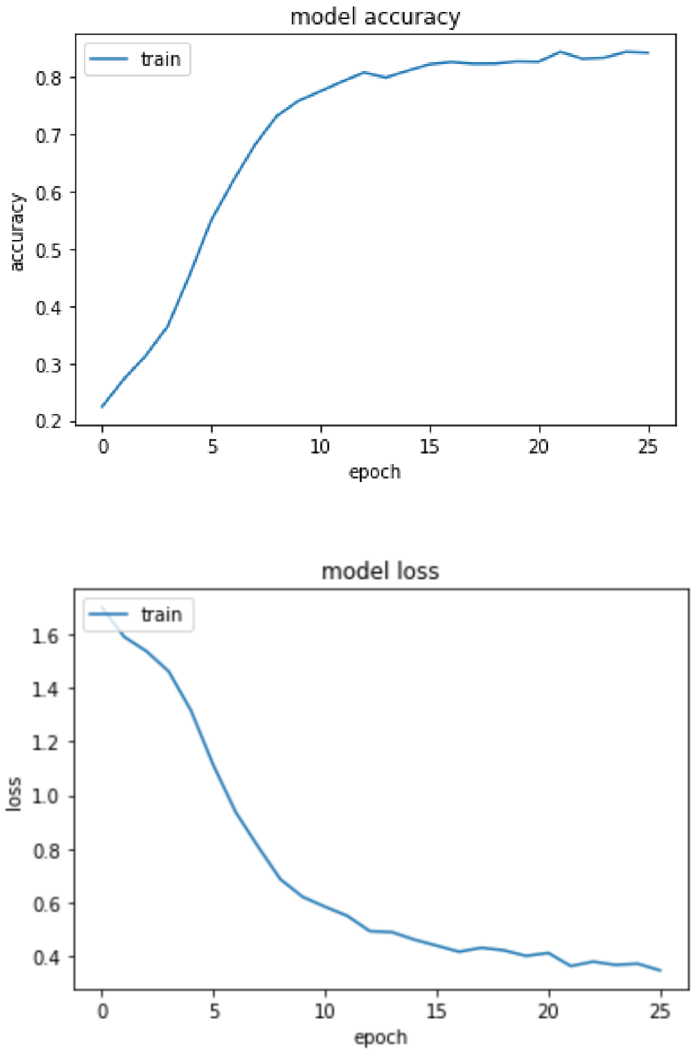

Fig 7.5 Train accuracy and loss of Modified CNN Model

7.2.1.2 CONFUSION MATRIX

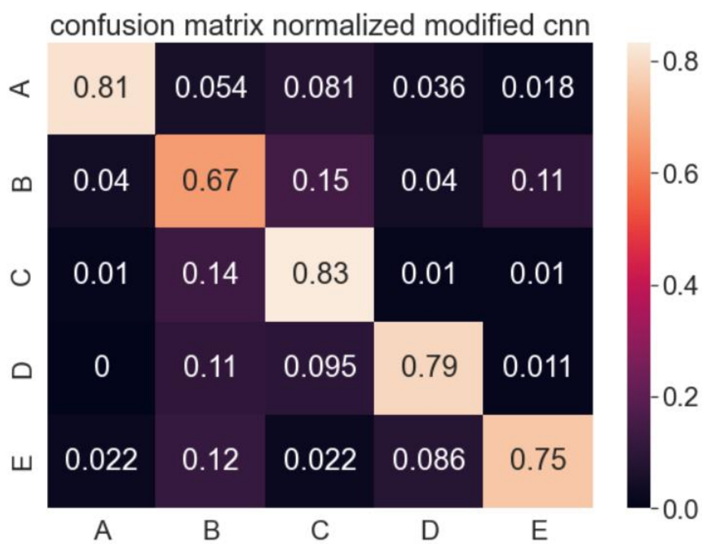

Fig. 7.6 Confusion matrix for alexnet

7.2.1.3 TEST SET

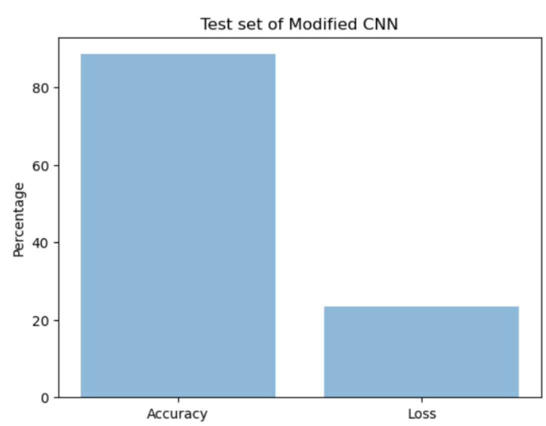

Fig 7.6 Test accuracy of Modified CNN Model

7.2.1.4 Output

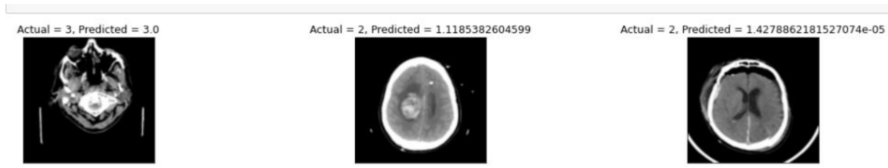

Fig 7.7 Output from testing a Modified CNN Model

CHAPTER 8

CONCLUSION

The proposed system for identification of Intracranial Hemorrhages using an effective Neural Network successfully compared pre-trained networks and an updated CNN network learning technique in classifying five types of intracranial haemorrhages. Both models were trained using data obtained from the Kaggle database. In terms of precision, the updated CNN model outperformed the Alexnet model.

Trauma or a head injury are the most common causes of intracranial bleeding. It's a severe health problem that necessitates immediate and intense medical attention, so detecting brain bleeding at an early stage is critical. The aim of the project is to use a simple CT scan and a high-efficiency neural network to detect bleeding within the brain. The network has been qualified to diagnose bleeding and identify the form of haemorrhage, assisting doctors in detecting bleeds earlier and providing prompt treatment to patients.

Overall, the initiative makes diagnosis and care simpler. The neural network's efficiency can be improved in the future by incorporating an LSTM layer into the CNN network, allowing the model to make more precise diagnoses and classifications of intracranial bleeding.

REFERENCES

1] M. Li, L. Kuang, S. Xu and Z. Sha, "Brain Tumor Detection Based on Multimodal Information Fusion and Convolutional Neural Network," in *IEEE Access*, vol. 7, pp. 180134-180146, 2019,
doi: 10.1109/ACCESS.2019.2958370.

2] Y. Liu *et al.*, "Deep C-LSTM Neural Network for Epileptic Seizure and Tumor Detection Using High-Dimension EEG Signals," in *IEEE Access*, vol. 8, pp. 37495-37504, 2020,
doi: 10.1109/ACCESS.2020.2976156.

3] P. Kumar Mallick, S. H. Ryu, S. K. Satapathy, S. Mishra, G. N. Nguyen and P. Tiwari, "Brain MRI Image Classification for Cancer Detection Using Deep Wavelet Autoencoder-Based Deep Neural Network," in IEEE Access, vol. 7, pp. 46278-46287, 2019,
doi: 10.1109/ACCESS.2019.2902252.

4] H. H. Sultan, N. M. Salem and W. Al-Atabany, "Multi-Classification of Brain Tumor Images Using Deep Neural Network," in *IEEE Access*, vol. 7, pp. 69215-69225, 2019,
doi: 10.1109/ACCESS.2019.2919122.

5] J. Zheng, D. Lin, Z. Gao, S. Wang, M. He and J. Fan, "Deep Learning Assisted Efficient AdaBoost Algorithm for Breast Cancer Detection and Early Diagnosis," in

IEEE Access, vol. 8, pp. 96946-96954, 2020,
doi:10.1109/ACCESS.2020.2993536.

6] Singh SP, Wang L, Gupta S, Goli H, Padmanabhan P, Gulyás B. 3D Deep Learning on Medical Images: A Review. Sensors (Basel). 2020 Sep 7;20(18):5097. doi: 10.3390/s20185097. PMID: 32906819; PMCID: PMC7570704.

7]Woźniak, M., Siłka, J. & Wieczorek, M. Deep neural network correlation learning mechanism for CT brain tumor detection. Neural Comput & Applic (2021).

8] T. Hossain, F. S. Shishir, M. Ashraf, M. A. Al Nasim and F. Muhammad Shah, "Brain Tumor Detection Using Convolutional Neural Network," 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), Dhaka, Bangladesh, 2019, pp. 1-6, doi: 10.1109/ICASERT.2019.8934561.

9] Gumaei, M. M. Hassan, M. R. Hassan, A. Alelaiwi and G. Fortino, "A Hybrid Feature Extraction Method With Regularized Extreme Learning Machine for Brain Tumor Classification," in IEEE Access, vol. 7, pp. 36266-36273, 2019, doi: 10.1109/ACCESS.2019.2904145.

10] P. M. S. Kumar and S. Chatteijee published a paper titled "Computer assisted diagnostic for cancer diagnosis using MRI images of the brain (Brain tumour detection and classification system)," which was presented at the 2016 IEEE Annual India Conference (INDICON) in Bangalore, India.

11] M. Gurbină, M. Lascu and D. Lascu, "Tumor Detection and Classification of MRI Brain Image using Different Wavelet Transforms and SVMs," 2019 42nd International Conference on Telecommunications and Signal Processing (TSP), Budapest, Hungary, 2019, pp. 505-508, doi: 10.1109/TSP.2019.8769040.

12] Machine Learning Approach enhances the Quality of the MRI Images in Tumor Detection and Diagnosis: A PSO dependent Cluster Analysis" was published by V. Sivakumar, N. Janakiraman, and S. Naganandhini.

13] C. Han *et al.*, "Combining Noise-to-Image and Image-to-Image GANs: Brain MR Image Augmentation for Tumor Detection," in *IEEE Access*, vol. 7, pp. 156966-156977, 2019, doi: 10.1109/ACCESS.2019.2947606.

14] W. Wang, F. Bu, Z. Lin and S. Zhai, "Learning Methods of Convolutional Neural Network Combined With Image Feature Extraction in Brain Tumor Detection," in IEEE Access, vol. 8, pp. 152659-152668, 2020, doi: 10.1109/ACCESS.2020.3016282.

15] G. Manogaran, P. M. Shakeel, A. S. Hassanein, P. Malarvizhi Kumar and G. Chandra Babu, "Machine Learning Approach-Based Gamma Distribution for Brain Tumor Detection and Data Sample Imbalance Analysis," in IEEE Access, vol. 7, pp. 12-19, 2019, doi: 10.1109/ACCESS.2018.2878276.

16] Ye, H., Gao, F., Yin, Y. et al. Precise diagnosis of intracranial hemorrhage and subtypes using a three-dimensional joint convolutional and recurrent neural network. Eur Radiol 29, 6191–6201 (2019).

17]Patel, Ajay & van de Leemput, Sil & Prokop, Mathias & Ginneken, Bram & Manniesing, Rashindra. (2019). Image Level Training and Prediction: Intracranial Hemorrhage Identification in 3D Non-Contrast CT. IEEE Access. PP. 1-1. 10.1109/ACCESS.2019.2927792.

18] https://www.healthline.com/health/extradural-hemorrhage

19] https://towardsdatascience.com/pre-processing-in-natural-language-machine-learning-898a84b8bd47

20] https://towardsdatascience.com/data-preprocessing-and-network-building-in-cnn-15624ef3a28b

21] https://www.javatpoint.com/dip-histograms

22] https://www.aionlinecourse.com/tutorial/machine-learning/deep-learning

23] https://www.analyticsvidhya.com/blog/2020/10/what-is-the-convolutional-neural-network-architecture/

24] https://towardsdatascience.com/alexnet-the-architecture-that-challenged-cnns-e406d5297951

25] https://machinelearningmastery.com/transfer-learning-for-deep-learning/

26] https://www.kaggle.com/