Table of Contents:

Abbreviations

WHO	World Health Organization
ML	Machine Learning
DL	Deep Learning
AI	Artificial Intelligence
RL	Reinforcement Learning
ANN	Artificial Neural Network
ReLU	Rectified Linear Unit Function
CNN/ConvNet	Convolutional Neural Networks
RCNN	Region Based Convolutional Neural Networks
Fast RCNN	Fast Region Based Convolutional Neural Networks
SVM	Support Vector Machine
ROI	Region Of Interest
SSD	Single Shot detector
RPN	Region Proposal Network
IoU	Intersection-Over-Union
AR	Augmented Reality
VR	Virtual Reality
MR	Mixed Reality
STT	Speech to Text
TTS	Text to Speech
API	Application Programming Interface

SDK	Software Development Kit
ASL	American Sign Language

Chapter One

1. Introduction

1.1 History

The miscommunication between deaf people and normal people made a psychological gap so in 1892 electrical hearing aids were invented such as the ear trumpet, a funnel-shaped device which collects sound waves and leads them into the ear.

Since then, technology has improved our hearing abilities with digital hearing aids and cochlear implants. In 1985 the cochlear implant was approved for people aged 18 and older.**[1]**

1.2 Problem Statement

There is a communication gap between the normal people and deaf people. Deaf people suffer from misunderstanding of normal people as they communicate by sign language while normal people communicate with their normal language. Normal people can’t understand sign language as it is not a traded language and it is difficult to learn. We can't communicate with just the language but we need some gestures to understand each other.

- **Sign Language Features:**

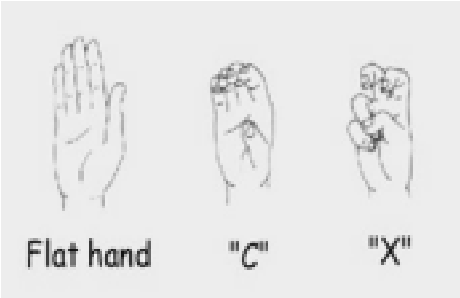
  Sign Language is used by Deaf people to communicate with each other . Sign Language is a full featured language (with lexicons, a grammar, etc) and it is the most advanced form of gestural communication.

- **Sign Composition:**

  Any manual sign can be broken up into four parameters.

Each of these parameters is independent of each other, and is dynamic or static during a sign, Static gesture is used for alphabet and number representation, whereas dynamic gesture is used for specific concepts. Dynamic also includes words, sentences etc. Static gesture consists of poses of hand, whereas the latter include motion of hands, head or both.

1- The Handshape : is defined by fingers and palm.



Figure(1): Examples of handshapes

2- The Orientation : is defined by two axes of the hand.

3- The Movement : is the hand trajectory (line, circle, curve, etc...).

4- The Location : is the hand position in relation to the body , The location is mainly used to express spatial and temporal information or relationship and the location granularity is more or less according to needs. Each of the four parameters carries information and is part of the sign meaning.

1.3 Problems that are facing the Deaf-Mute people

1. Social Effects :

- Reducing the misCommunication with normal people.
- Reducing the bullying on deaf people.

2. Environmental Effects:

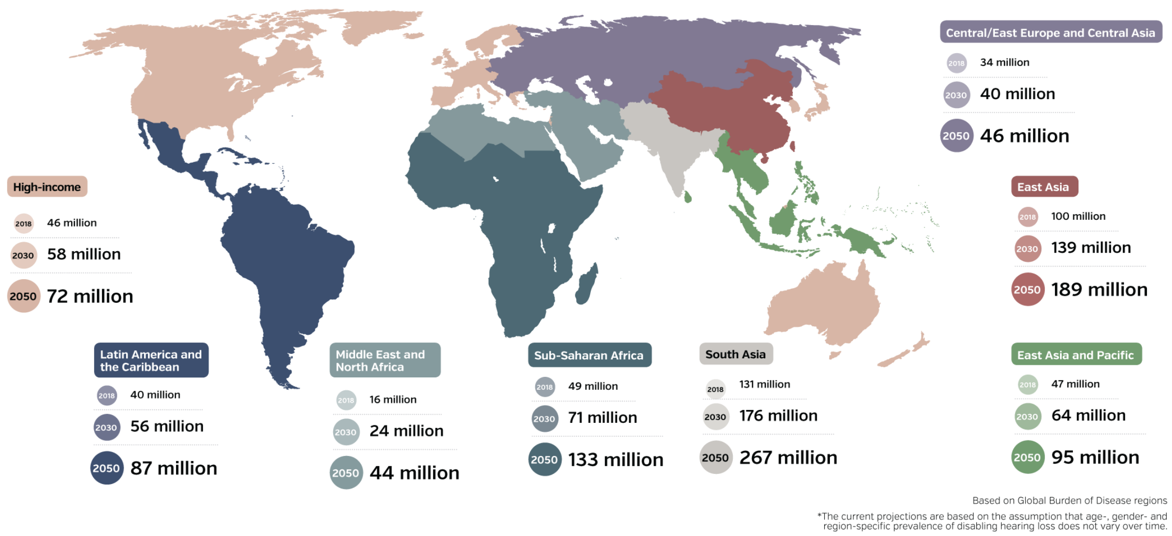
- Remote communication (COVID-19 effect).

3. Commercial Effects:

- Solve some educational problems.
- Reducing unemployment.

1.4 Facts about Deaf People

According to WHO , Over 5% of the world's population – or 430 million people – require rehabilitation to address their 'disabling' hearing loss (432 million adults and 34 million children). It is estimated that by 2050 over 700 million people – or one in every ten people – will have disabling hearing loss.**[2]**



Figure(2):Projected number of people with hearing loss in different world regions until 2050

The map shows the current and projected number of people with hearing loss in different regions.

Projections show that the number of people with disabling hearing loss will increase in all regions.**[3]**

1.5 Project Description

- BI Translator is a mobile application that has two phases to improve the communication between normal people and deaf people.
- The first phase helps normal people to understand the deaf people by translating signs into text using Deep Learning models.
- The second phase helps deaf people to understand normal people by translating speech into signs using Augmented Reality and Graphic Motion.

1.6 Objectives and Goals

- To facilitate the communication between the Deaf-Mute people and the Normal people.
- Reducing the social effects that affect the Deaf-Mute people.
- Facilitate the communication between the lip readers (deaf people who read the lips to communicate) and normal people during Covid-19 by using our translator.
- Minimum cost.

1.7 Overview

Chapter (2) Machine Learning : in this chapter we propose a brief introduction of machine learning showing its life cycle , uses and its various fields: supervised, unsupervised and reinforcement learning, and giving an introduction for each one of them.

Chapter (3) Deep Learning : This chapter shows the difference between machine learning and deep learning ,The details of neural networks and the explanation of various deep learning algorithms such as CNN , Fast R-CNN.

Chapter (4) Animation : This chapter shows the definition of animation , its uses and its types.

Chapter (5) Augmented Reality: This chapter shows the differences of AR , VR and MR, AR Benefits and how we used it.

Chapter (6) Cloud : This chapter describes what is cloud and its types and also describes the cloud services we used.

Chapter (7) Previous Solutions : This chapter discusses the previous solutions that were used before to help the deaf people.

Chapter (8) proposed Methods : This chapter discusses our solution and its diagrams and models used to translate sign to text, the chapter shows the flowcharts and pseudo codes for each model Also shows the two ways we used to translate speech to sign.

Chapter (9) Results and Discussion: shows our results and discusses them.

Chapter (10) Conclusion and Future work: shows the whole conclusion and future work that can be done to improve the system.

Chapter (11) References: provides the references that are used in our project.

Chapter Two

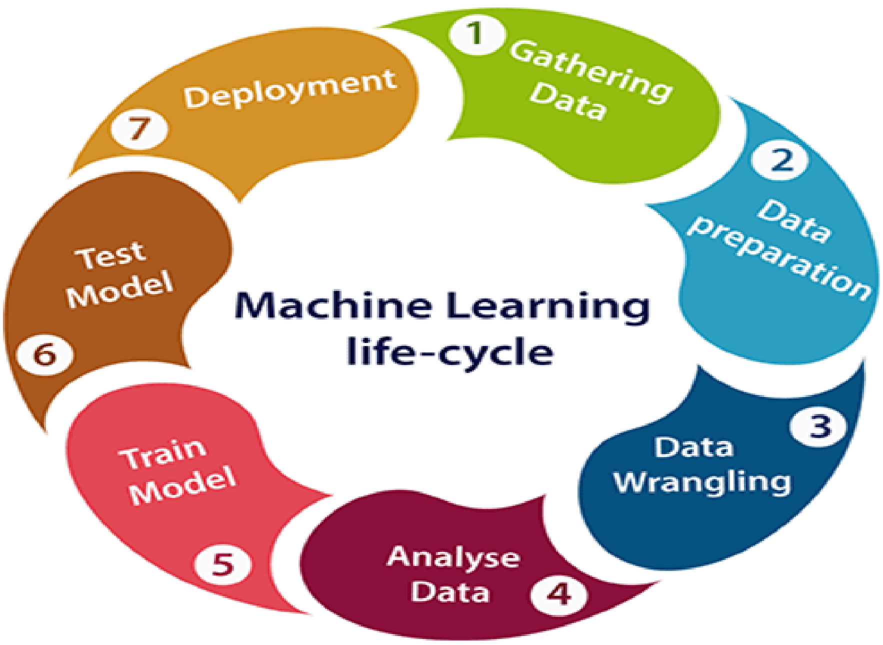
2. Machine Learning

Machine learning is one modern innovation that has helped man enhance not only many industrial and professional processes but also advances everyday living from driving cars to translating speech.

2.1 ML Definition

Machine learning is a branch of artificial intelligence(AI) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy.**[4]**

2.2 ML life cycle



Figure(3): ML life cycle [5]

1) **Gathering Data:**

Data Gathering is the first step of the machine learning life cycle. The goal of this step is to identify and obtain all data-related problems. data can be collected from various sources such as **files**, **database**, **internet**, or **mobile devices**.

This step includes the below tasks:

- Identify various data sources
- Collect data
- Integrate the data obtained from different sources

After performing these tasks , we will get a coherent set of data which is called a **dataset**

2) Data Preparation:

Data preparation is a step where we put our data into a suitable place and prepare it to use in our machine learning training.

In this step, first, we put all data together, and then randomize the ordering of data.

This step can be further divided into two processes:

- **Data exploration:**
  It is used to understand the nature of data that we have to work with. We need to understand the characteristics, format, and quality of data.
  A better understanding of data leads to an effective outcome. In this, we find Correlations, general trends, and outliers.

- **Data pre-processing:**
  Now the next step is preprocessing of data for its analysis.

3) Data Wrangling

Data wrangling is the process of cleaning and converting raw data into a usable format. It is the process of cleaning the data, selecting the variable to use, and transforming the data in a proper format to make it more suitable for analysis in the next step. It is one of the most important steps of the complete process. Cleaning of data is required to address the quality issues.

In real-world applications, collected data may have various issues, including:

- Missing Values
- Duplicate data
- Invalid data
- Noise

4) Data Analysis:

This step involves:

- Selection of analytical techniques
- Building models
- Review the result

The aim of this step is to build a machine learning model to analyze the data using various analytical techniques and review the outcome. It starts with the determination of the type of the problems, where we select the machine learning techniques such as Classification, Regression, Cluster analysis, Association, etc. then build the model using prepared data, and evaluate the model.

5) Train Model:

In this step we train our model to improve its performance for better outcome of the problem.

We use datasets to train the model using various machine learning algorithms. Training a model is required so that it can understand the various patterns, rules, and features.

6) Test Model:

In this step, we check for the accuracy of our model by providing a test dataset to it.

Testing the model determines the percentage accuracy of the model as per the requirement of project or problem.

7) Deployment :

The last step of the machine learning life cycle is deployment, where we deploy the model in the real-world system.

If the above-prepared model is producing an accurate result as per our requirement with acceptable speed, then we deploy the model in the real system. But before deploying the project, we will check whether it is improving its performance using available data or not. The deployment phase is similar to making the final report for a project.

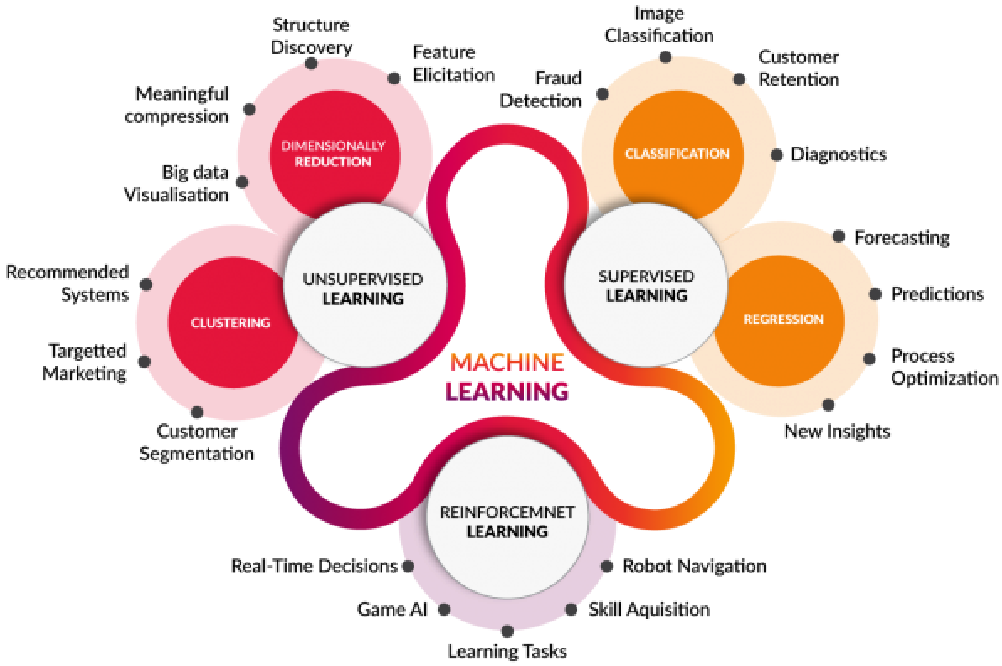
2.3 ML Uses

ML is used to build algorithms that can receive the input data and use statistical analysis to predict the output so there are limitless applications of machine learning. They are available in every form from simple to highly complex . The system can perform the following tasks by Machine Learning:

- **Image Recognition**
- **Speech Recognition**
- **Medical diagnosis**
- **Statistical Arbitrage**
- **Learning associations**
- **Classification**

- **Prediction**
- **Extraction**
- **Regression**
- **Financial Services.**

2.4 ML Algorithms Types



Figure(4): ML Algorithms Types [6]

2.4.1 Supervised Learning

Supervised learning is defined by its use of labeled datasets to train algorithms that to classify data or predict outcomes accurately. It uses a training set to teach models to yield the desired output. This training dataset includes inputs and correct outputs, which allow the model to learn over time.

The objective of a supervised learning model is to predict the correct label for newly presented input data a supervised learning algorithm can be written simply as:

$$Y = f(x)$$

The function used to connect input features to a predicted output is created by the machine learning model during training.

Types of Supervised Learning:

Supervised learning can be separated into two types of problems when data mining which are classification and regression.

- ## Classification

Classification: A classification problem is when the output variable is a category, such as “red” or “blue” or “disease” and “no disease” .**[7]**

Classification algorithm examples:

- **Binary Classification.**
    - **Logistic Regression.**
    - **k-Nearest Neighbors.**
    - **Decision Trees.**
    - **Support Vector Machine.**
    - **Naive Bayes.**
- **Multi-Class Classification.**
    - **Plant species classification.**
    - **Face classification.**

- ## Regression

Regression analysis is the process of estimating the relationship between a dependent variable and independent variables.

Regression algorithms examples :

- **Linear regression.**
- **Logistic regression.**
- **Ridge regression.**
- **Lasso regression.**
- **Polynomial regression.**

2.4.2 Unsupervised Learning

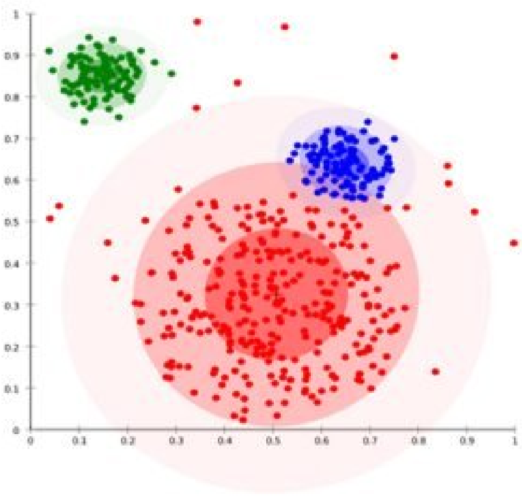
Unsupervised learning uses machine learning algorithms to analyze and cluster unlabeled datasets. These algorithms discover hidden patterns or data groupings without the need for human intervention. Its ability to discover similarities and differences in information make it the ideal solution for exploratory data analysis, cross-selling strategies, customer segmentation, and image recognition.**[8]**

Examples of unsupervised learning

1) **Clustering:**it is a data mining technique which groups unlabeled data based on their similarities or differences. Clustering algorithms are used to process raw, unclassified data objects into groups represented by structures or patterns in the information. Clustering algorithms can be categorized into a few types, specifically exclusive, overlapping, hierarchical, and probabilistic.

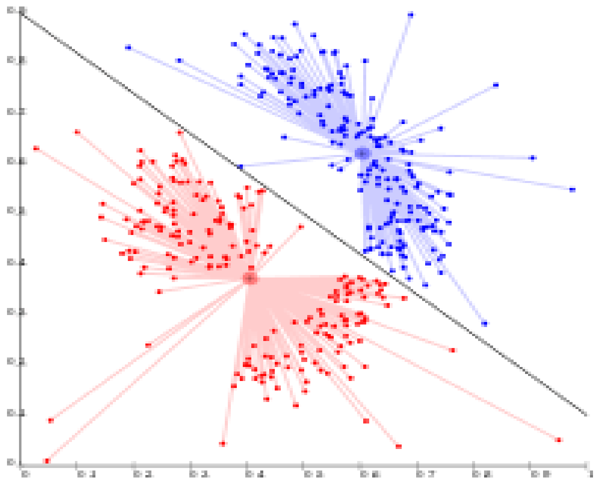
Popular Clustering Types:[9]

- **Distribution based methods :** It is a clustering model in which we will fit the data on the probability that it may belong to the same distribution. The grouping done may be normal or gaussian .



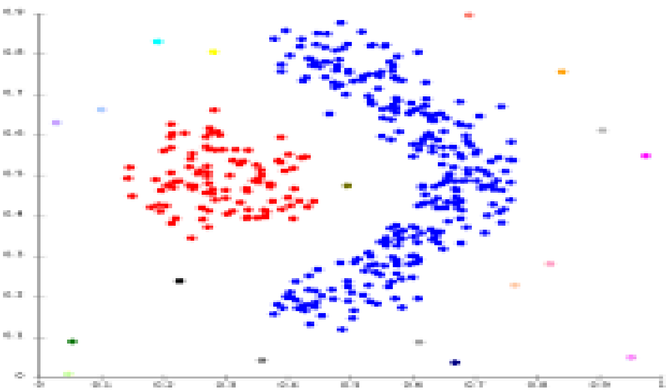
Figure(5): Distribution based methods Clustering

- **Centroid based methods :** This is basically one of iterative clustering algorithms in which the clusters are formed by the closeness of data points to the centroid of clusters. K – means algorithm is one of the popular examples of this algorithm.



Figure(6): Centroid based methods Clustering

- **Density Models :** In this clustering model there will be a search of data space for areas of varied density of data points in the data space . It isolates various density regions based on different densities present in the data space .



Figure(7): Density Models Clustering

- **Connectivity Based Clustering :** which is known as Hierarchy Clustering which constructs trees of clusters of objects, in which any two clusters are disjoint, or one includes the other. The cluster of all objects is the root of the tree.

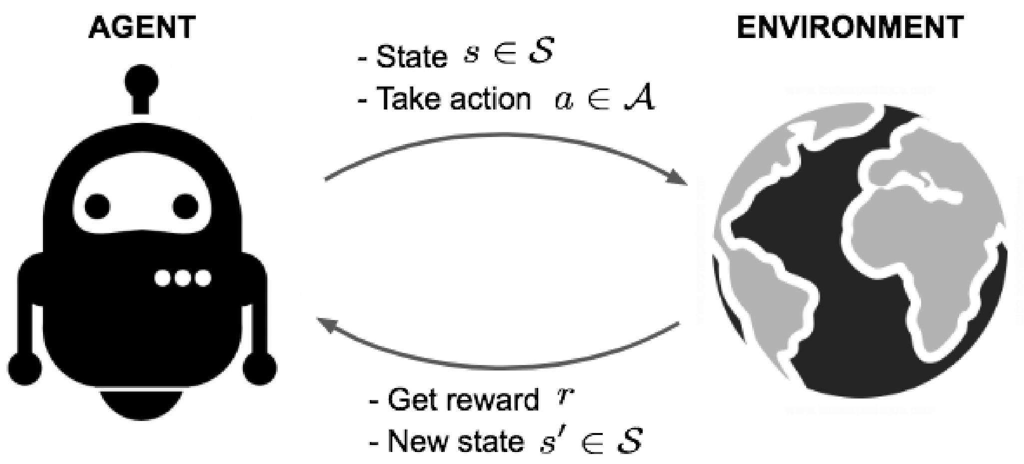
2) **Association:** An association rule is an unsupervised learning method which is used for finding the relationships between variables in the large database. It determines the set of items that occur together in the dataset.**[8]**

   Association rule learning can be divided into three types of algorithms:**[11]**

   - Apriori.
   - Eclat.
   - F-P Growth Algorithm.

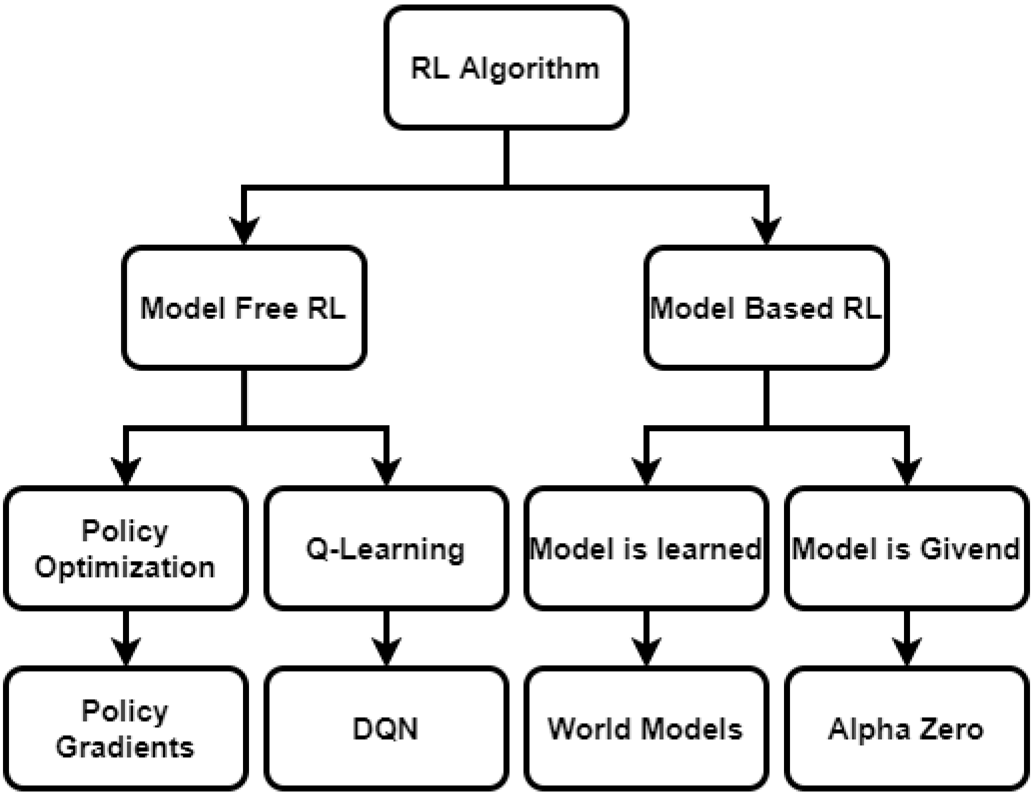
2.4.3 Reinforcement Learning (RL)

Is an area of machine learning concerned with how intelligent agents ought to take actions in an environment in order to maximize the notion of cumulative reward , The goal of reinforcement learning is to pick the best known action for any given state.**[10]**



Figure(8):RL scenario

- **Agent:** It is an assumed entity which performs actions in an environment to gain some reward.
- **Environment (e):** A scenario that an agent has to face.
- **Reward (R):** An immediate return given to an agent when he or she performs specific action or task.
- **State (s):** State refers to the current situation returned by the environment.
- **Policy (π):** It is a strategy which is applied by the agent to decide the next action based on the current state.



Figure(9):RL algorithms

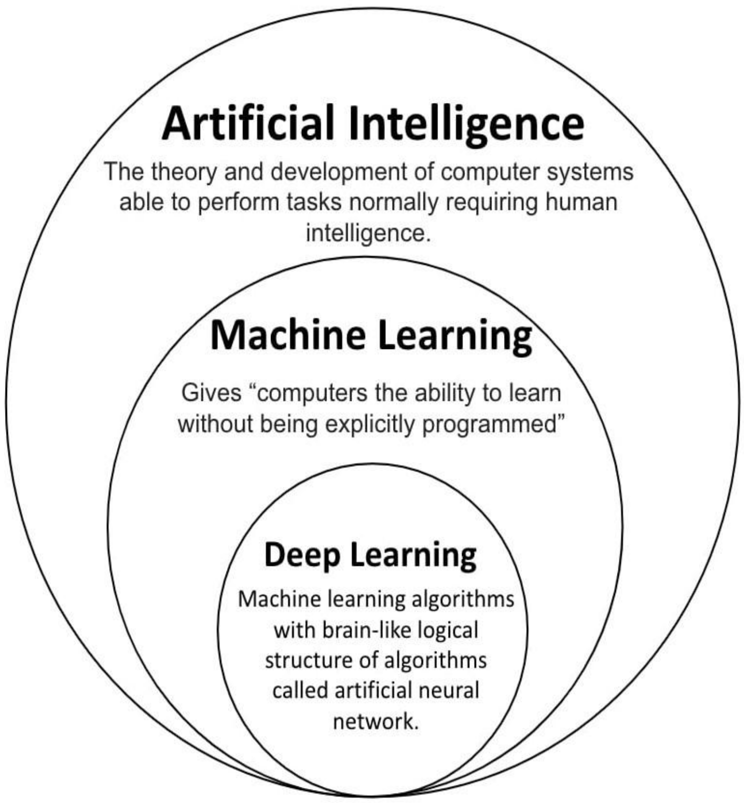
Chapter Three

3. Deep Learning

3.1 DL Definition

Deep learning is an artificial intelligence (AI) 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 in artificial intelligence that has networks capable of learning unsupervised from data that is unstructured or unlabeled. Also known as deep neural learning or deep neural network.**[12]**



Figure(10):Understanding Deep Learning.

3.2 ML vs DL

1) **Data dependencies:** The most important difference between deep learning and traditional machine learning is its performance as the scale of data increases. When the data is small, deep learning algorithms don't perform that well. This is because deep learning algorithms need a large amount of data to understand it perfectly. On the other hand, traditional machine learning algorithms with their handcrafted rules prevail in this scenario.**[13]**

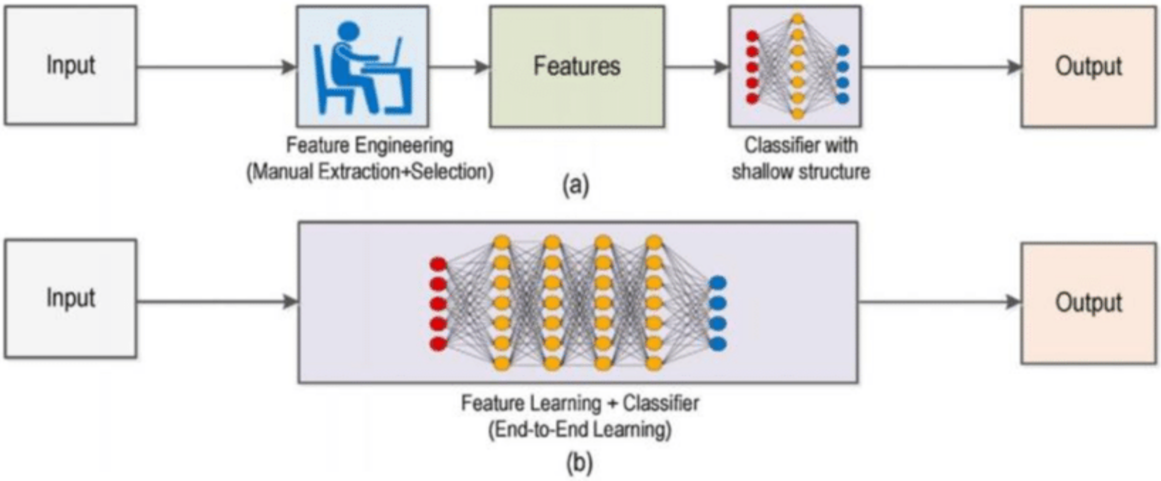
2) **Hardware dependencies:** Deep learning algorithms heavily depend on high-end machines, contrary to traditional machine learning algorithms, which can work on low-end machines. This is because the requirements of deep learning algorithms include GPUs which are an integral part of its working. Deep learning algorithms inherently do a large amount of matrix multiplication operations. These operations can be efficiently optimized using a GPU because GPU is built for this purpose.**[13]**

3) **Feature engineering:** Feature engineering is a process of putting domain knowledge into the creation of feature extractors to reduce the complexity of the data and make patterns more visible to learning algorithms to work. This process is difficult and expensive in terms of time and expertise.**[13]**

   In Machine learning, most of the applied features need to be identified by an expert and then hand-coded as per the domain and data type.

   For example, features can be pixel values, shape, textures, position and orientation. The performance of most of the Machine Learning algorithms depends on how accurately the features are identified and extracted.

Deep learning algorithms try to learn high-level features from data. This is a very distinctive part of Deep Learning and a major step ahead of traditional Machine Learning. Therefore, deep learning reduces the task of developing new feature extractor for every problem. Convolutional NN will try to learn low-level features such as edges and lines in early layers then parts of faces of people and then high-level representation of a face.



Figure(11):Feature Engineering.

4) **Execution time :** a deep learning algorithm takes a long time to train. This is because there are so many parameters in a deep learning algorithm that training them takes longer than usual whereas machine learning comparatively takes much less time to train, ranging from a few seconds to a few hours.

   This in turn is completely reversed on testing time. At test time, deep learning algorithms take much less time to run. Whereas, if you compare it with k-nearest neighbors (a type of machine learning algorithm), test time increases on increasing the size of data. Although this is not applicable on all machine learning algorithms, as some of them have small testing times too.**[14]**

3.3 Neural Network

The inspiration for deep learning is the way that the human brain filters the information. Its main motive is to simulate human-like decision making.

Neurons in the brain pass the signals to perform the actions.

artificial neurons connect in a neural network to perform tasks clustering, classification, or regression. **[14]**

3.3.1 Layers

Neurons are grouped into three different types of layers :

- **Input layer**

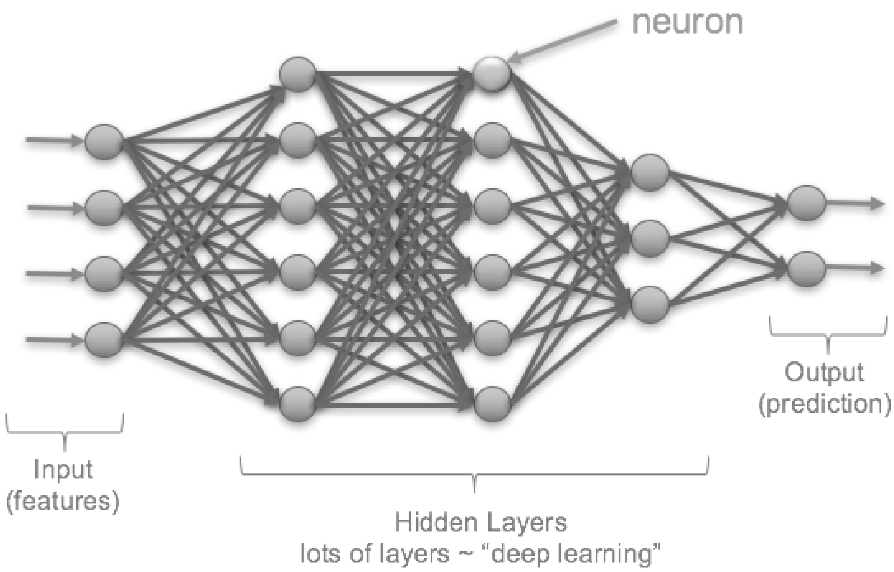
  It receives the input data from the observation.

- **Hidden layer**

  It performs mathematical computations on input data.

- **Output layer**

  It gives the desired result.



Figure(12):Types of Neurons Layers .

3.3.2 Weight

The connection between neurons is called weight, which is the numerical value . The weight between neurons determines the learning ability of the neural network.**[14]**

3.3.3 Activation function

It is used for standardizing the output from the neuron , Activation functions are the mathematical equations that calculate the output of the neural network. It also helps to normalize the output in a range between 0 to 1 or -1 to 1. **[14]**

- **Types of Activation Functions**
    - **Linear Function:[17]**

    Equation:

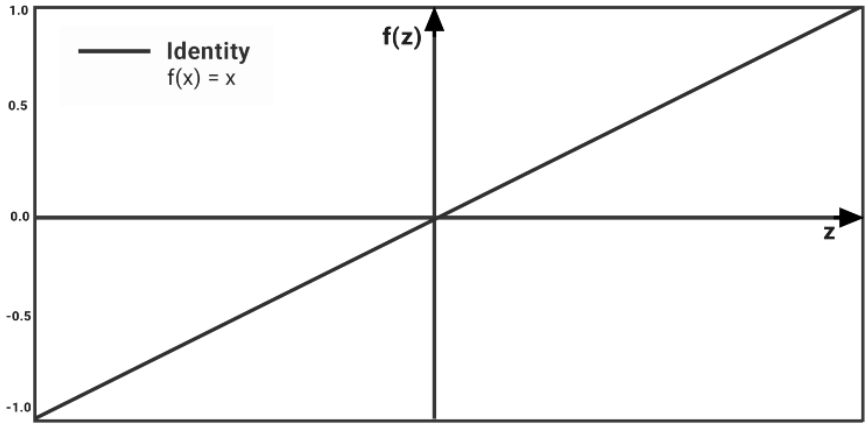
$$f(x) = ax + c$$

Explanation: The value of f(x) increases proportionally with the value of x .

Range: The input value is the weighted sum of the weighted and bias of the neurons in a layer .

Usage: solves the issue of a binary step function where it reports only a value of 0 and 1.

Issue: this function is not fit to handle complexes.



Figure(13):Linear Function .

- **Sigmoid Function[17]**

Equation:

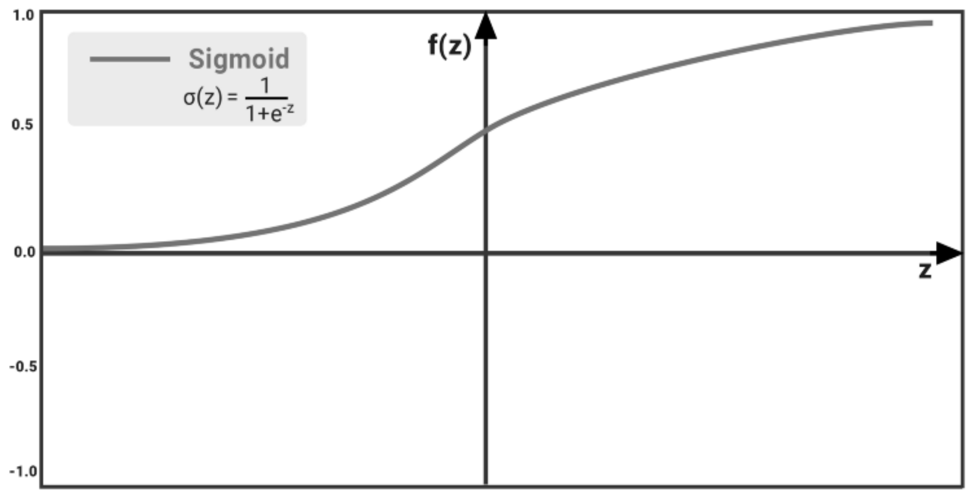
$$\frac{1}{1+e^{-x}}$$

Explanation: The Sigmoid function takes a value as input and outputs another value between 0 and 1 , It is non-linear and easy to work with when constructing a neural network model , The good part about this function is that continuously differentiable over different values of *z* and has a fixed output range.

Range : from 0 to 1 .

Usage : especially in binary classification models as part of the output layer to capture the probability ranging from 0 to 1.

Issue: function arises when we have multiple hidden layers in our neural network.



Figure(14):Sigmoid Function .

- **Tanh Function[17]**

  Equation:

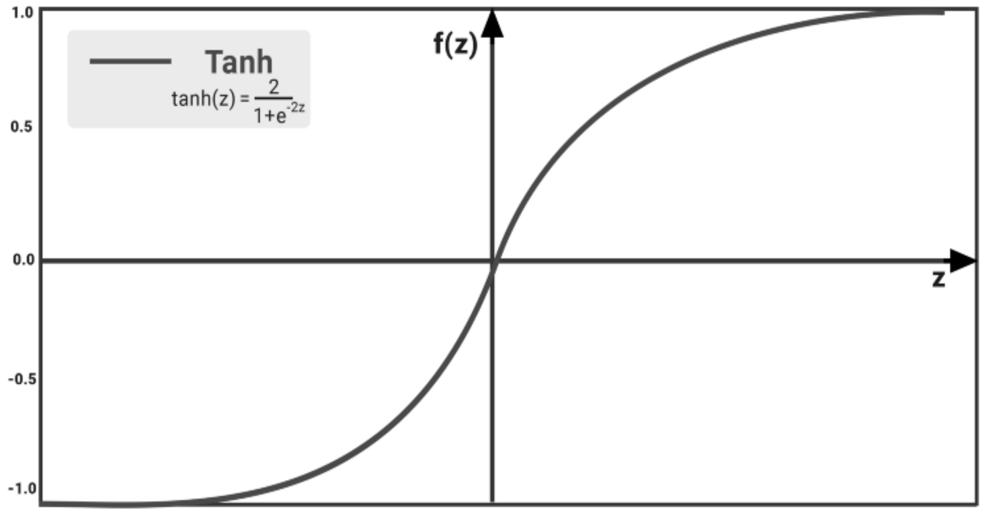
$$\frac{e^{2x} - 1}{e^{2x} + 1}$$

Explanation : The Tanh function is a modified or scaled up version of the sigmoid function.

Range : from -1 to 1

Usage : helps us in establishing which scores to consider in the next layer and which to ignore.

Issue: this function still has the vanishing gradient problem.



Figure(15):Tanh Function .

- **Rectified Linear Unit Function (ReLU)**

  Equation:

$$f(x) = max(x, 0)$$

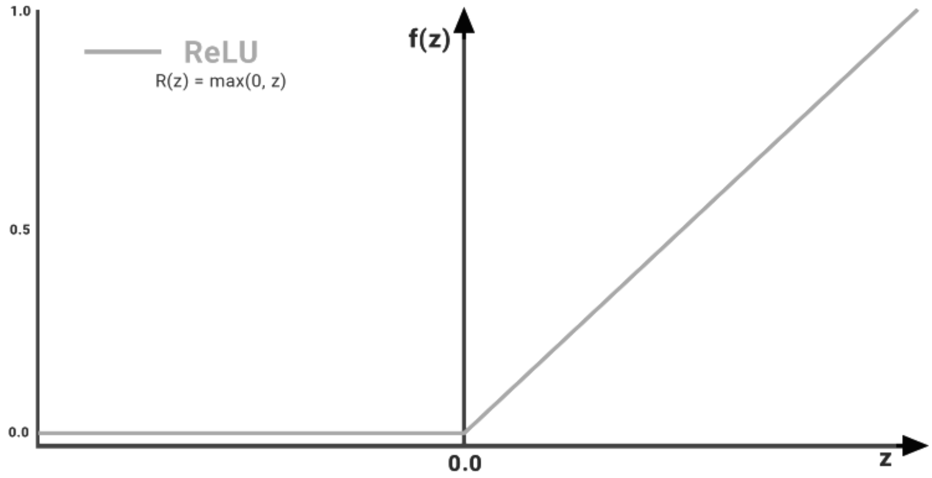
Explanation: The Rectified Linear Unit or ReLU for short would be considered the most commonly used activation function in deep learning models .

The function simply outputs the value of 0 if it receives any negative input, but for any positive value *z,* it returns that value back like a linear function.

Range : from 0 to ∞ .

Usage : backpropagate the errors and have multiple layers of neurons being activated by the ReLU function.

Issues : The function suffers from the dying ReLU problem For activations corresponding to values of $z< 0$, the gradient will be 0 because of which the weights will not get adjusted during the gradient descent in backpropagation.



Figure(16):ReLU Function .

- **Softmax Activation Function**

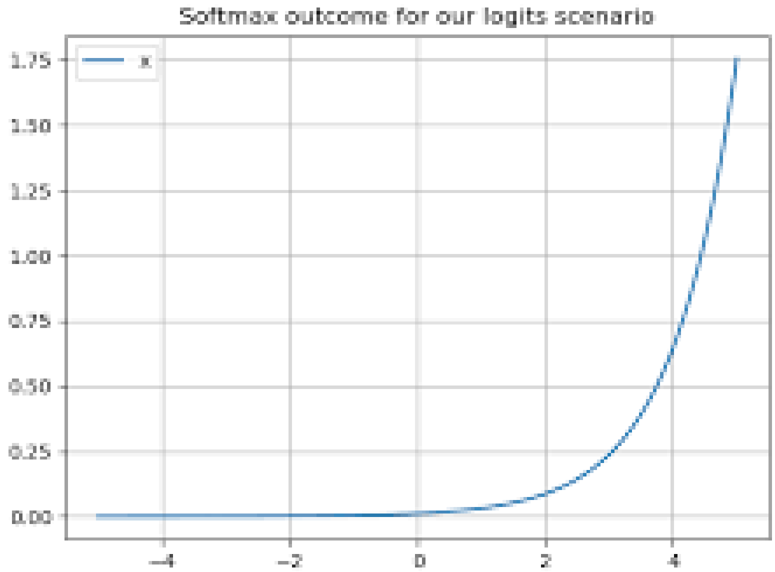
Equation:

$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}$$

Explanation : softmax function and the sigmoid function are similar. The softmax operates on a vector while the sigmoid takes a scalar.

Range : it used in classification problem which output either 0 or 1.

Usage :it converts a vector of numbers into a vector of probabilities,where the probabilities of each value are proportional to the relative scale of each value in the vector softmax function is used to normalize the outputs.



Figure(17):Softmax Function .

In an ANN, the sigmoid function is a non-linear AF used primarily in feedforward neural networks . It is a differentiable real function, defined for real input values, and containing positive derivatives everywhere with a specific degree of smoothness. The sigmoid function appears in the output layer of the deep learning models and is used for predicting probability-based outputs.

3.4 DL Algorithms Types

3.4.1 Artificial Neural Network

An artificial neural network (ANN) is the piece of a computing system designed to simulate the way the human brain analyzes and processes information.**[15]**

It is the foundation of artificial intelligence (AI) and solves problems that would prove impossible or difficult by human or statistical standards. ANNs have self-learning capabilities that enable them to produce better results as more data becomes available.

An ANN has hundreds or thousands of artificial neurons called processing units, which are interconnected by nodes. These processing units are made up of input and output units. The input units receive various forms and structures of information based on an internal weighting system, and the neural network attempts to learn about the information presented to produce one output report.**[15]**

In ANN , Data is processed in two different propagations which are forward propagation and backpropagation.

- **Forward Propagation**

  The input X provides the initial information that then propagates to the hidden units at each layer and finally produces the output $\widehat{Y}$. The architecture of the network entails determining its depth, width, and activation functions used on each layer. **Depth** is the number of hidden layers. **Width** is the number of units (nodes) on each hidden layer since we don’t control neither input layer nor output layer dimensions. There are quite a few sets of activation functions such as Rectified Linear Unit, Sigmoid, Hyperbolic tangent, etc.

Forward propagation (or forward pass) refers to the calculation and storage of intermediate variables (including outputs) for a neural network in order from the input layer to the output layer.**[16]**

For the sake of simplicity, let us assume that the input example is $X \in R^{d}$ and that our hidden layer does not include a bias term. Here the intermediate variable is:**[16]**

$$\mathbf{z} = \mathbf{W}^{(1)}\mathbf{x},$$

where $W^{(1)} \in R^{h*d}$ is the weight parameter of the hidden layer. After running the intermediate variable $Z \in R^{h}$ through the activation function
ϕ we obtain our hidden activation vector of length h,**[16]**

$$\mathbf{h} = \phi(\mathbf{z}).$$

The hidden variable h is also an intermediate variable. Assuming that the parameters of the output layer only possess a weight of $W^{(2)} \in R^{q*h}$,we can obtain an output layer variable with a vector of length O.**[17]**

$$\mathbf{o} = \mathbf{W}^{(2)}\mathbf{h}.$$

Assuming that the loss function is l and the example label is y, we can then calculate the loss term for a single data example,**[16]**

$$L = l(\mathbf{o}, y).$$

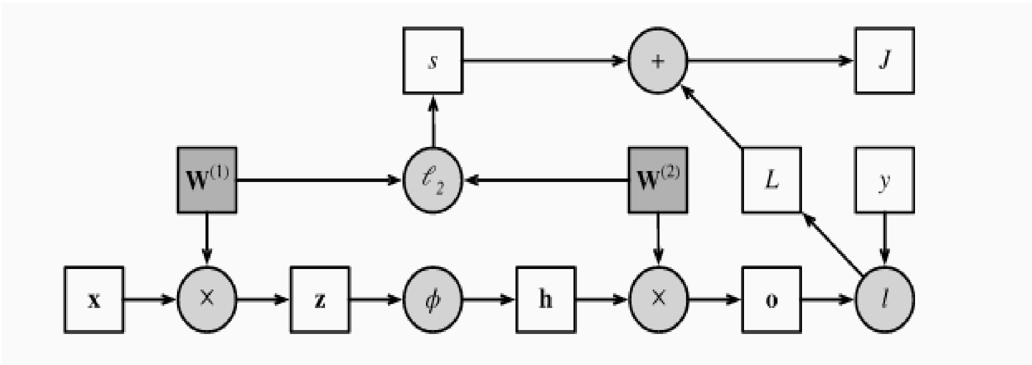
According to the definition of L2 regularization, given the hyperparameter λ, the regularization term is

$$s = \frac{\lambda}{2}\left(\|\mathbf{W}^{(1)}\|_F^2 + \|\mathbf{W}^{(2)}\|_F^2\right),$$

where the Frobenius norm of the matrix is simply the L2 norm applied after flattening the matrix into a vector. Finally, the model's regularized loss on a given data example is:**[16]**

$$J = L + s.$$

We refer to J as the objective function.



Figure(18):Computational graph of forward propagation .

- **Backpropagation:**

Backpropagation refers to the method of calculating the gradient of neural network parameters. In short, the method traverses the network in reverse order, from the output to the input layer, according to the chain rule from calculus. The algorithm stores any intermediate variables (partial derivatives) required while calculating the gradient with respect to some parameters. Assume that we have functions Y=f(X) and Z=g(Y), in which the input and the output X,Y,Z are tensors of arbitrary shapes. By using the chain rule, we can compute the derivative of Z with respect to X via

$$\frac{\partial \mathsf{Z}}{\partial \mathsf{X}} = \text{prod}\left(\frac{\partial \mathsf{Z}}{\partial \mathsf{Y}}, \frac{\partial \mathsf{Y}}{\partial \mathsf{X}}\right).$$

Here we use the prod operator to multiply its arguments after the necessary operations, such as transposition and swapping input positions, have been carried out. For vectors, this is straightforward: it is simply matrix-matrix multiplication. For higher dimensional tensors, we use the appropriate counterpart. The operator prod hides all the notation overhead.

Recall that the parameters of the simple network with one hidden layer, whose computational graph is in **Fig.18**, are$W^{(1)}$ and $W^{(2)}$.The objective of backpropagation is to calculate the gradients $\frac{\partial J}{\partial W^{(1)}}$and $\frac{\partial J}{\partial W^{(2)}}$. To accomplish this, we apply the chain rule and calculate, in turn, the gradient of each intermediate variable and parameter. The order of calculations are reversed relative to those performed in forward propagation, since we need to start with the outcome of the computational graph and work our way towards the parameters. The first step is to calculate the gradients of the objective function J=L+s with respect to the loss term L and the regularization term s.**[16]**

$$\frac{\partial J}{\partial L} = 1 \text{ and } \frac{\partial J}{\partial s} = 1.$$

Next, we compute the gradient of the objective function with respect to variable of the output layer o according to the chain rule:**[17]**

$$\frac{\partial J}{\partial \mathbf{o}} = \text{prod}\left(\frac{\partial J}{\partial L}, \frac{\partial L}{\partial \mathbf{o}}\right) = \frac{\partial L}{\partial \mathbf{o}} \in \mathbb{R}^q.$$

Next, we calculate the gradients of the regularization term with respect to both parameters:**[16]**

$$\frac{\partial s}{\partial \mathbf{W}^{(1)}} = \lambda \mathbf{W}^{(1)} \text{ and } \frac{\partial s}{\partial \mathbf{W}^{(2)}} = \lambda \mathbf{W}^{(2)}.$$

Now we are able to calculate the gradient $\frac{\partial J}{\partial W^{(2)}} \in R^{q*h}$ of the model parameters closest to the output layer. Using the chain rule yields:**[16]**

$$\frac{\partial J}{\partial \mathbf{W}^{(1)}} = \text{prod}\left(\frac{\partial J}{\partial \mathbf{z}}, \frac{\partial \mathbf{z}}{\partial \mathbf{W}^{(1)}}\right) + \text{prod}\left(\frac{\partial J}{\partial s}, \frac{\partial s}{\partial \mathbf{W}^{(1)}}\right) = \frac{\partial J}{\partial \mathbf{z}} \mathbf{x}^\top + \lambda \mathbf{W}^{(1)}.$$

To obtain the gradient with respect to $W^{(1)}$ we need to continue back propagation along the output layer to the hidden layer. The gradient with respect to the hidden layer's outputs $\frac{\partial J}{\partial h} \in R^{h}$ is given by

$$\frac{\partial J}{\partial \mathbf{h}} = \text{prod}\left(\frac{\partial J}{\partial \mathbf{o}}, \frac{\partial \mathbf{o}}{\partial \mathbf{h}}\right) = {\mathbf{W}^{(2)}}^\top \frac{\partial J}{\partial \mathbf{o}}.$$

Since the activation function ϕ applies elementwise, calculating the gradient $\frac{\partial J}{\partial Z} \in R^{h}$ of the intermediate variable z requires that we use the element wise multiplication operator, which we denote by ⊙:

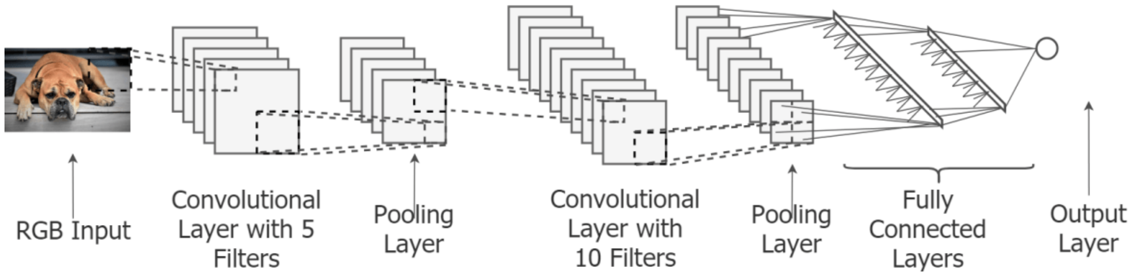
$$\frac{\partial J}{\partial \mathbf{z}} = \text{prod}\left(\frac{\partial J}{\partial \mathbf{h}}, \frac{\partial \mathbf{h}}{\partial \mathbf{z}}\right) = \frac{\partial J}{\partial \mathbf{h}} \odot \phi'\left(\mathbf{z}\right).$$

Finally, we can obtain the gradient $\frac{\partial J}{\partial W^{(1)}} \in R^{h*d}$ of the model parameters closest to the input layer. According to the chain rule, we get

$$\frac{\partial J}{\partial \mathbf{W}^{(1)}} = \text{prod}\left(\frac{\partial J}{\partial \mathbf{z}}, \frac{\partial \mathbf{z}}{\partial \mathbf{W}^{(1)}}\right) + \text{prod}\left(\frac{\partial J}{\partial s}, \frac{\partial s}{\partial \mathbf{W}^{(1)}}\right) = \frac{\partial J}{\partial \mathbf{z}} \mathbf{x}^\top + \lambda \mathbf{W}^{(1)}.$$

3.4.2 Convolutional Neural Network

A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other.**[18]**



Figure(19):CNN Architecture

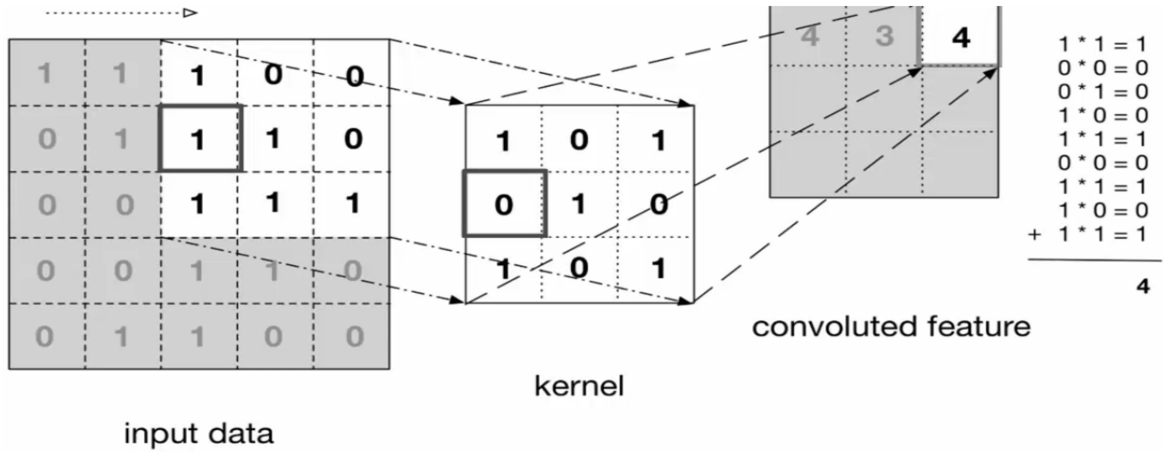
Convolutional neural network (CNN), a class of artificial neural networks that has become dominant in various computer vision tasks, is attracting interest across a variety of domains, including radiology.

CNN is designed to automatically and adaptively learn spatial hierarchies of features through backpropagation by using multiple building blocks, such as convolution layers, pooling layers, and fully connected layers.

CNN Layers:

1) **Convolution Layer :** Convolution is the first layer to extract features from an input image. Convolution preserves the relationship between pixels by learning image features using small squares of input data. It is a mathematical operation that takes two inputs such as image matrix and a filter or kernel.

   Convolution is a mathematical operation on two objects to produce an outcome that expresses how the shape of one is modified by the other.

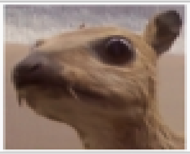
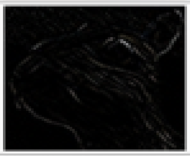
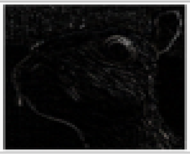
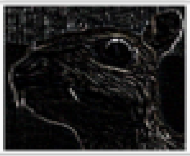
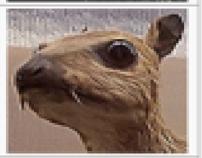
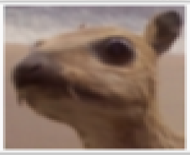
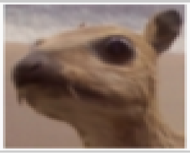


Figure(20): Convolution Layer

CNN uses filters (Kernels) to detect specific features in an image in order to get useful features .

Filter (Kernel) is a set of weights in a matrix applied on an image or a matrix to obtain the required features.

We can obtain different output using different filters and here are some types of filters

Operation	Filter	Convolved Image
Identity	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	
Edge detection	$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix}$	
	$\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$	
	$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$	
Sharpen	$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix}$	
Box blur (normalized)	$\frac{1}{9}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$	
Gaussian blur (approximation)	$\frac{1}{16}\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$	

Figure(21): Types of filters

Strides:

Stride is the number of pixels shifted over the input matrix. When the stride is 1 then we move the filters to 1 pixel at a time. When the stride is 2 then we move the filters to 2 pixels at a time and so on.**[19]**

Padding:

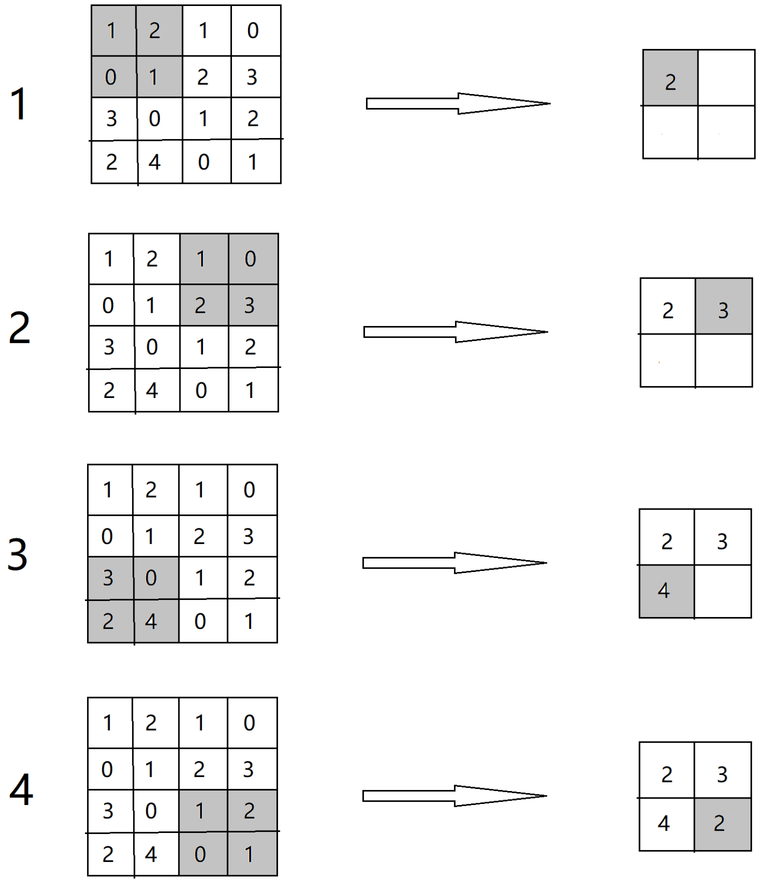
Sometimes the filter does not fit the input image. We have two options:

- Pad the picture with zeros (zero-padding) so that it fits
- Drop the part of the image where the filter did not fit. This is called valid padding which keeps only valid part of the image.**[19]**

2) **Pooling Layer:** Pooling layers section would reduce the number of parameters when the images are too large. Spatial pooling is also called subsampling or downsampling which reduces the dimensionality of each map but retains important information. Spatial pooling can be of different types:

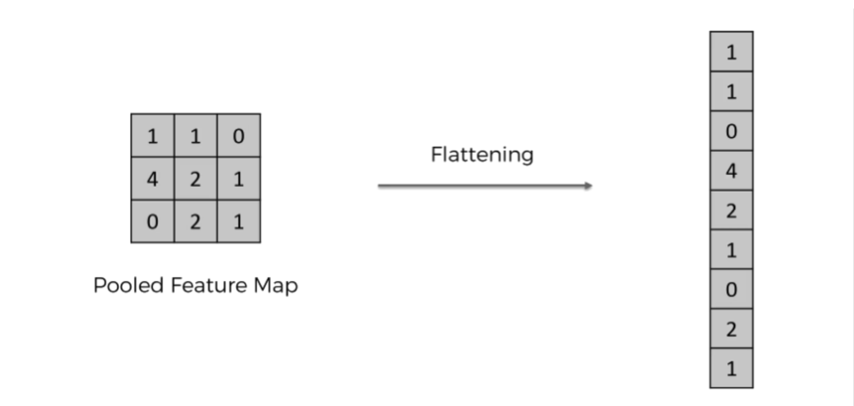
    - Max Pooling
    - Average Pooling
    - Sum Pooling

    Max pooling takes the largest element from the rectified feature map. Taking the largest element could also take the average pooling. Sum of all elements in the feature map call as sum pooling.**[19]**



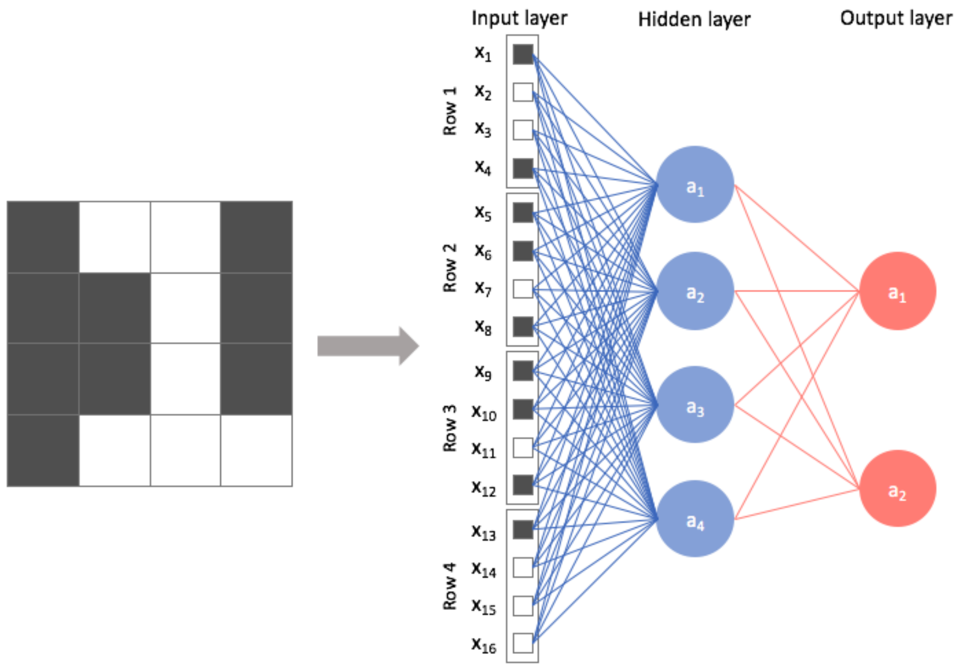
Figure(22): Example Of Pooling (Max pooling)

3) **Flatten layer:** Flattening is converting the data into a 1-dimensional array for inputting it to the next layer. We flatten the output of the convolutional layers to create a single long feature vector.**[20]**



Figure(23): Flatten layer

4) **Fully Connected Layer:** which is feedforward neural networks , neurons in a fully connected layer have full connections to all activations in the previous layer, as it is in regular ANN, and their activation can be computed via a matrix multiplication followed by a bias offset.



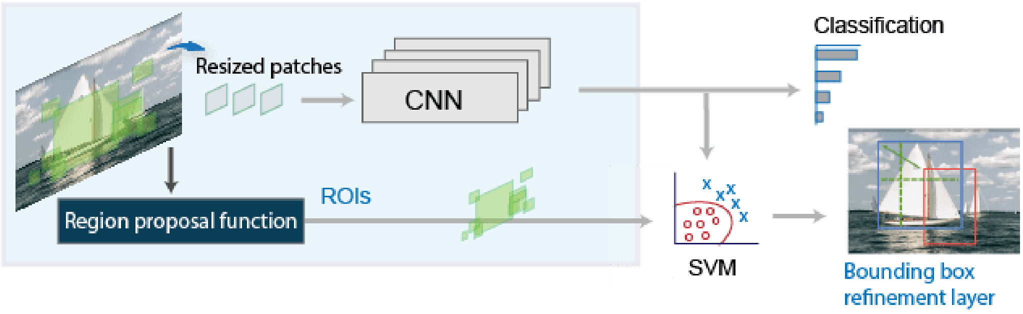
Figure(24):Fully connected layer

The problem with CNN is that the objects of interest might have different spatial locations within the image and different aspect ratios. And this leads us to get a huge number of regions and this could computationally blow up. Therefore, algorithms like R-CNN have been developed to find these occurrences and find them fast.

3.4.3 Region-Based Convolutional Neural Network :

R-CNN is short for “Region-based Convolutional Neural Networks”. The main idea is composed of two steps. First, using selective search, it identifies a manageable number of bounding-box object region candidates (“region of interest” or “RoI”). And then it extracts CNN features from each region independently for classification.

The CNN acts as a feature extractor and the output dense layer consists of the features extracted from the image and the extracted features are fed into an SVM to classify the presence of the object within that candidate region proposal.



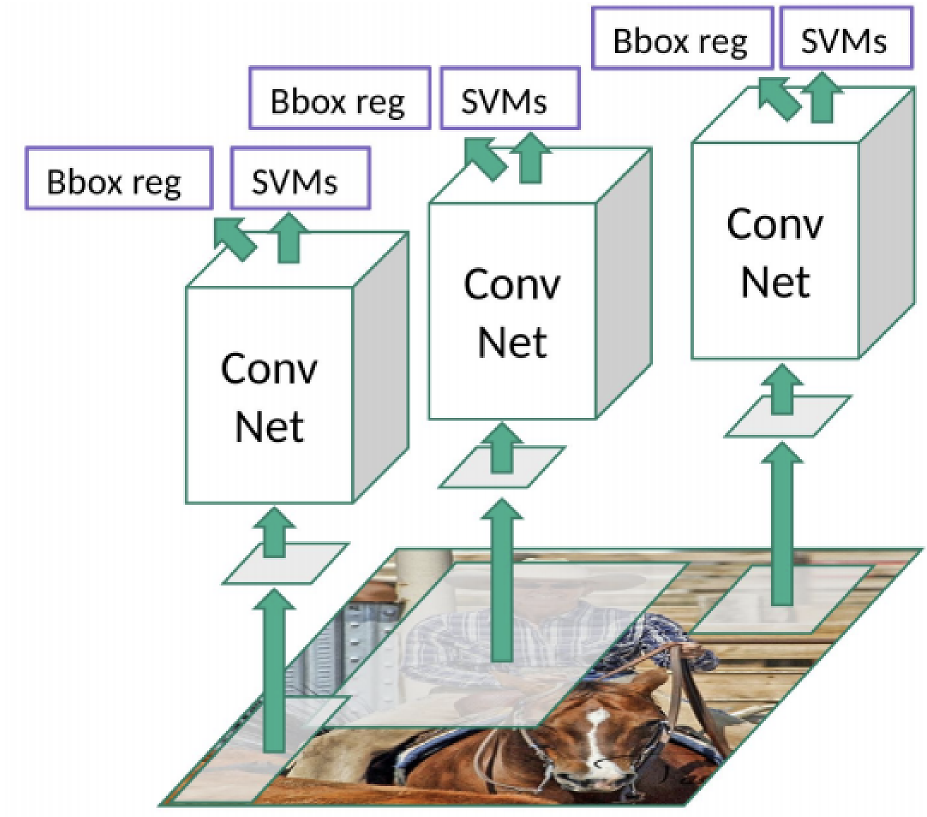
Figure(25): R-CNN Architecture [23]

Model Workflow[21]:

- Pre-train a CNN network on image classification tasks.
- Propose category-independent regions of interest by selective search . Those regions may contain target objects and they are of different sizes.
    - Selective search is a common algorithm to provide region proposals that potentially contain objects. It is built on top of the image segmentation output and uses region-based characteristics.
- Region candidates are warped to have a fixed size as required by CNN.
- Continue fine-tuning the CNN on warped proposal regions for K + 1 classes.
- Given every image region, one forward propagation through the CNN generates a feature vector. This feature vector is then consumed by a binary SVM trained for each class independently.

The positive samples are proposed regions with IoU (intersection over union) overlap threshold >= 0.3, and negative samples are irrelevant.

- To reduce the localization errors, a regression model is trained to correct the predicted detection window on the bounding box correction offset using CNN features.

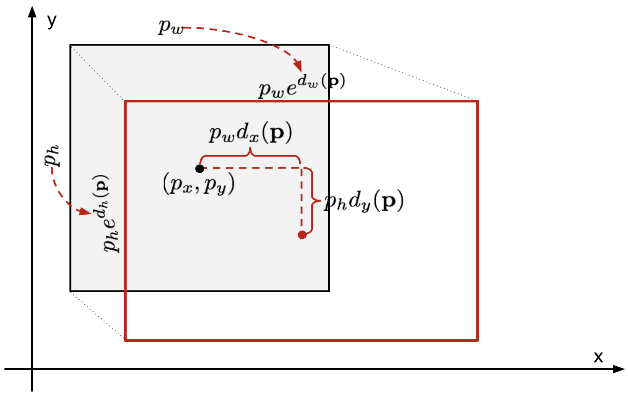


Figure(26): R-CNN Model [24]

Bounding Box Regression

Given a predicted bounding box coordinate p=(px,py,pw,ph) (center coordinate, width, height) and its corresponding ground truth box coordinates g=(gx,gy,gw,gh) the regressor is configured to learn scale-invariant transformation between two centers and log-scale transformation between widths and heights. All the transformation functions take p as input.

$$\hat{g}_x = p_w d_x(\mathbf{p}) + p_x$$
$$\hat{g}_y = p_h d_y(\mathbf{p}) + p_y$$
$$\hat{g}_w = p_w \exp(d_w(\mathbf{p}))$$
$$\hat{g}_h = p_h \exp(d_h(\mathbf{p}))$$



Figure(27): Bounding Box[22]

An obvious benefit of applying such transformation is that all the bounding box correction functions di (p) where i∈{x,y,w,h} can take any value between [-∞, +∞] The targets for them to learn are:

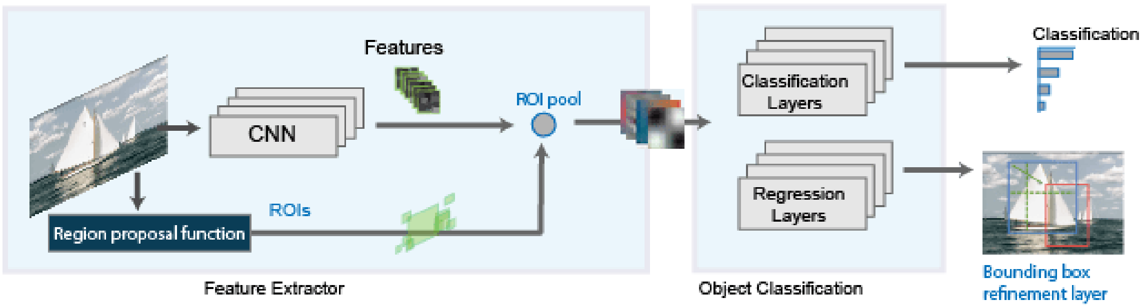
$$t_x = (g_x - p_x)/p_w$$
$$t_y = (g_y - p_y)/p_h$$
$$t_w = \log(g_w/p_w)$$
$$t_h = \log(g_h/p_h)$$

Disadvantage

- One of the most common defects of R-CNN is speed bottleneck as the model is expensive and slow.
- The main performance bottleneck of an R-CNN lies in the independent CNN forward propagation for each region proposal without sharing computation.

3.4.4 Fast R-CNN

To make R-CNN faster, Girshick improved procedure by unifying three independent models into one jointly trained framework and increasing shared computation results , One of the major improvements of the fast *R-CNN* from the R-CNN is that the CNN forward propagation is only performed on the entire image and the region proposals share this feature matrix, Then the same feature matrix is branched out to be used for learning the object classifier and the bounding-box regressor.

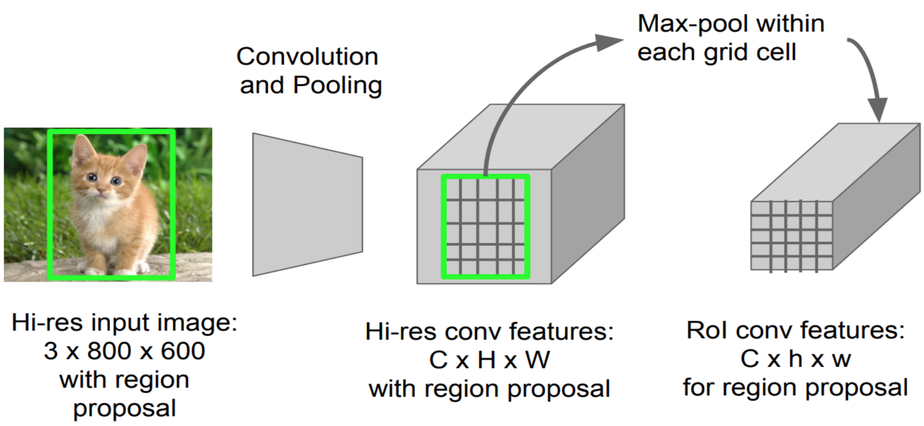


Figure(28): Fast R-CNN Architecture [23]

RoI Pooling

It is a type of max pooling to convert features in the projected region of the image of any size, h x w, into a small fixed window, H x W.

The input region is divided into H x W grids, approximately every subwindow of size h/H x w/W. Then apply max-pooling in each grid.



Figure(29): ROI Pooling[22]

Model Workflow[22]

- First, pre-train a convolutional neural network on image classification tasks.
- Propose regions by selective search (~2k candidates per image).

  Alter the pre-trained CNN:

  - Replace the last max pooling layer of the pre-trained CNN with a ROI Pooling layer. The RoI pooling layer outputs fixed-length feature vectors of region proposals. Sharing the CNN computation makes a lot of sense, as many region proposals of the same images are highly overlapped.
  - Replace the last fully connected layer and the last softmax layer (K classes) with a fully connected layer and softmax over K + 1 classes.

  - Finally the model branches into two output layers:
    - A softmax estimator of K + 1 classes (same as in R-CNN, +1 is the “background” class), outputting a discrete probability distribution per RoI.
    - A bounding-box regression model which predicts offsets relative to the original RoI for each of K classes.

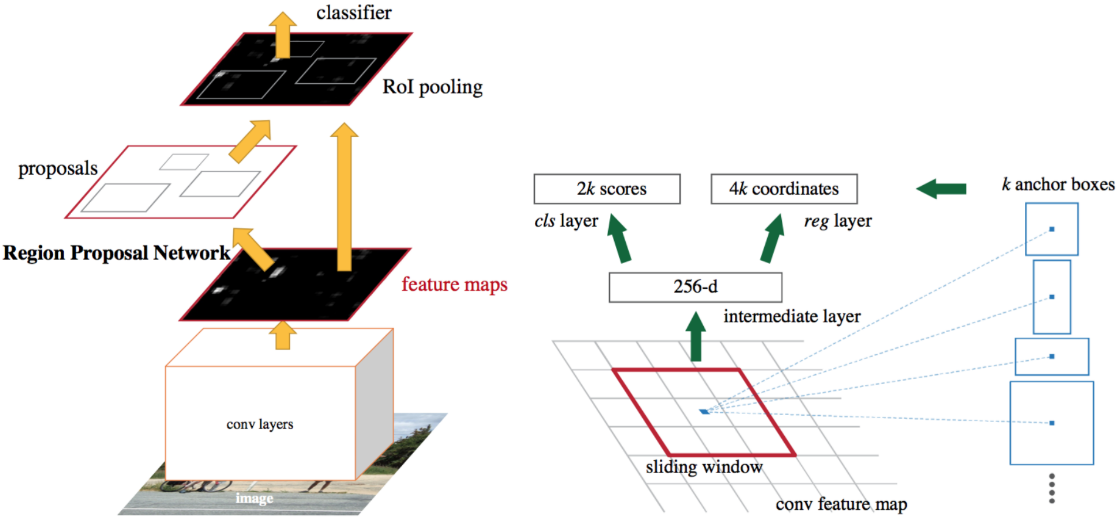
  Fast R-CNN is much faster in both training and testing time.

Disadvantage

- The improvement is not dramatic because the regional proposals are generated separately by another model and that is very expensive.

3.4.5 Faster R-CNN

An intuitive speedup solution is to integrate the region proposal algorithm into the CNN model. Faster R-CNN Ren is doing exactly this: construct a single, unified model composed of RPN (region proposal network) and fast R-CNN with shared convolutional feature layers.



Figure(30): Faster R-CNN architecture .[23]

Model Workflow [24]

- Pre-train a CNN network on image classification tasks.
- Fine-tune the RPN (region proposal network) end-to-end for the region proposal task, which is initialized by the pre-train image classifier. Positive samples have IoU (intersection-over-union) > 0.7, while negative samples have IoU < 0.3.
    - Slide a small n x n spatial window over the conv feature map of the entire image.
    - At the center of each sliding window, we predict multiple regions of various scales and ratios simultaneously. An anchor is a combination of (sliding window center, scale,

ratio). For example, 3 scales + 3 ratios => k=9 anchors at each sliding position.

- Train a Fast R-CNN object detection model using the proposals generated by the current RPN
- Then use the Fast R-CNN network to initialize RPN training. While keeping the shared convolutional layers, only fine-tune the RPN-specific layers.
- Finally fine-tune the unique layers of Fast R-CNN.

Finally , in this chapter we took a background about machine learning and deep learning , so in chapter 9 we will talk about the solutions we used in BI-Transaltor System .

Chapter Four

4. Animation

Animation comes from the Latin words “anima,” which means “life,” and “animare” which means “to breathe life into.”Animation consists of still images (called “frames”) with slight differences between them.

When viewed together in a sequence, they give the illusion of motion **[25]**

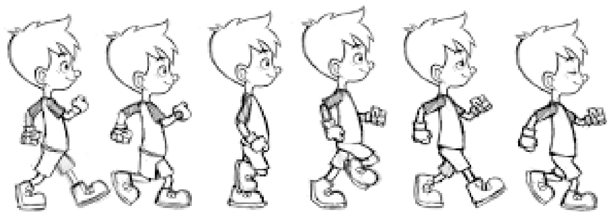
4.1 Animation Definition

Animation is a method of photographing successive drawings and models to create an illusion of movement in a sequence. Because our eyes can only retain an image for approx. 1/10 of a second, when multiple images appear in fast succession, the brain blends them into a single moving image.**[26]**

Different Types of Animation:

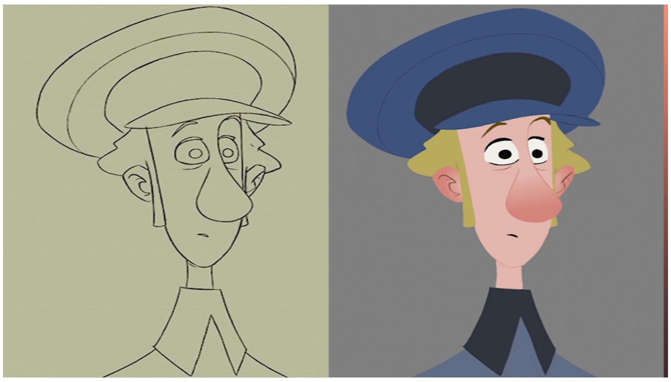
- **Traditional Animation:**

  usually refers to animation hand-drawn on paper. It was the process used for most of the productions throughout the 20th century.



Figure(31): Traditional Animation

- **Vector-based Animation:** it can be 2D without being traditional.Vector animation refers to animation where the art or motion is controlled by vectors rather than pixels. It often allows cleaner, smoother animation because images are displayed and resized using mathematical values instead of stored pixel values.**[27]**



Figure(32): 2D Vector Based Animation

- **3D Animation:** 3D animation is when computer generated objects appear to move through three dimensional space. In 3D animation, objects can be moved and rotated following the same principles as in real life.**[28, 29]**

Figure(33): 3D Animation

- **Motion Graphics:** Motion graphics is animation, but with text as a major component. Essentially, it's animated graphic design. Which motion graphics, where the marriage of sound, motion, and graphic design come together exceptionally well.

- **Stop Motion:** Stop motion animation (also called stop frame animation) is animation that is captured one frame at time, with physical objects that are moved between frames. When you play back the sequence of images rapidly, it creates the illusion of movement. If you understand how 2D drawn animation works, stop motion is similar, except using physical objects instead of drawings.

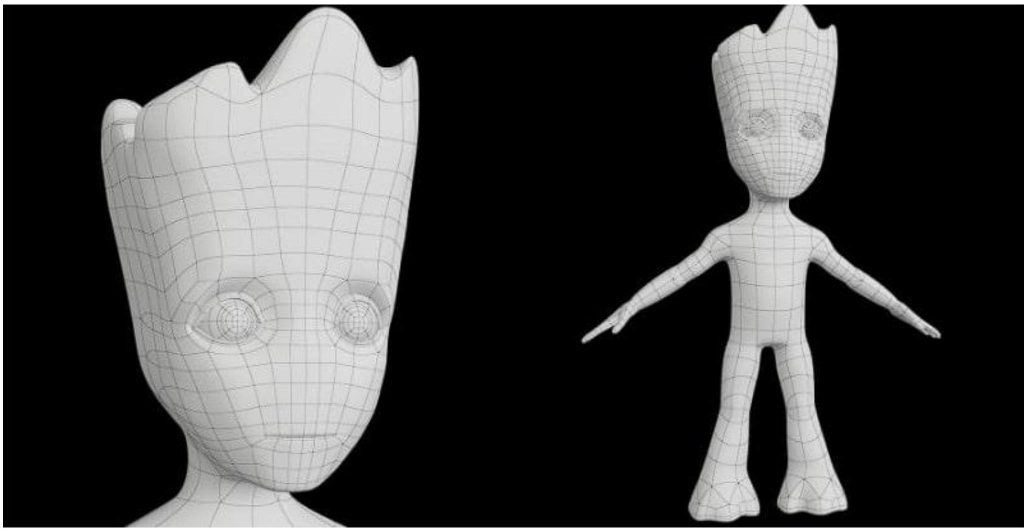
4.2 3D Animation

4.2.1 How does 3D animation work?

3D Animation can be done by using some processes which are:**[30]**

- **Modeling**

  Before we can have a functioning scene with character interactions, we first need to build our 3D characters. Characters are based on 3D computer models, and their creation is called 3D modeling. A basic 3D model is a mesh of points, lines, and curves arranged in a way to map out an object. A computer sees models as pure geometric shapes. It's not until colors and textures are added that the map starts to resemble a real object.

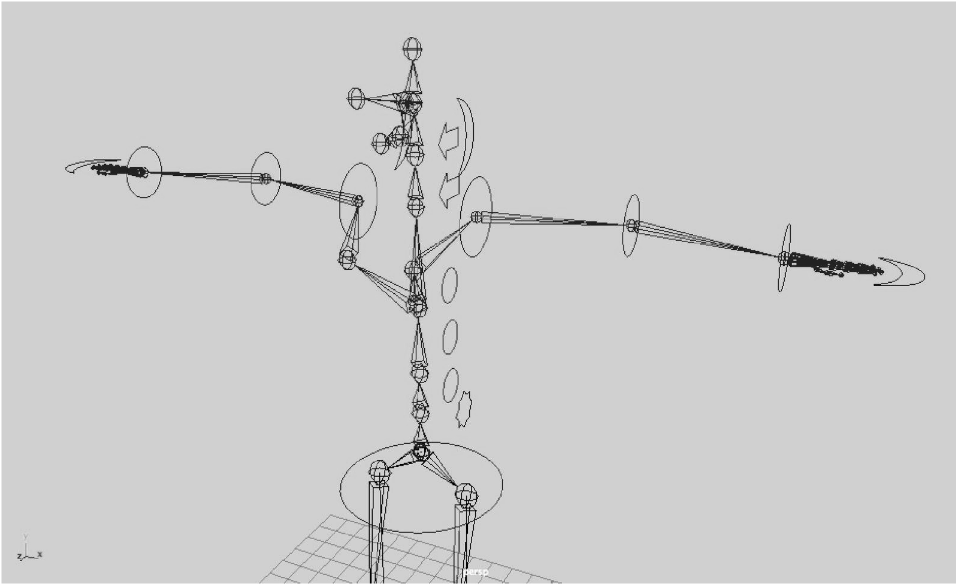


Figure(34): modelling Process

- **Rigging and skinning:**

  Rigging is the process of producing that skeleton. Once the skeleton is formed, the 3D model (also called skin) is attached to the rig to get the character ready to be moved around.

Rigging is the process to set up a controllable skeleton for the character that is intended for animation. Depending on the subject matter, every rig is unique and so is the corresponding set of controls.



Figure(35): Rigging Process

Skinning is the process of attaching the 3D model (skin) to the rigged skeleton so that the 3D model can be manipulated by the controls of the rig.**[32]**

- **Texturing**

The art of giving clothes to the 3D models.
When a 3D model is created, 2D images can be overlaid on it to add colours, designs, and textures. This is called mapping, and often the entirety of a model's colour comes from this.**[32]**

- **Animation**

putting characters in their respective scenes and animating their movements with 3D animation software.
Animation is the process of taking a 3D object and getting it to move. Animation comes in a few different flavors. There's keyframe animation, where the animator manipulates the objects on a frame-by-frame basis, similar to old hand-drawn cartoons. Other methods of animation include placing objects

on splines and setting them to follow the path of the curve, or importing motion capture data and applying it to a character rig. **[32]**

- **Rendering**
  The last part of the animation production process is rendering. This is when the animation is finalized and exported. This step requires great attention to detail to ensure the final render is perfectly polished.

4.2.2 Uses of 3D Animation:

- **Character animation** for film, TV and gaming.**[31]**
- **Education:**This is because 3D animation is fun and memorable for students learning a new process or idea. In a way, young people are more familiar with 3D animation because they see it so much in entertainment.**[31]**
- **Medicine:** When 3D animation is used in medical videos, the need for as much hands-on experience decreases.3D animation uses in medicine are also a step-up from the stationary, physical 3D models used in the past. Now, medical students can see all the movements a real human body makes.**[31]**
- **Business:**In business, the uses of 3D animation usually include
  - Presentations.
  - Marketing and advertising.
  - Education and training.**[31]**
- **The military and police:** 3D animation is used for the military and police training to avoid the risk of injury. In addition, the military and police might also use 3D animation to design their machining and engineering.**[31]**

- **Engineering and technology prototyping:** Speaking of machinery, engineering and technology are industries where 3D animation is massive. This is because it saves so much time. Engineers can view their design on a screen before they waste materials. They are also able to customise and modify their design as they see fit without any consequences.**[31]**
- **Architecture and design:**With 3D animations, architects can now experience their buildings in Virtual Reality (VR) or Augmented Reality (AR).**[31]**

4.2.3 why to use 3D animation

- 3D animation is eye-catching.
- it allows businesses to communicate in a memorable and immersive way.
- It’s applicable to anything.

Chapter Five

5. Augmented Reality

Figure(36): AR

5.1 What is AR ?

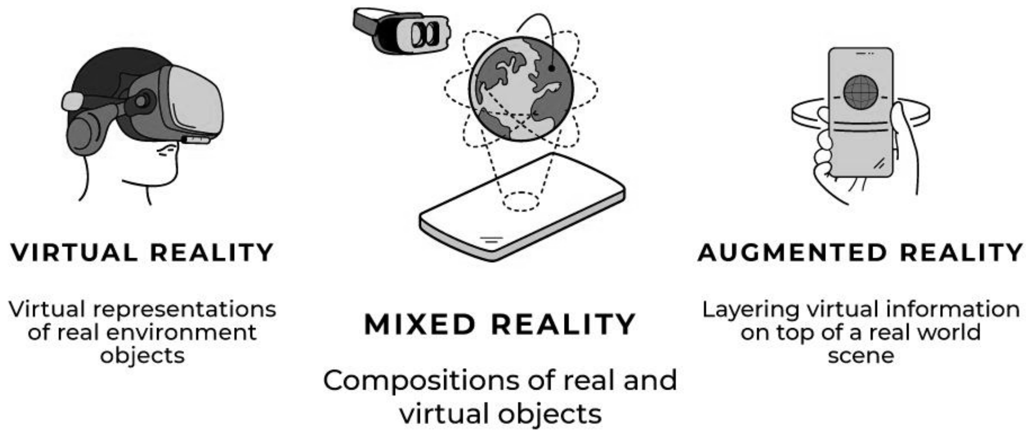
Augmented reality (AR) is an enhanced version of the real physical world that is achieved through the use of digital visual elements, sound, or other sensory stimuli delivered via technology.it is an interactive 3D experience that combines a view of the real world with computer-generated elements in real time and lets people superimpose digital content (images, sounds, text) over a real-world environment.**[32,33]**

5.2 VR & MR

- **Virtual Reality :** it refers to a computer-generated simulation in which a person can interact within an artificial three-dimensional environment using electronic devices, such as special goggles with a screen or gloves fitted with sensors. In this simulated artificial environment, the user is able to have a realistic-feeling experience.**[34]**

- **Mixed Reality:** it is a user environment in which physical reality and digital content are combined in a way that enables interaction with and among real-world and virtual objects.**[35]**

5.3 AR vs VR vs MR



Figure(37): AR vs VR vs MR

- **AR:** it can be designed for digital elements to appear over real-world views, sometimes with limited interactivity between them, often via smartphones. Examples include Apple's ARKit and Android's ARCore (developer kits), the Pokémon Go game.**[36]**

- **VR:** immersive experiences can be designed by VR which can isolate users from the real world, typically via headset devices. Examples include PSVR for gaming, Oculus and Google Cardboard, where users can explore, e.g. Stonehenge using headset-mounted smartphones.**[36]**

- **MR:** AR and VR elements are combined so that digital objects can interact with the real world; therefore, elements are designed that are anchored to a real environment. Examples include Magic Leap and HoloLens, which users can use.**[36]**

5.4 AR Benefits:

- **Enhanced Communication & Collaboration:**

  Communication and visualization are the two important pillars of any company project. AR successfully showcases analyzed data in the form of tables, charts, infographics and brings them to life by the perfect visual effects.AR and VR also erase the language barrier by its language translation tools which assist in fluent communication.**[37]**

- **Time-Saving Technologies**

  Time is precious and everyone is working hard to manage their time schedules and advanced technologies such as AR and VR just make it super easy and convenient for people who hate wasting their time. These technologies don't only save your time but also provide you with the best of products and services from the comfort of your home.**[37]**

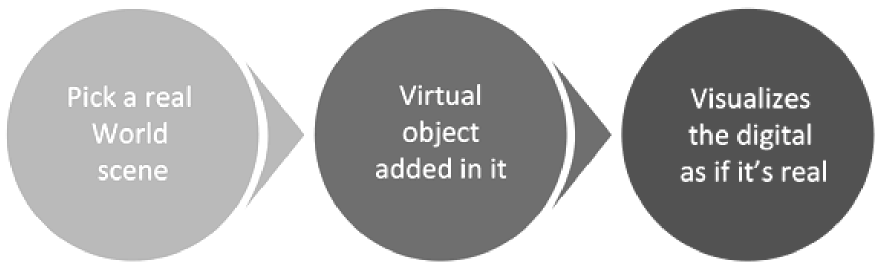
5.5 AR Types:

So far, there are four modes in which augmented reality can be used for practical purposes:**[38]**

- **Marker AR:** has a lot to do with image recognition and requires a camera for scanning.

- **Markerless AR:** provides data on a user's location by utilizing location-tracking devices like GPS, compass, gyroscope, accelerometer.
- **Projection AR:** allows user interaction with holographic images.
- **Superimposed AR:** an object recognition method that replaces an item fully or partially.

5.6 How does AR work?



Figure(38): AR workflow

- **Pick a real world scene :** by collecting data about user's interactions and sending it for processing. Cameras on devices are scanning the surroundings and with this info, a device locates physical objects and generates 3D models. **[39]**
- **Making virtual objects (Processing):** AR devices eventually should act like little computers, something modern smartphones already do.**[39]**
- **Adding virtual objects in it :** This refers to a miniature projector on AR headsets, which takes data from sensors and projects digital content (result of processing) onto a surface to view.**[39]**
- **Visualizes the digital as if it's real :** Some AR devices have mirrors to assist human eyes to view virtual images. The goal of such reflection paths is to perform a proper image alignment.**[39]**

5.7 Helpers used for building AR:

1) **Vuforia engine:**

- Vuforia Engine is a software development kit (SDK) for creating Augmented Reality apps. Developers can easily add advanced computer vision functionality to any application, allowing it to recognize images and objects, and interact with spaces in the real world.**[40]**
- Vuforia Engine offers a variety of trackable targets and capabilities that can be categorized as Images, Objects, and Environments.

- **Vuforia engine Features:[41]**

1) **Tracking Images:**

- **Image Targets:**Attach content onto flat images, such as print media and product packaging.

Figure(39): Image Target

- **Cylinder Targets:**Recognize images wrapped onto objects that are cylindrical or close to in shape (e.g. beverage bottles, coffee cups, soda cans).
- **Multi-Targets :** Use more than one Image Target and arrange them into regular geometric shapes (e.g. boxes) or in any arbitrary arrangement of planar surfaces with Multi Targets.
- **VuMarks :** These are customized markers that can encode a range of data formats. They support both unique identification and tracking for AR applications.

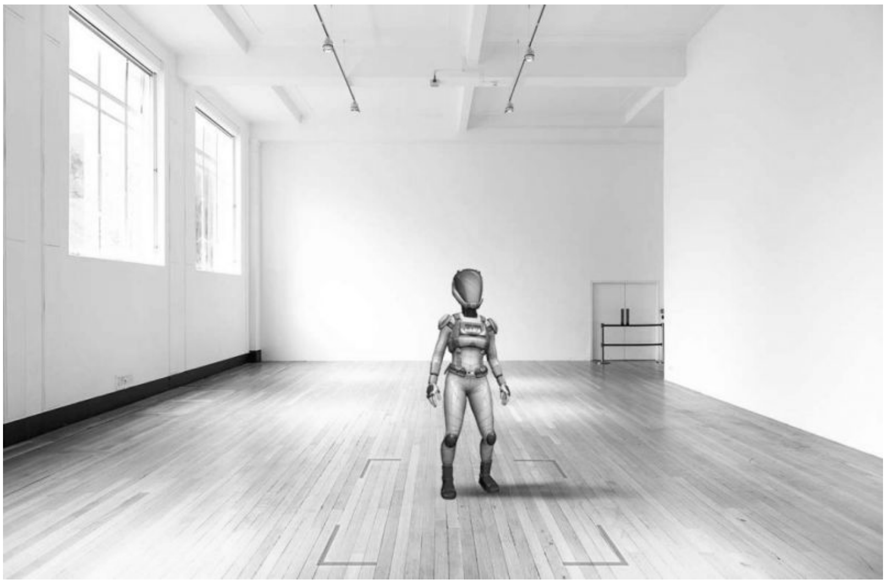
2) Tracking Objects

- **Model Targets :**objects can be recognized by shape using pre-existing 3D models. Place AR content on a wide variety of items like industrial equipment, vehicles, toys, and home appliances.
- **Object Targets :** Created by scanning an object with the Vuforia Object Scanner. It is a suitable option for toys and other products with rich surface details and a consistent shape.

3) Tracking Environments

- **Area Targets :** Augment real environments that you scan using the Vuforia Area Target Creator app or a commercially available 3D scanner. Create accurately aligned persistent content into a wide variety of commercial, public, and fun places to enrich spaces with augmented experiences.

- **Ground Plane :** Enables us to place content on horizontal surfaces in the environment like tables and floors.



Figure(40): Ground Plane

2) ARCore : ARCore is Google's platform for building augmented reality experiences. Using different APIs, ARCore enables your phone to sense its environment, understand the world and interact with information.**[42]**

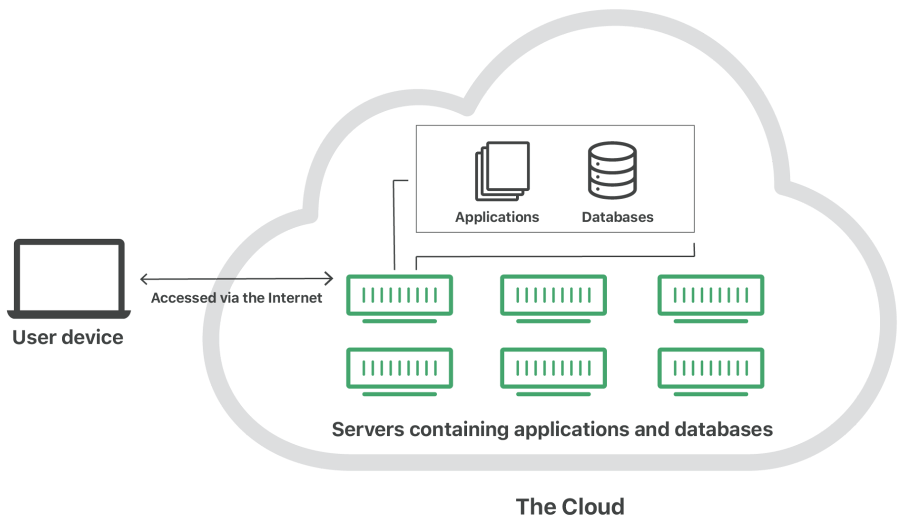
ARCore uses three key capabilities to integrate virtual content with the real world as seen through the phone's camera:**[42]**

- **Motion tracking** allows the phone to understand and track its position relative to the world.
- **Environmental understanding** allows the phone to detect the size and location of all types of surfaces: horizontal, vertical and angled surfaces like the ground, a coffee table or walls.
- **Light estimation** allows the phone to estimate the environment's current lighting conditions.

Chapter Six

6. Cloud

The cloud refers to servers that are accessed over the Internet, and the software and databases that run on those servers. Cloud servers are located in data centers all over the world. By using cloud computing, users and companies don't have to manage physical servers themselves or run software applications on their own machines.



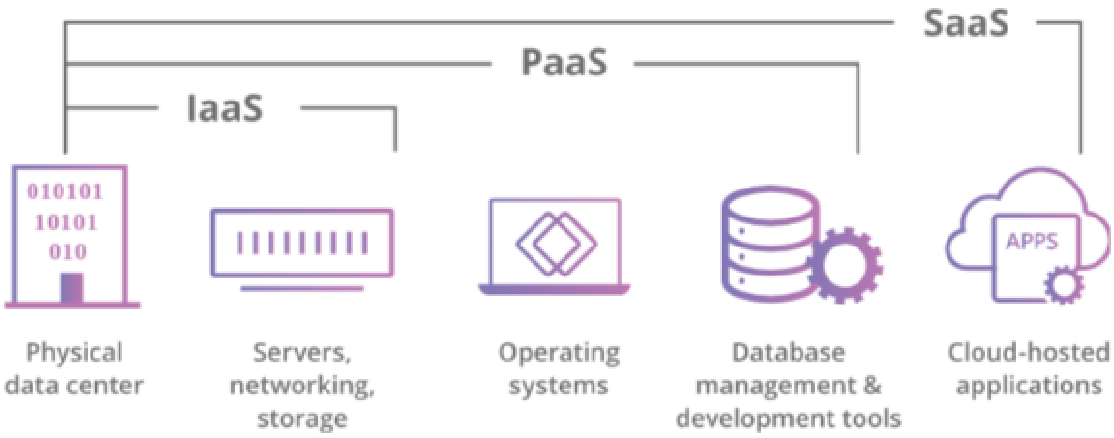
Figure(41): Cloud archtiture

6.1 The main service models of cloud computing

- **Software-as-a-Service (SaaS):**Instead of users installing an application on their device, Saas applications are hosted on cloud servers, and users access them over the Internet Examples of SaaS applications include Salesforce, MailChimp, and Slack.

- **Platform-as-a-Service (PaaS) :** It allows programmers to develop ,test , run and manage the application . PaaS examples include Heroku and Microsoft Azure.
- **Infrastructure-as-a-Service (IaaS) :**

  Iaas is also known as Hardware as Services(Haas). It is computing infrastructures managed over the internet. The main advantage of using Iaas is that it helps users to avoid the cost and complexity of purchasing and managing the physical servers .



Figure(42): Service models of cloud computing

6.2 Types of cloud

There are three main cloud options, based on different access models: public, private and hybrid.

Types	Explanation
Public Cloud	• It supports services that provide a multi-tenant storage environment that is most suited for unstructured data on a subscription basis.
Private Cloud	• It is an in-house storage resource deployed as a dedicated environment protected behind a firewall. • Internally hosted private cloud storage implementations

	emulate some of the features of commercial public cloud services, providing easy access and allocation of storage resources for business users, as well as object storage protocols.
Hybrid Cloud	● This cloud storage option is a mix of private cloud storage and third-party public cloud storage services, with a layer of orchestration management to operationally integrate the two platforms.

6.3 Cloud Storage

Cloud storage is a service model in which data is transmitted and stored on remote storage systems, where it is maintained, managed, backed up and made available to users over a network, typically, the internet.

Cloud storage is based on a virtualized storage infrastructure with accessible interfaces, near-instant elasticity and scalability, multi-tenancy,and metered resources.

Cloud-based data is stored in logical pools across disparate, commodity storage servers located on premises or in a data center managed by a third-party cloud provider.**[58]**

6.3.1 Types of Cloud Storage:

Type	Explanation
Block storage	● It divides large volumes of data into smaller units called blocks. ● Each block is associated with a unique identifier and placed on one of the system's storage drives. ● Block storage has multiple advantages: ○ Fast. ○ efficient.

	○ provides the low latency required by applications such as databases. ○ high-performance workloads.
File storage	● organizes data in a hierarchical system of files and folders; it is commonly used with personal computer storage drives and network-attached storage (NAS). ● Data in a file storage system is stored in files, and the files are stored in folders. ● Directories and subdirectories are used to organize the folders and locate files and data. ● A file storage-based cloud can make data access and retrieval easier, with this hierarchical format being familiar to users and required by some applications.
Object storage	● stores data as objects. ● which consist of three components: ○ data stored in a file. ○ metadata associated with the data file. ○ unique identifier ● Object-based storage systems allow metadata to be customized, which can streamline data access and analysis.

6.3.2 Cloud Storage Firebase

Firebase is a platform for web and mobile application development. It was created by Firebase, Inc in 2011 and acquired in 2014 by Google.Firebase storage is storage incorporated into a fire-base platform.**[57]**

- There are multiple methods which are used for storage to :
  - Simple Management .
  - Monitor our states.
  - improve our security.

- **Key capabilities**
  - Robust operations.
  - Strong security
  - High scalability

6.3.3 Advantages and Disadvantages

Advantages

- **Off-site management**

  This frees your staff from tasks associated with storage, such as procurement, installation, administration, and maintenance.

- **Quick implementation**

  Using a cloud service accelerates the process of setting up and adding to your storage capabilities

- **Cost-effective**

  you pay for the capacity you use.

- **Scalability**

  Growth constraints are one of the most severe limitations of on-premise storage

Disadvantages

- **Security :**

  Cloud storage providers try to secure their infrastructure with up-to-date technologies and practices, but occasional breaches have occurred, creating discomfort with users.

- **Administrative control:**

  Being able to view your data, access it, and move it at will is another common concern with cloud resources

- **Latency**: : Delays in data transmission to and from the cloud can occur as a result of traffic congestion, especially when you use shared public internet connections.

6.4 IBM Cloud

IBM Cloud offers the most open and secure public cloud platform for business, a next-generation hybrid multi cloud platform, advanced data and AI capabilities, and deep enterprise expertise across 20 industries.**[45]**

IBM Cloud: is a full stack cloud platform with over 170 products and services covering data, containers, AI, IoT, and blockchain.**[44]**

It provides many different types of services such as:**[43]**

- **Speech:**Convert text and speech with the ability to customize models.

  Examples : Speech To Text (STT) , Text To Speech (TTS)

- **Language:**Analyze text and extract meta-data from unstructured content.

  Examples :Language Translator , Natural Language Classifier.

- **Empathy:** Understand tone, personality, and emotional state.

  Examples : Tone Analyzer , Personality insights.

6.4.1 IBM Watson Speech To Text (STT)

Figure(43): IBM Speech To Text

The IBM Watson Speech to Text service provides APIs that use IBM's speech-recognition capabilities to produce transcripts of spoken audio. The service can transcribe speech from various languages and audio formats. In addition to basic transcription, the service can produce detailed information about many different aspects of the audio. It returns all JSON response content in the UTF-8 character set.**[46]**

This service uses deep-learning AI to apply knowledge of grammar, language structure, and the composition of audio and voice signals to accurately**[46]**

Speech to Text: Customer Care is ideal for clients who need to extract high-quality speech transcripts from call center audio.**[46]**

Watson Speech to Text(STT) features

- **Powerful real-time speech recognition**

Automatically transcribe audio from 7 languages in real-time. Rapidly identify and transcribe what is being discussed, even from lower quality audio, across a variety of audio formats**[47]**

- **Highly accurate speech engine**

  Customize your model to improve accuracy for language and content you care most about, such as product names, sensitive subjects or names of individuals. Recognize different speakers in your audio and spot specified keywords in real-time with high accuracy and confidence.**[47]**

- **Built to support various use cases**

  Transcribe audio for various use cases ranging from real-time transcription for audio from a microphone, to analyzing 1000s of audio recording from your call center to provide meaningful analytics.**[47]**

Advantages:

- Processes unstructured data.
- Assists humans instead of replacing them.
- Helps overcome human limitations.
- Improves productivity by delivering relevant data.
- Improves user experience.
- Can process large quantities of data.

- Easy to set up and get started with.**[48]**

Disadvantages:

- Doesn’t directly support structured data.
- Expensive to switch to.
- Requires maintenance.
- Only supports a limited number of languages.
- Takes time to implement fully.
- Requires education and training to make full use of its resources.**[48]**

.

Chapter Seven

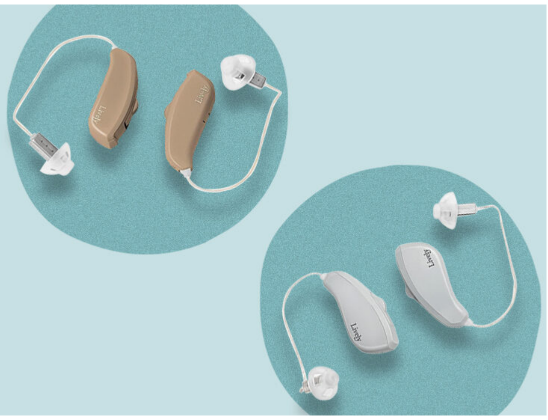
7. Previous Solutions

7.1 Traditional Solutions

7.1.1 Hearing Aids:

History: The first electric hearing aid was invented in 1898 by Miller Reese Hutchison. His design used an electric current to amplify weak signals.In 1913, the world was introduced to the first commercially manufactured hearing aids. These devices were cumbersome and not very portable. In the 1920s vacuum-tube hearing aids were produced; these tubes were able to turn speech into electric signals and then the signal itself was amplified.**[51]**

What is hearing Aids?



Figure(44): Hearing Aids

A hearing aid is a battery-powered electronic device designed to improve your hearing. Small enough to wear in or behind your ear, they make some sounds louder. They may help you hear better when it's quiet and when it's noisy. It works by using:**[52]**

- **A microphone** picks up sound around you.
- **An amplifier** makes the sound louder.

- **A receiver** sends these amplified sounds into your ear.

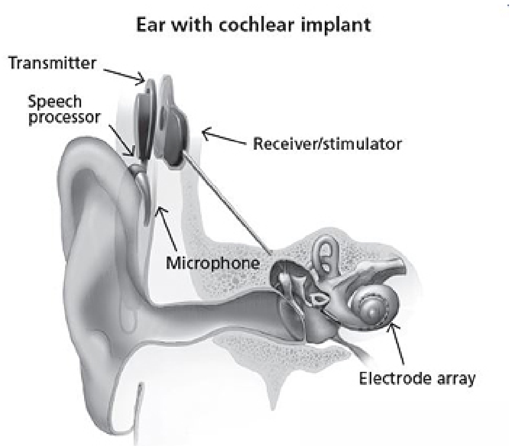
Disadvantages of Hearing Aids:

1) Hearing aids require an adjustment period that may take several months. Follow-up visits with the licensed hearing aid dispenser may be needed to take full advantage of the hearing aids.
2) Hearing Aids have batteries with various battery life so the deaf person must change batteries regularly.
3) These small hearing aids aren't suitable for people with severe, more advanced hearing loss.
4) Hearing aids can be expensive.**[53, 54]**

7.1.2 cochlear implant:

History: The first cochlear implant was performed by Professor Kurt Burien in Vienna on December 16, 1977. A few months later a 48-year-old patient denoted MC-1 in was implanted in Melbourne in 1978.5 Since that implantation, advances in medical technology have led to continual miniaturization of the components.

What is a cochlear implant?



Figure(45): Cochlear implant

A cochlear implant is a small, complex electronic device that can help to provide a sense of sound to a person who is profoundly deaf or severely hard-of-hearing.

The implant consists of an external portion that sits behind the ear and a second portion that is surgically placed under the skin (see figure).**[49]**

An implant has the following parts:

- A microphone, which picks up sound from the environment.
- A speech processor, which selects and arranges sounds picked up by the microphone.
- A transmitter and receiver/stimulator, which receive signals from the speech processor and convert them into electric impulses.
- An electrode array, which is a group of electrodes that collects the impulses from the stimulator and sends them to different regions of the auditory nerve.

- An implant does not restore normal hearing. Instead, it can give a deaf person a useful representation of sounds in the environment and help him or her to understand speech.

The Difference between cochlear implant and Hearing Aids:

A hearing aid is also a medical device for hearing loss. But unlike a cochlear implant, it doesn't transmit sound signals via electrodes.

Instead, hearing aids use a microphone, amplifier, and speaker to make sounds louder. This can help you hear things better.

Also, hearing aids aren't surgically implanted. They're worn inside or behind the ear.

Hearing aids are typically ideal if you have mild to moderate hearing loss. The device's level of amplification depends on your degree of hearing loss.

Certain hearing aids may help severe hearing loss, but sometimes they still won't benefit speech understanding. In this case, a cochlear implant might be the better choice.**[50]**

Disadvantage of cochlear implant:

- Losing remaining natural hearing in the ear with the implant
- Regularly recharging batteries or using new ones
- Cochlear implants require ongoing maintenance.
- It can be prohibitively expensive.
- Damage to the implant during sports activity or accidents.

7.2 Modern Solutions:

7.2.1 Hardware gloves

The basic idea dates to the 1980s, when researchers started exploring how humans could interact with computers using gestures. In 1983, a Bell Labs engineer named Gary Grimes invented a glove for data entry using the 26 manual gestures of the American Manual Alphabet, used by speakers of American Sign Language. But the first glove intended to make interactions between deaf and non-deaf people easier was announced in 1988 by the Stanford University researchers James Kramer and Larry Leifer. It was called the "talking glove"

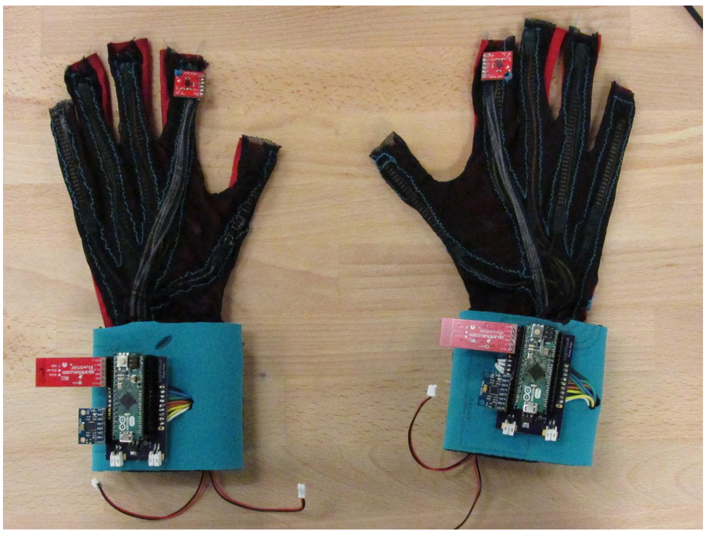
1) Wearable-tech glove

Figure(46): Wearable-tech glove

The system includes a pair of gloves with thin, stretchable sensors that run the length of each of the five fingers. These sensors, made from electrically conducting yarns, pick up hand motions and finger placements that stand for individual letters, numbers, words and phrases.

The device then turns the finger movements into electrical signals, which are sent to a dollar-coin–sized circuit board worn on the wrist. The board transmits those signals wirelessly to a smartphone that translates them into spoken words at the rate of about a one word per second.**[56]**

2) SignAloud Gloves:



Figure(47): SignAloud glove

“SignAloud,” is a pair of gloves that can recognize hand gestures that correspond to words and phrases in American Sign Language. Each glove contains sensors that record hand position and movement and send data wirelessly via Bluetooth to a central computer. The computer looks at the gesture data through various sequential statistical regressions, similar to a neural network. If the data match a gesture, then the associated word or phrase is spoken through a speaker.**[55]**

- **Disadvantages of Hardware Gloves:**
  1) Hardware gloves would not have access to facial expressions which is an important feature in sign language.
  2) Hardware costs a lot of money so this solution is so expensive.
  3) It is not comfortable to deaf people to wear gloves, especially on summer days.

7.2.2 DeepASL

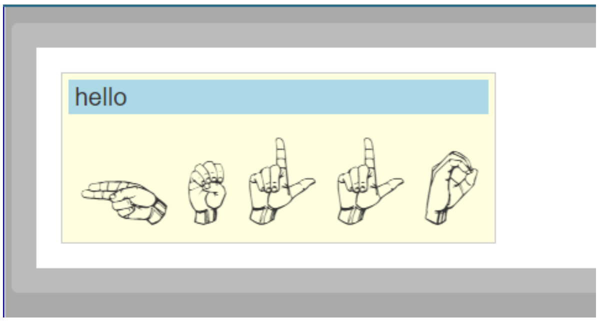
DeepASL, uses a camera device to capture hand motions, then feeds the data through a deep learning algorithm, which matches it to signs of ASL. Unlike many previous devices, DeepASL can translate whole sentences rather than single words, and doesn’t require users to pause between signs.

DeepASL can help people who are deaf and hard of hearing by serving as a real-time translator. It could be especially useful in emergency situations.

7.2.3 WeCapable website

WeCapable Tool easily converts English text into sign language symbols. This tool will be very useful for both teaching and learning American sign language.

Translation of text to sign language is also given as a task during sign language study sessions. This tool can easily produce the correct answers and because the visual stays on screen, students can follow the hand movements at their own pace.

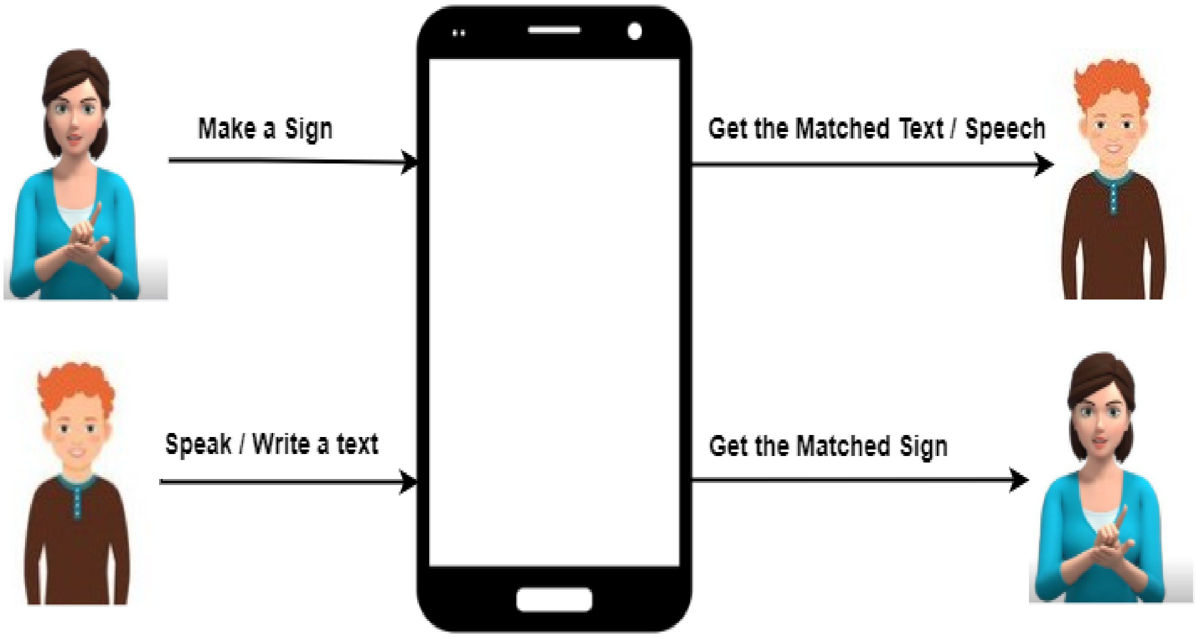


Figure(48): WeCapable website

Chapter Eight

8. Proposed Methods

8.1 System Overview



Figure(49): BI Translator overview

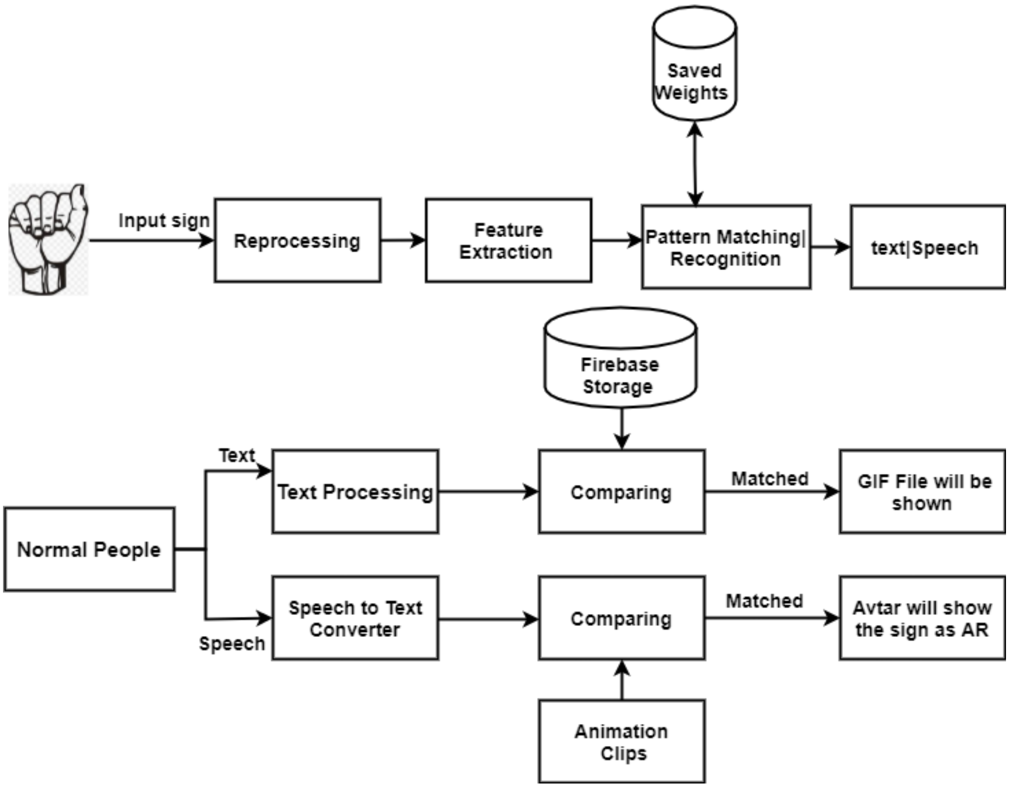
BI Translator is a mobile application that helps deaf and normal people to communicate with each other . It helps them by presenting two phases:

- The first phase helps them by translating from ASL to Speech / Text
- The second phase helps them by translating from Speech / Text to ASL.

In this chapter the two phases will be discussed in detail.

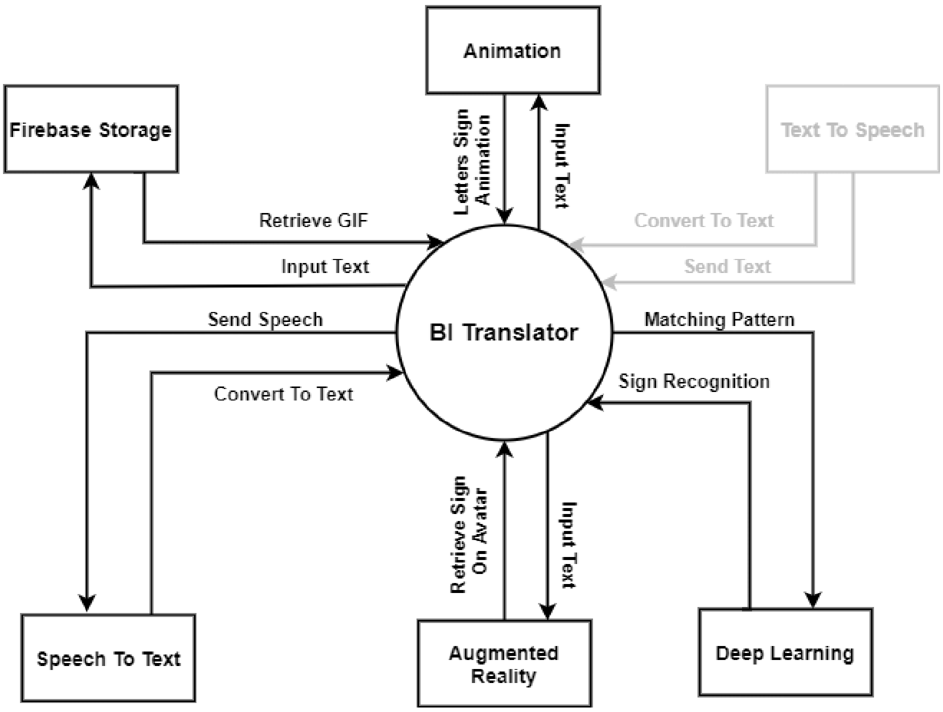
8.2 System Diagrams

8.2.1 System Block Diagram



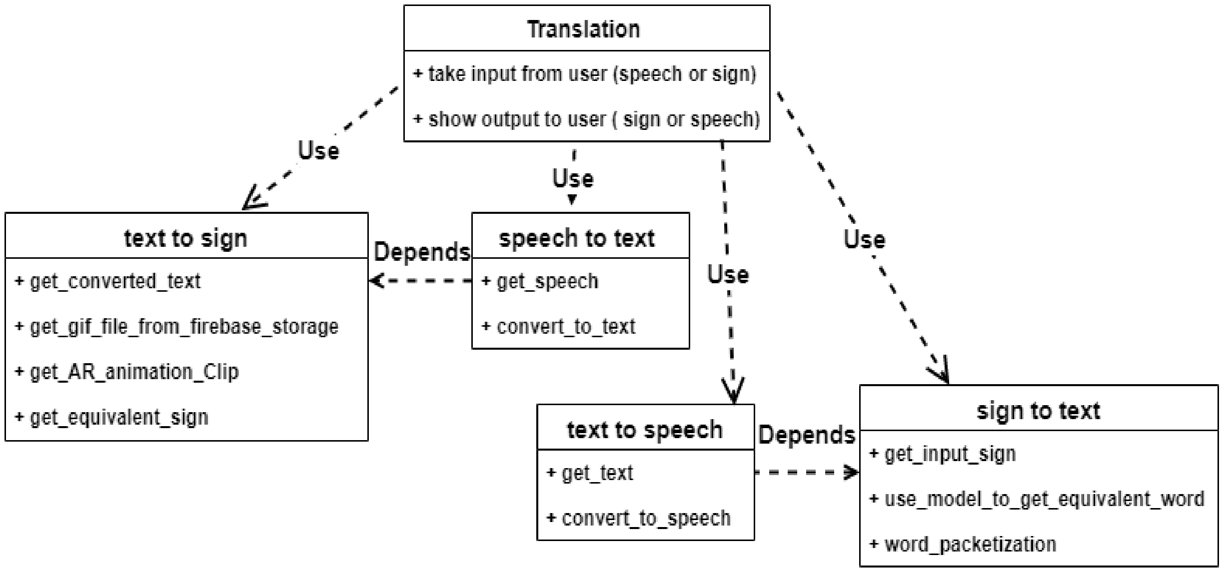
Figure(50): System Block Diagram

8.2.2 System Context Diagram



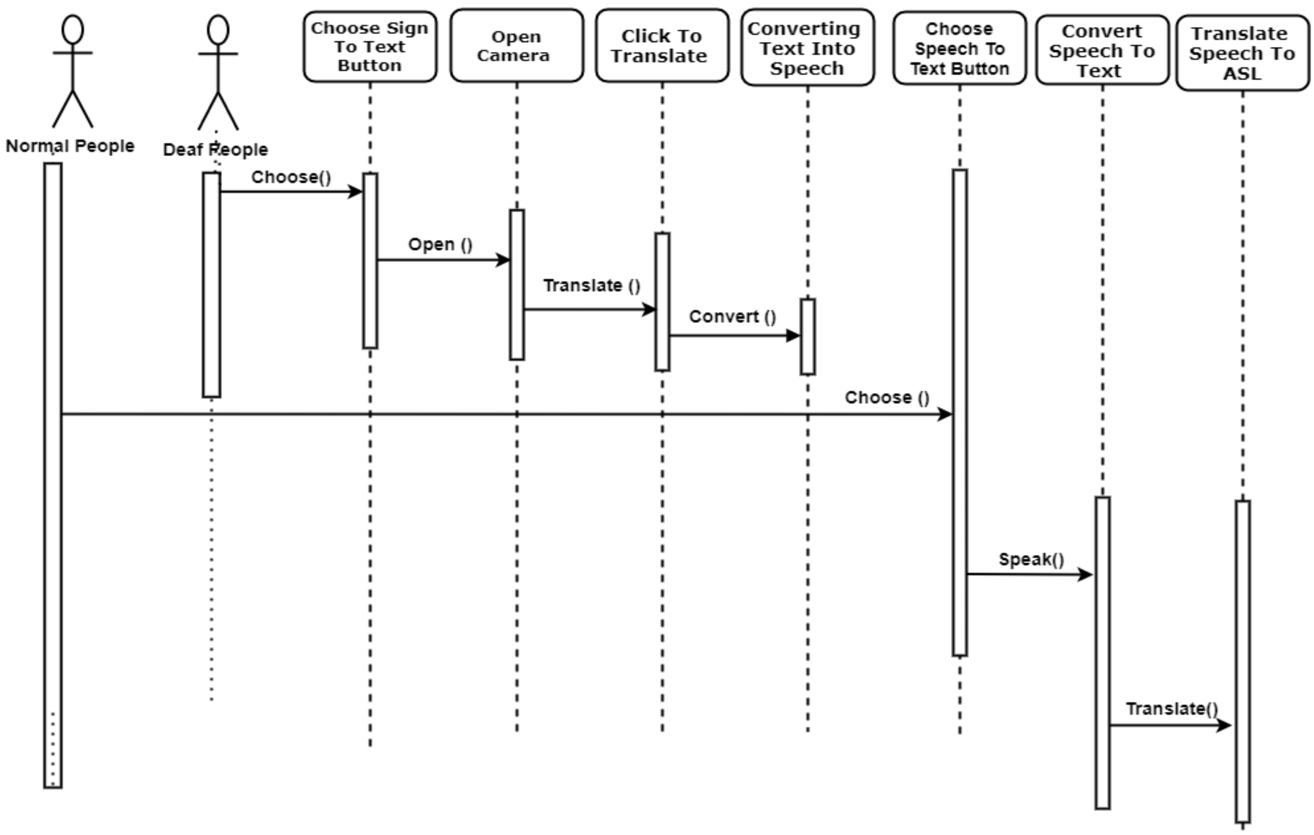
Figure(51): System Context Diagram

8.2.3 System Class Diagram



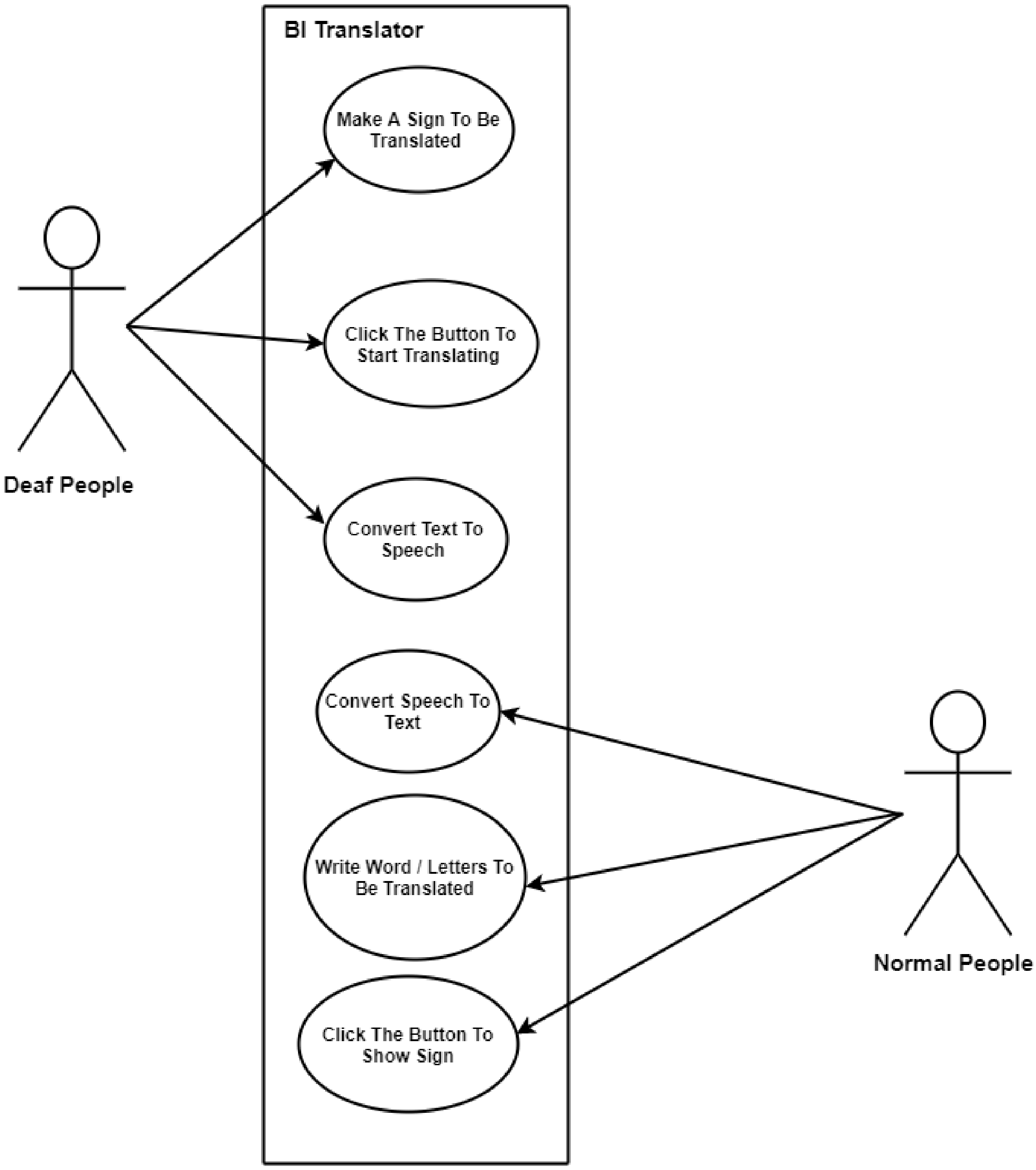
Figure(52): System Class Diagram

8.2.4 System Sequence Diagram



Figure(53): System Sequence Diagram

8.2.5 System Use Case Diagram



Figure(54): System Use Case Diagram

8.3 ASL To Text | speech

8.3.1 CNN Approach

8.3.1.1 Flowchart

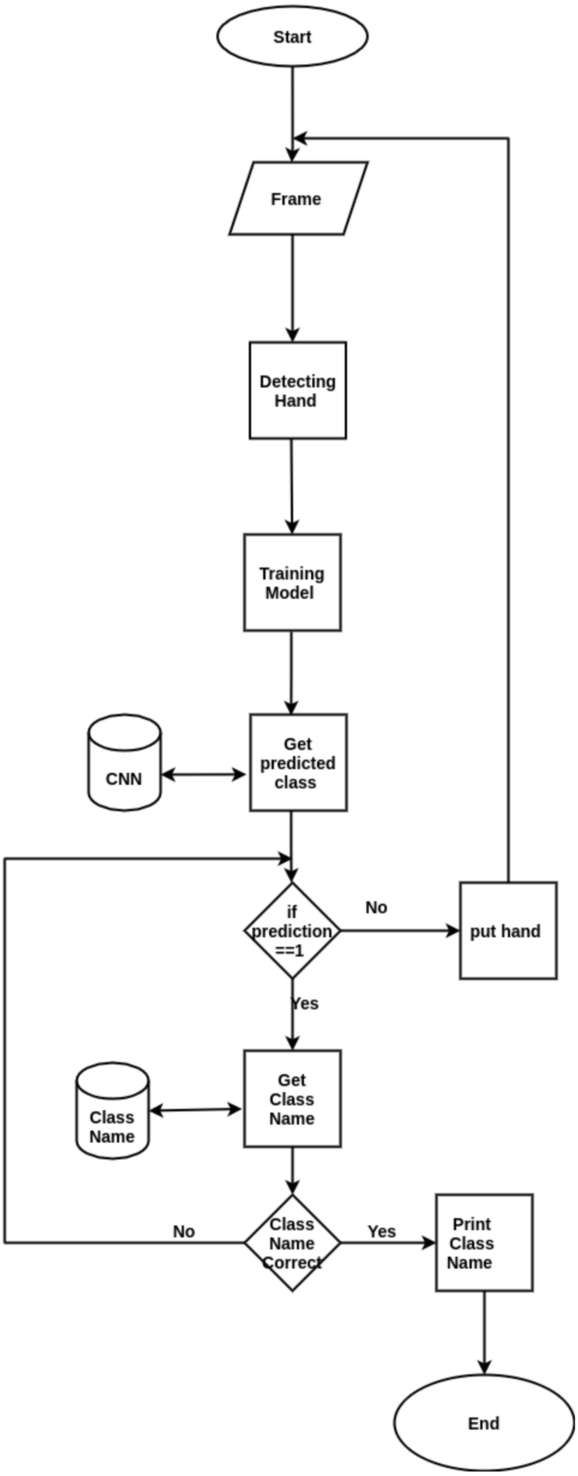


Figure (55):First approach flowchart

8.3.1.2 Pseudo code

Procedure Cnn-model (data-Set-Path , Epoch_numbers)

Begin
Data set reprocessing
Model training get callbacks
Model fitting get history of training
Plot learning curve
End

8.3.1.3 Implementation and Results

1-Model Implementation .

Layer (type)	Output Shape	Parameters
conv2d (Conv2D)	**(None, 28, 28, 32)**	**320**
max_pooling2d (MaxPooling2D)	**(None, 14, 14, 32)**	**0**
conv2d_1 (Conv2D)	**(None, 14, 14, 64)**	**18496**
max_pooling2d_1 (MaxPooling2	**(None, 7, 7, 64)**	**0**
conv2d_2 (Conv2D)	**(None, 7, 7, 128)**	**73856**
max_pooling2d_2 (MaxPooling2	**(None, 3, 3, 128)**	**0**
conv2d_3 (Conv2D)	**(None, 3, 3, 128)**	**147584**
max_pooling2d_3 (MaxPooling2	**(None, 1, 1, 128)**	**0**

dropout (Dropout)	**(None, 1, 1, 128)**	**0**
flatten (Flatten)	**(None, 128)**	**0**
dense (Dense)	**(None, 128)**	**16512**
dense_1 (Dense)	**(None, 26)**	**3354**

Total Parameter

Total params: 260,122

Trainable params: 260,122

Non-trainable params: 0

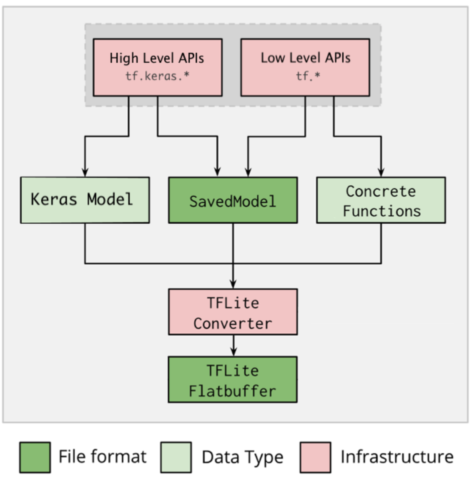
Data Set : Sign Language Mnist.

Model Integration

Model integration applied by using a TensorFlow-Lite to apply classification in mobile applications.

TensorFlow-Lite

The TensorFlow Lite converter takes a TensorFlow model and generates a TensorFlow Lite model (an optimized Flat -Buffer format identified by the .tflite file extension).



Figure(56): TensorFlow-Converter Process.

Android Application

In this approach we use mobile applications
as an Interface for showing sign recognition.

1- CNN Results

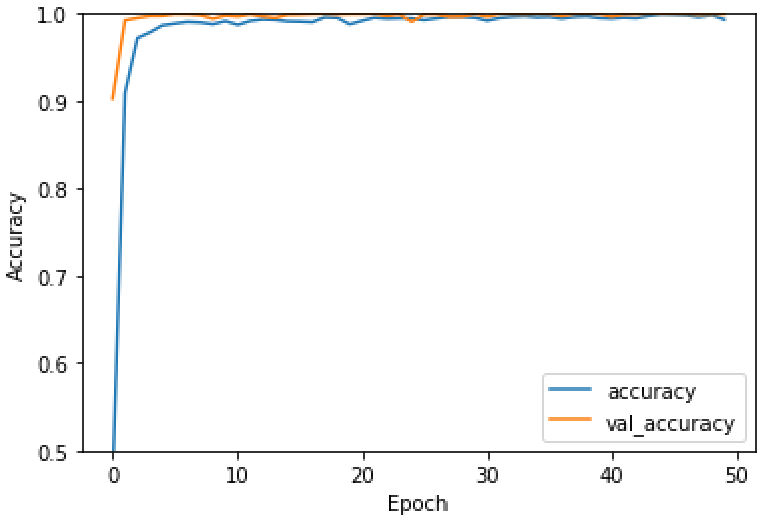


Figure (57): CNN Learning Curve.

Accuracy : 94 %

1- Translation Results

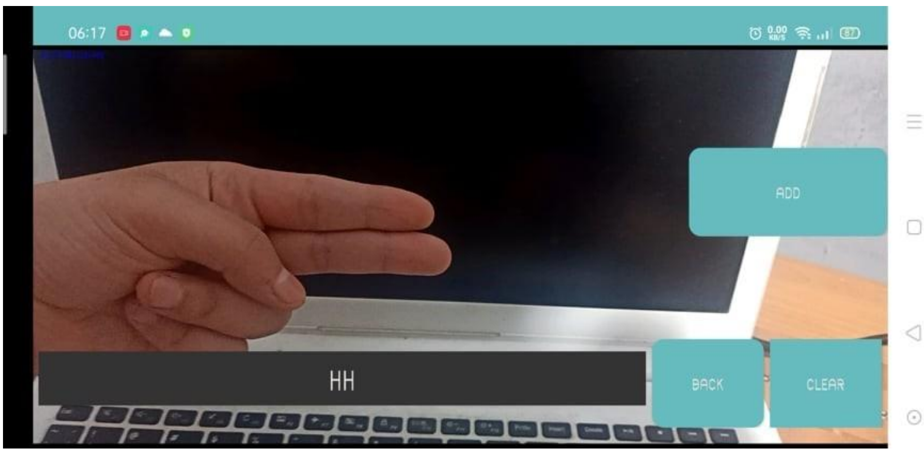


Figure (58):Bi-Translator Translation

8.3.1.4 Limitations of CNN approach

1-Simple Architecture.

2- Translate simple static alphabets.

3- Slow Translation.

4- poor dataset .

8.3.2 SSD - Faster R-CNN Approach

In this approach the system receives frames and sends them to SSD which apply hand detection then apply Faster R-CNN to start recognition signs.

8.3.2.1 Flowchart

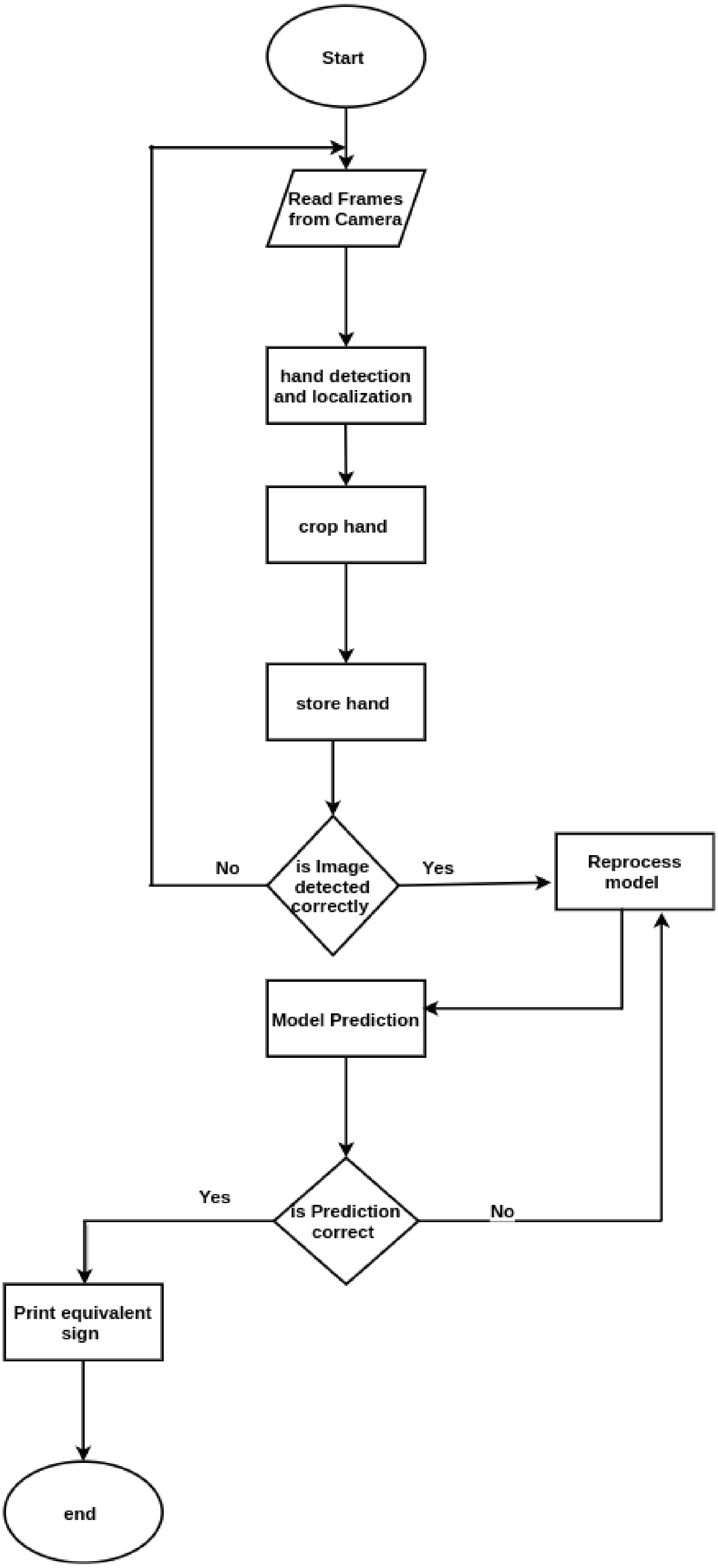


Figure (59):second approach flowchart

8.3.2.2 Pseudo code

Procedure Model_Training(dataset_path , epochs)
    Begin
        Data_set reprocessing
        SSD hand detection
        Extract hand frames
        Start contouring hand
        Store contouring hand
        Model training get callbacks
        Model testing and get history
    End

8.3.2.3 Implementation and Results

SSD hand detection

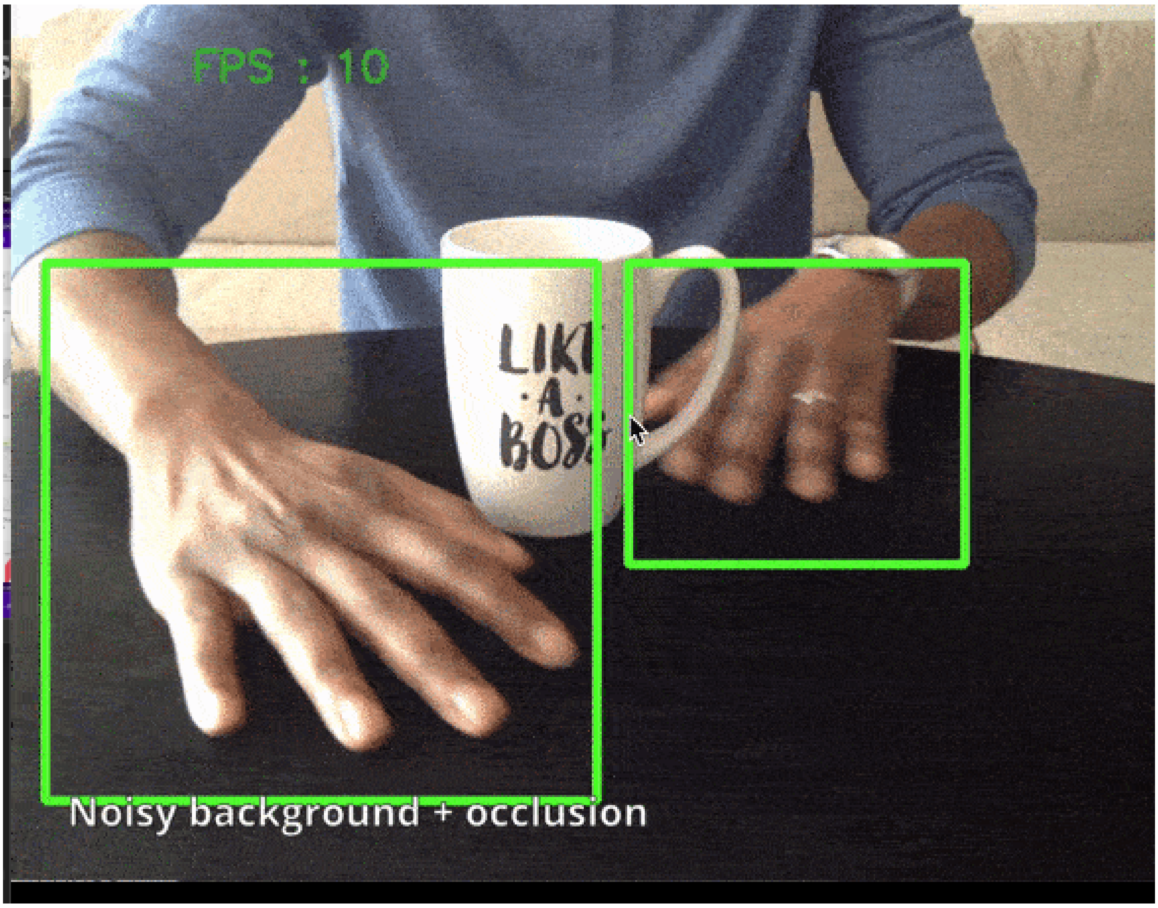


Figure (60):SSD Hand Detection

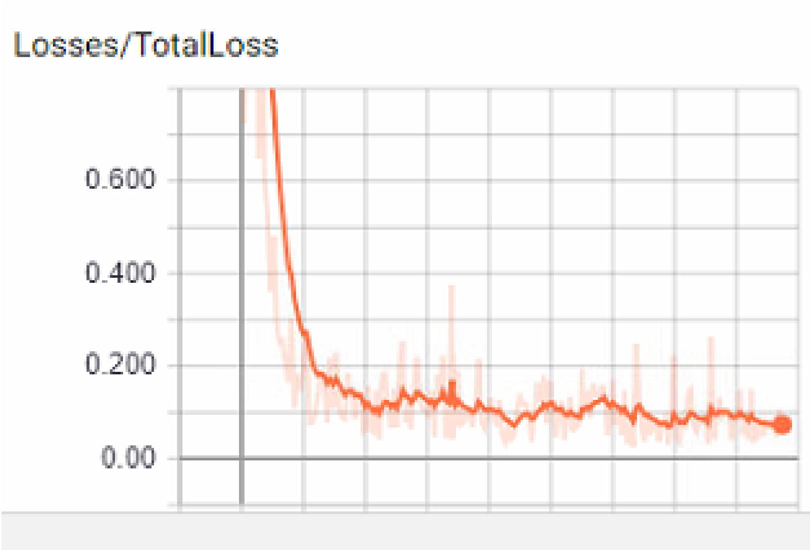


Figure (61): Fast-RCNN Learning Curve.

Accuracy : 93 %

8.3.2.4 Limitation

1- It needs Too much data set for the training process.
2- Very slow in Training process.
3- Slow translation of each sign .
4- bad detection of sign language for the similar signs like (

8.3.3 mediaPipe and sklearn Approach

In This approach we will discuss the main parameters to describe words in ASL and applying mediapipe with sklearn to recognize sign language.
The parameters are:

- handshape .
- palm orientation - movement.
- location .
- expression/non-manual signals.

All five parameters must be performed correctly to sign the word Accurately .

Now we will know the role of facial expressions as an important part of communication by knowing a simple sign like “YOU”

This version of “YOU” is actually a question that will be equivalent to the below facial expression.

 [63]

The same sign with a different movement for the sign “YOU” with the question “ "It's *you?”*

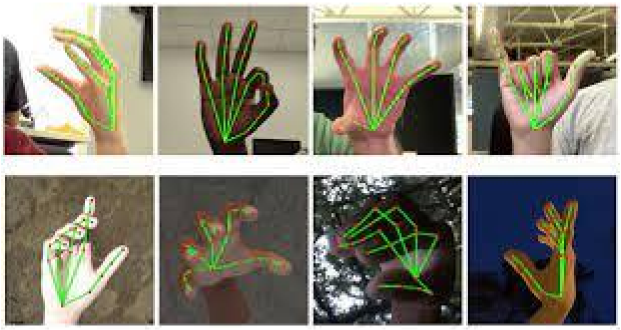
 [63]

From the two different images we will discuss the facial recognition role with sign language recognition .

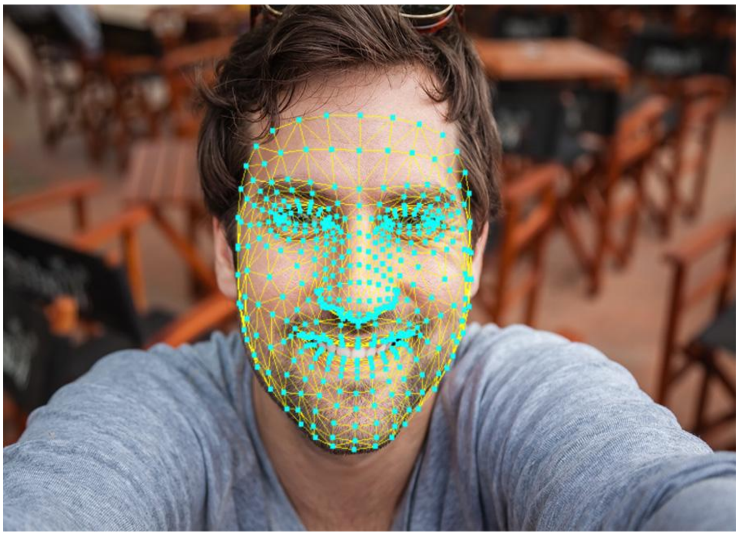
8.3.3.1 MediaPipe

The hand tracking is done with the open source Google MediaPipe technology. This is a framework that provides real time computer vision technologies such as hand detection, hand tracking, face detection or object detection. Launched in 2019, MediaPipe is the latest technology advancement that makes this project possible.

The Google MediaPipe technology provides detailed real time finger tracking with multiple hands , face detection and multiple others detection models. In the images below it is shown how the hand detection works. **1 [64]**



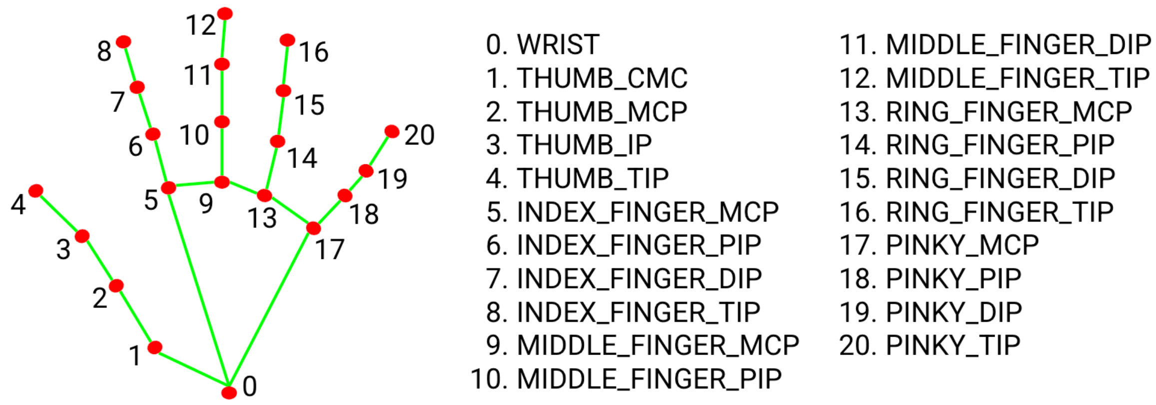
figure(62):mediapipe hand detection



figure(63):mediapipe face-mesh detection .

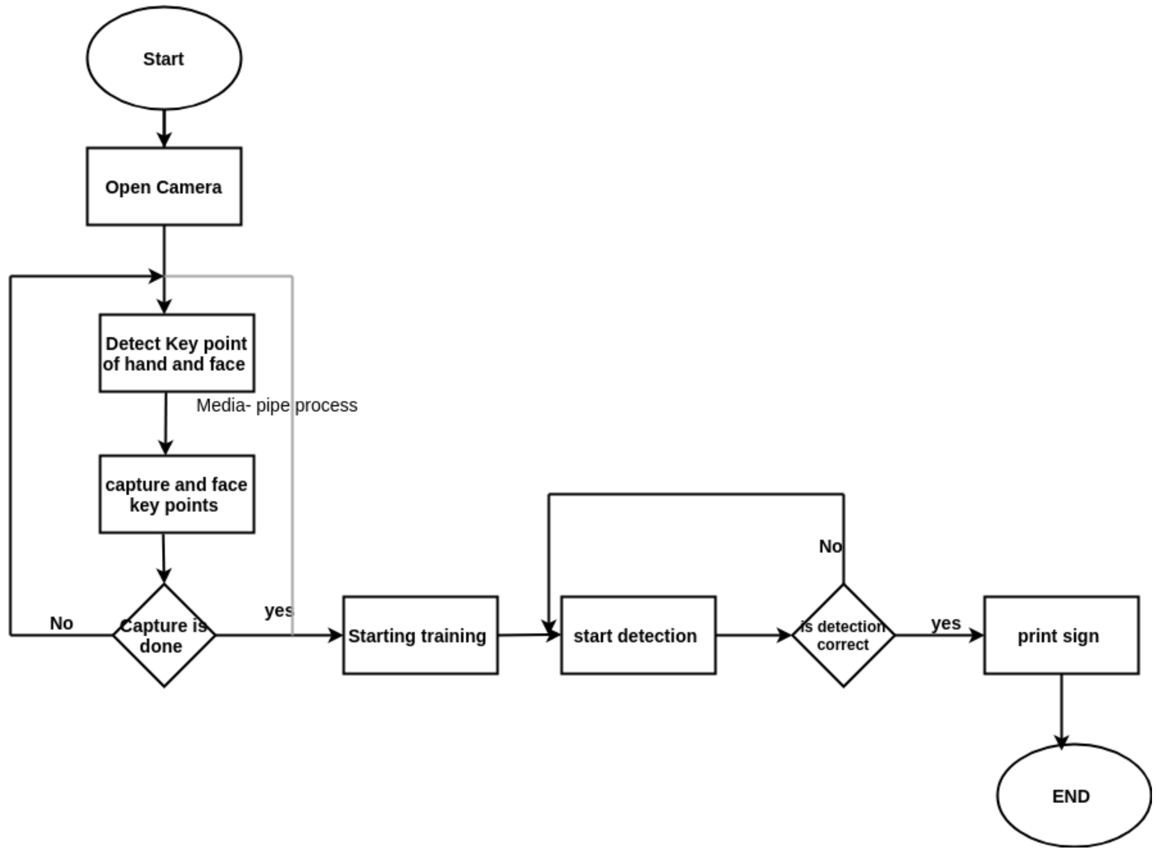
Hand Landmark Model

Hand has a localized keypoint of 21 3D hand-knuckle coordinates inside the detected hand regions .



figure(64):hand landmarks

8.3.3.2 Flowchart



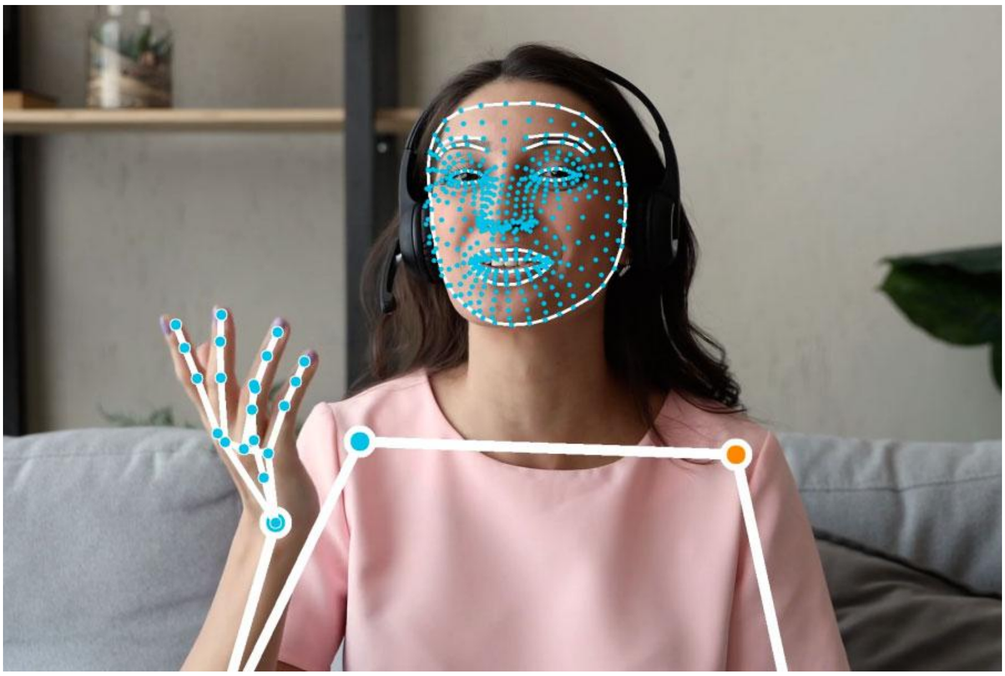
figure(65):Third approach flow-chart

8.3.3.3 Pseudo code

```
Procedure Mediapipe_Traning():
        Begin
          Open camera
           If there is no hand and face
                End
           Else
                Detect hand and face keypoints
          Capture key points
          Store key points
          Start training
          Recognize sign
          If recognition is true
                Print sign
        Else
                Detect sign
End
```

8.3.3.4 Implementation and Results

Accuracy of model : 99 %



figure(66):Media-pipe Holistic solution .

8.4 Speech | Text to ASL

In this phase we will translate natural speech to ASL by two different techniques.To achieve these two techniques we need to create an **Avatar** to represent signs.

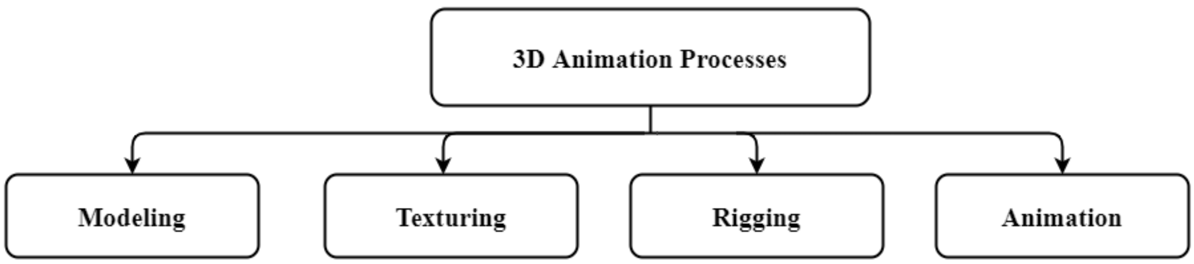
To create this Avatar we used a software tool called **Blender** which supports 3D Animation.

Blender

- Blender is a free and open source 3D modeling and animation suite.
- Blender is a computer graphics program that allows you to produce high-quality still images and animations using three-dimensional geometry.

8.4.1 Avatar Creation

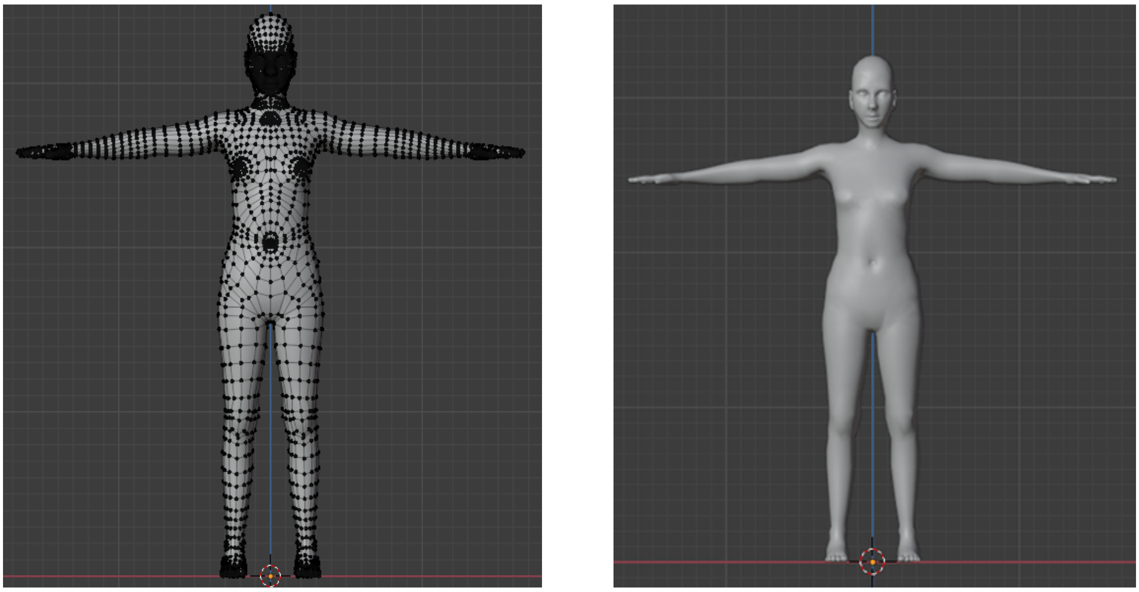
- We used 3D Animation Processes to complete the creation of the Avatar.



Figure(67): 3D Animation Process

1) **Modeling:** is a mesh of points, lines, and curves arranged in a way to map out an object.

   But for our avatar this step is done by Searching and getting the appropriate model.**[59]**



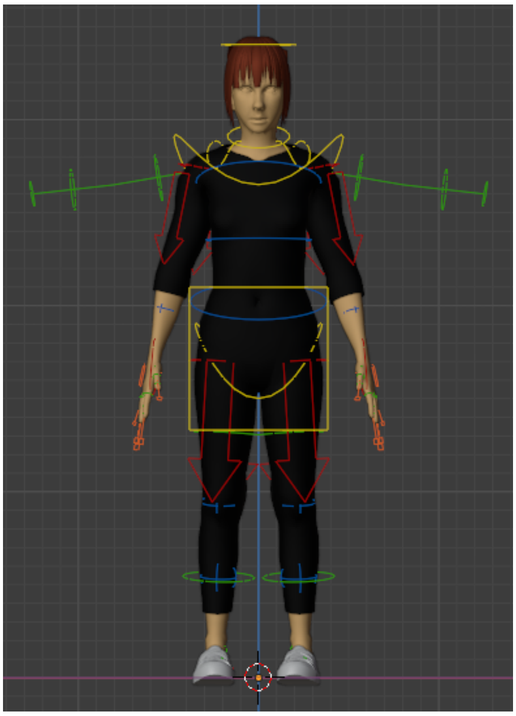
Figure(68): Modelling Process

2) **Texturing:**Adding clothing and accessories to your character model.
   By creating new layers , that makes the moving of the outlines for the new layer to reflect how clothes are. **[60]**

Figure(69):Texturing Process

3) **Rigging:**

- Rigging is the process of connecting an armature(Skeleton) to a mesh to make it more purposeful for animation and make the avatar alive. Rigging step happens in Edit mode.
- Rigging has multiple features as:
    1) Armatures
    2) Constraints
    3) Object Modifiers
    4) Shape Keys
    5) Drivers



Figure(70):Rigging Process

4) **Animation:** has an effective role in our translation by using the avatar , which is by making an object move or change shape over time. Animation is happening after rigs generations in Pose Mode.

- We can apply animation in multiple ways as:
    - **Moving as a whole object:** which means that changing their position , orientation or size in time

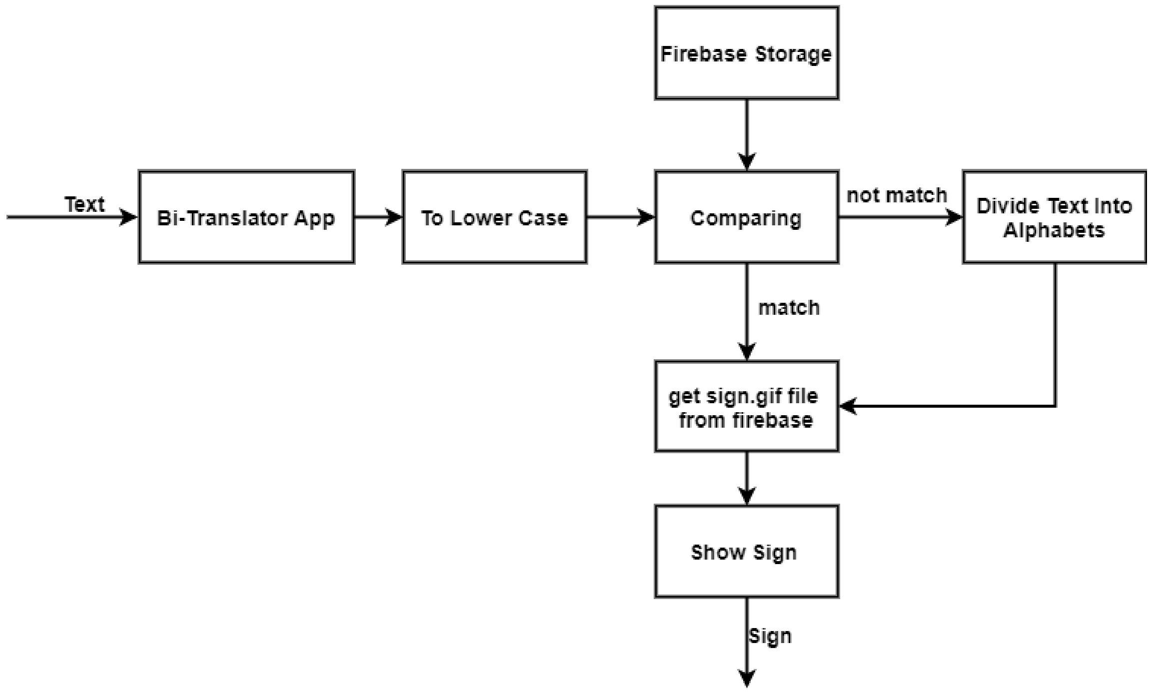
- **Deforming them:** which means Animating their vertices or control points.

- We made animations to letters from A to Z and words (hello , i , need , coffee).

Figure(71):Animation Process (Letter K)

- We used the animations of this **Avatar** with two different roles in these ways .

8.4.2 First Way : Letter Translation



Figure(72):Letter Translation

- In this technique we will translate **text** into sequence of signs by using gif files of generated animations of letters that are stored in the firebase storage. The equivalent gif file (sign) will be shown if the input text is a letter (in case of storage matching) ,otherwise the input text will be tokenized to its letters and equivalent gif files (sign) will be shown (in case of storage mismatching).

1) **Usage of firebase in this way:**

   - We use Firebase Storage to store and manage media and files generated by bi-translator applications and retrieve data from storage to our BI Translator application .

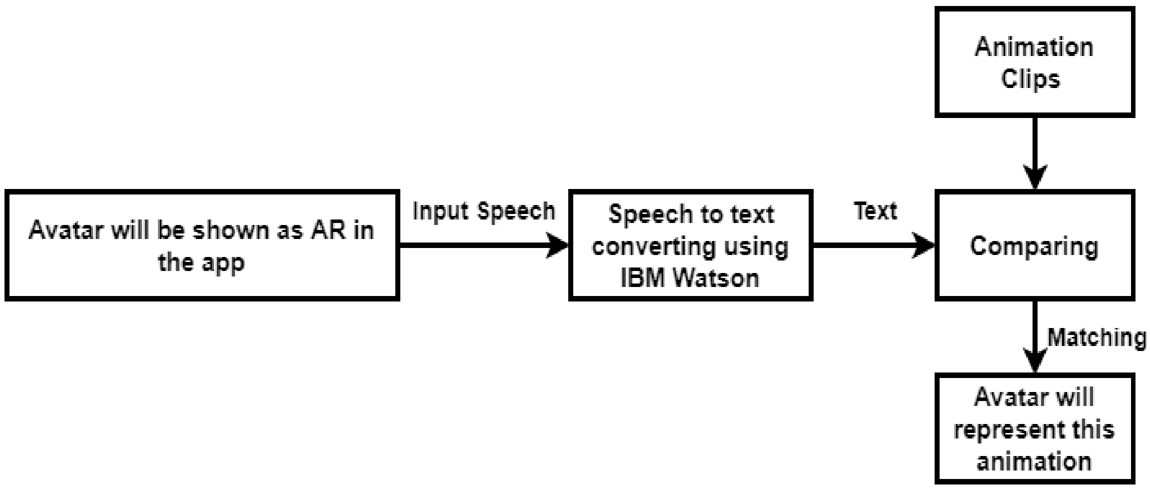
   - We used it by integrating fire-base SDK into android studio and used it by programming.

2) **GIF Files Generation:**

We generate GIF files using **krita** that takes frames of each letter animation from blender with the help of ffmpeg.exe.

8.4.3 Second Way: Words Translation

- In this technique we used Augmented Reality to translate words into signs by using unity3D platform.
- **Unity**: it is a game development ecosystem of powerful rendering engine, intuitive tools, rapid workflows for 2D and 3D games. It supports multiple platforms such as android , ios,...etc
- Our avatar will be shown in BI Translator using AR Core. If IBM Watson recognizes any speech it will convert it into text. This text will be compared with animation clips. Avatar will represent this equivalent animation for this text(in case of matching).



Figure():Words Translation

1) **Importing avatar from Blender to unity**

- Rigging of Blender Platform is not suitable for unity platform so we used a converter to help us convert Rigify rig to Humanoid to get the right rigging.**[61]**

- When we import the avatar , Animation clips made by blender are imported with it and we can control the transitions between animation clips by using Animator Controller.
- An Animator Controller allows you to arrange and maintain a set of Animation Clips and associated Animation Transitions for a character or object.**[62]**

2) Allocation of the Avatar:

Firstly we used Vuforia Engine to allocate our Avatar , secondly we used AR Foundation and ARCore to improve its performance.

1) Vuforia engine:

- **Image Target**

We used an image target from vuforia engine which depends on allocation of the Avatar on a specific image.

Figure(73): Avatar Allocation Using Image Target

Limitations of Image Target

- According to our Project , it is not the best way to use a specific image in real time.

- **Ground Plane**
  We also used ground plane from vuforia engine which depends on detection of horizontal or vertical plane to allocate our avatar on the detected plane

Figure(74):Avatar Allocation Using Ground Plane

Limitations of Ground Plane

- It doesn’t support all mobile devices
- Detection of the planes wasn’t stable

2) AR Foundation and ARCore

- AR Foundation allows us to work with augmented reality platforms in a multi-platform way within Unity. This package presents an interface for Unity developers to use, but doesn't implement any AR features itself. To use AR Foundation on a target device, we need separate packages for the target platforms officially supported by Unity:

  - ARCore XR Plugin on Android
  - ARKit XR Plugin on iOS

  AR Foundation is a set of MonoBehaviours and APIs for dealing with devices that support many concepts such as device tracking , plane detection , 2D and 3D image tracking...etc.

- To Use AR Foundation in unity we import it from the Package Manager in unity.
- **ARCore XR Plugin**
  - ARCore XR Plugin package is used to enable ARCore support via Unity's multi-platform XR API.
  - To Use ARCore in unity we import it from the Package Manager in unity.
  - After we imported these packages , now we are able to use the AR Camera and use the augmented reality service.

Figure(75):Avatar Allocation Using ARCore

3) Speech to text

- Speech to text uses speech recognition technology to identify patterns in sound waves and match them to the phonemes of speech to translate them into text.

- **Why Speech to Text ?**
    - Ease of communication.
    - Time saved with increased efficiency: can transcribe even lengthy passages of text in minutes or even seconds.
    - Speech to text is cheap and available.
- **How speech to text is used in Bi Translator?**
    - We used **IBM Watson Speech to Text API** to achieve this feature in words translation technique.
    - An **API** is a set of programming code that enables data transmission between one software product and another. It also contains the terms of this data exchange.
    - We used IBM Watson speech to text API by integrating its API and URL to our unity Project then we relate our animation clips to it.

Chapter Nine

9. Results and Discussion

9.1 ASL to text | Speech Results

9.1.1 CNN first approach result

All static alphabets are translated by this approach

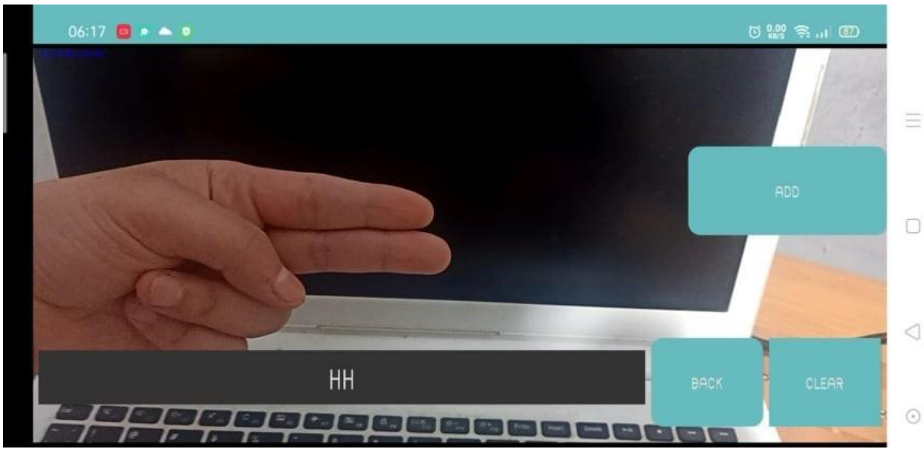


Figure (76)Translation of “H” sign

Figure (77)Translation of “L” sign

9.1.1 Faster-RCNN Second approach result

All alphabets are translated by this approach and some words like "hello" , "hi " and "Five" .

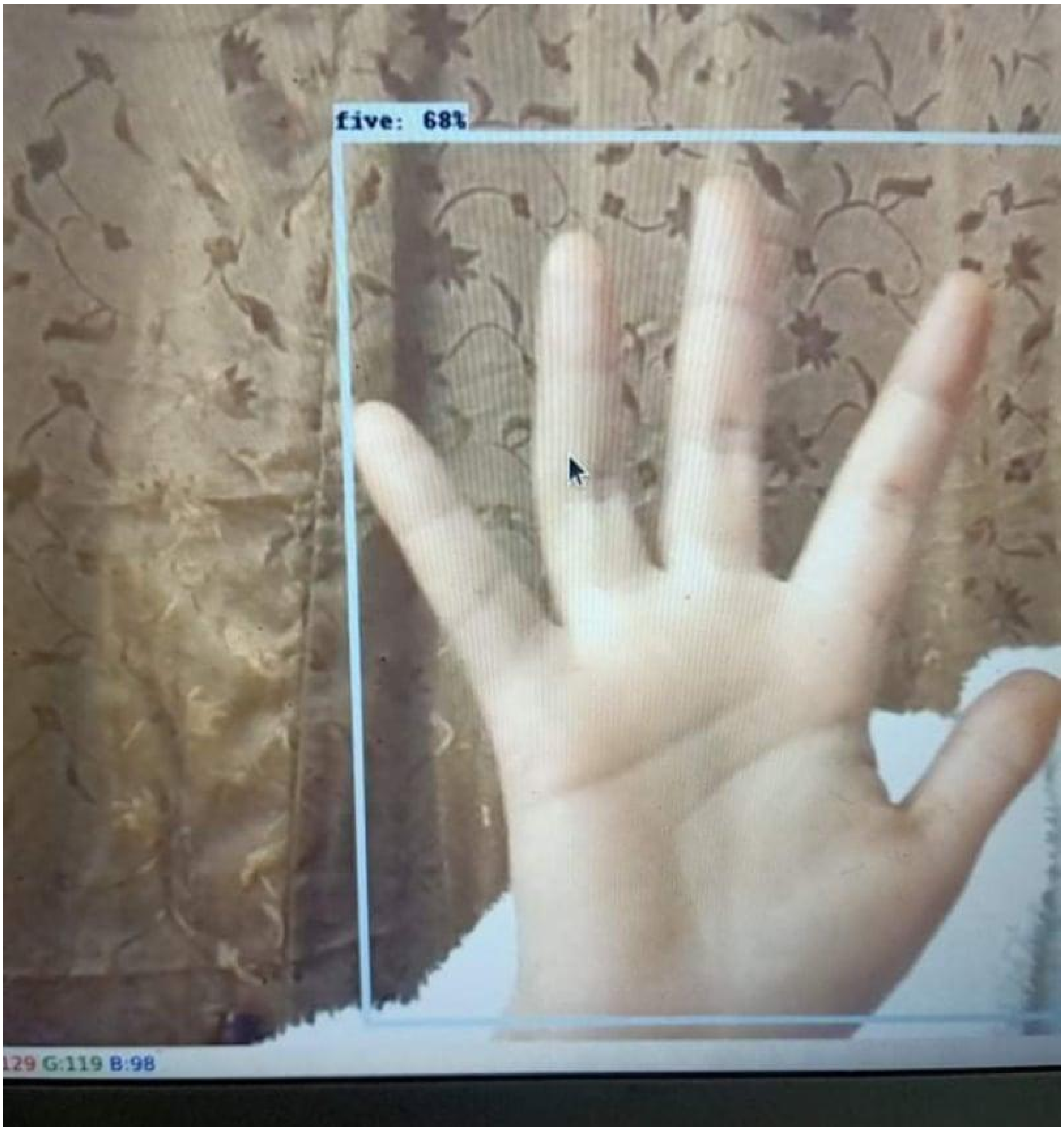


Figure (78)Translation of "Five" sign

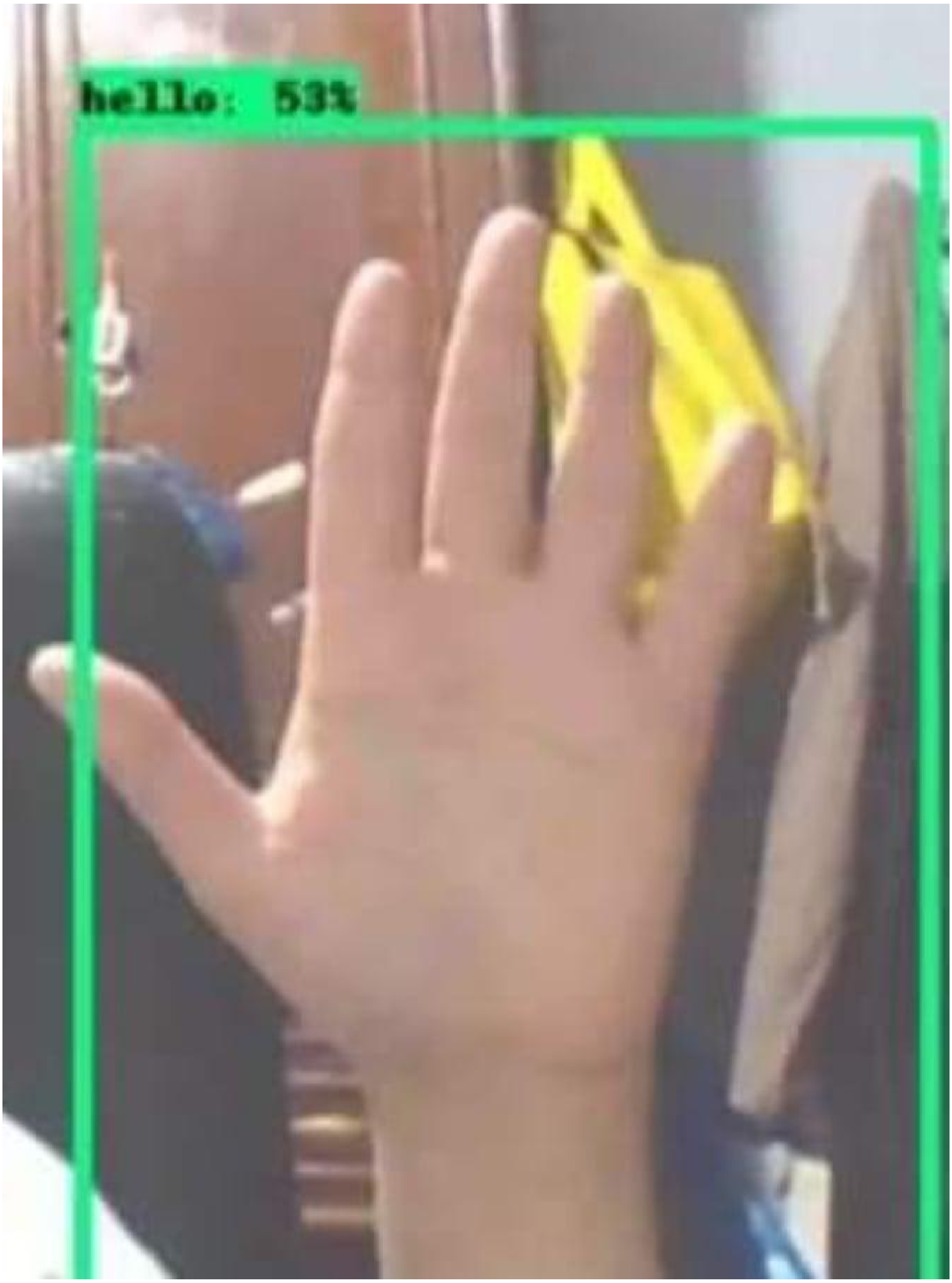


Figure (79)Translation of “Hello ” sign

9.1.1 Media-Pipe sklearn approach result

All alphabets are translated by this approach and some words like "hello" , "I love you " and "I " .

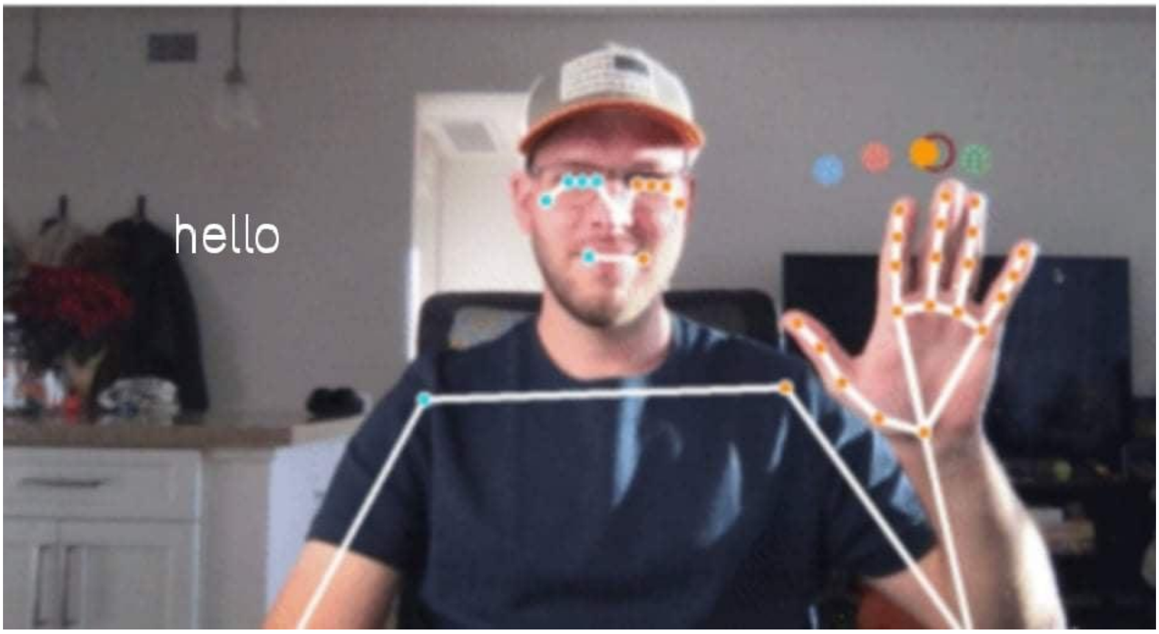
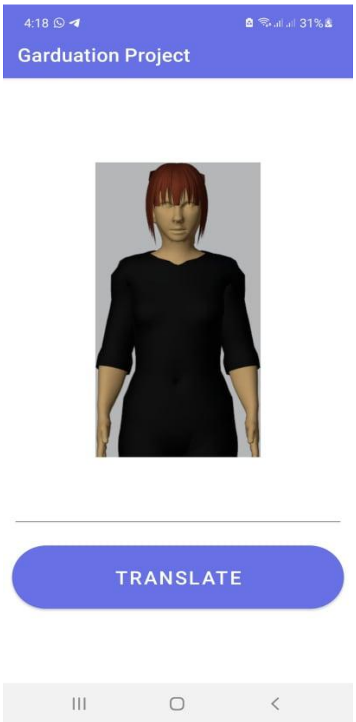


Figure (80)Translation of "Hello " sign

9.2 Speech | text to ASL Results

9.2.1 Letter Translation Results:

- When clicking on “Letter Translation” Button , this activity will be shown and it shows the avatar in its reset pose.



Figure(81): Reset Pose

- Translating the word “Hi”.

Figure(82): ‘H’ Sign **Figure(83): ‘I’ Sign**

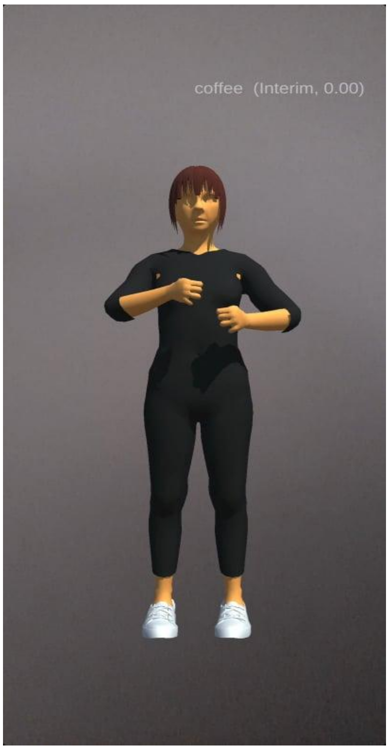
- After finishing translating the avatar will return back to its reset pose.

9.2.2 Words Translation Results:

We will show the results of the four words (hello , I, need ,coffee)

Figure(84): 'I' Sign

Figure(85): 'Hello' Sign



Figure(87): 'Coffee' Sign

Figure(86): 'Need' Sign

9.3 The used Tools

<table>
<tr><td>Anaconda</td><td>Definition :
Anaconda is a distribution of the Python and R programming languages for scientific computing that aims to simplify package management and deployment.
Usage:
An open-source package and environment management system called conda , which makes it easy to
install/update packages and
create/load environments.
support multiple useful IDEs like Spyder and jupyter .
Support machine/Deep learning libraries like tensorflow , sklearn and keras .
Version :
Our used version is 1.10.0.</td></tr>
<tr><td>Spyder</td><td>Definition :
Spyder is a free and open source scientific environment written in Python.
It is designed by scientists and is exclusively for scientists, data analysts, and engineers.
Features :
Customize syntax highlighting.
Availability of breakpoints (debugging and conditional breakpoints)
interactive extension which allows you to run lines , files and cells .
It provides a real time code introspection.
Usage :
It is used in training and testing of deep learning models .
Version :
We used the last version which is version 4 .</td></tr>
</table>

Blender	**Definition :** is a free and open-source 3D creation suite that supports pretty much every aspect of 3D development. **Features :** ● 3D Modeling . ● texturing . ● raster graphic editing . ● rigging and skinning. ● fluide and smoke simulation. ● practical simulation . ● soft body simulation . ● rendering . ● motion graphics. **Usage :** ● It is used in building the animated character. **Version :** ● We used blender's latest version 2.9.
Krita	**Definition :** Krita is a professional FREE and open source painting program. **Features :** ● Krita provides an OpenGL based canvas in addition to an unaccelerated canvas. ● Krita’s filters, histogram computation and image recomposion are multi-threaded and make use of multiple cores if available. - The effect of filters is previewed on-canvas **Usage :** ● It is used in building Gif files . **Version** ● We used Krita's latest version 4.4.3.

Unity	**Definition :** Unity is a game development ecosystem of powerful rendering engine, intuitive tools, rapid workflows for 2D and 3D games. **Features :** ● Easy workflow . ● Dedicated tools for both 2D , 3D game creation. ● a very unique and flexible animation system to create animation in very less time . ● Smooth frame rate with reliable Performance on all the platforms . ● reduce the time development by using already created reusable assets available on the huge asset store . **Usage :** ● It is used in Augmented reality. **Version :** ● Unity v.4
Android Studio	**Definition :** Android studio is the official integrated development environment (IDE) for Google's android operating system. **Features :** ● Instant App Run. ● Visual Layout Editor. ● Intelligence Code Editor ● Addition of New Activity as a Code Template. ● Help to Build Up App for All devices. - Help to Connect with Firebase **Usage** : ● It is used as an interface for translating. **Version :** ● We used the latest version 4.2

9.4 Environment

1- Ubuntu 20.4 OS .

2- NVIDIA GeForce 940MX

3- 12 GB RAM

4- Core i7-7500U 7th .

9.5 Constraints

- Augmented Reality required android Api 30 or above.
- Tensorflow api required android Api 27 or above .

Chapter Ten

10. Conclusion and Future Work

10.1 Conclusion

Finally , as we said before BI Translator System facilitates communication between normal and deaf people. This system acts as a mediator between them by using a mobile application which provides two phases:

1. The first phase (ASL to Text) which is achieved by Deep Learning approaches.
2. The second phase (Speech | Text to ASL) which is achieved by augmented reality (AR).

We will ask some questions and try to answer them to explain our choices in the system,firstly why our system is a mobile based application? Because it is available for everyone ,easy to be used and applicable in the real time translation.

Secondly , why did we use Augmented Reality ?Because AR allows us flexibility,AR devices don't always require high performing hardware and the world is constantly changing so we need something flexible that provides solutions to everyday problems that are becoming ever more complex.

10.2 Future Work

- Apply more dataset on translation ASL to Text.
- Integrate final approach of translating ASL to Text in mobile application.
- Integrate text to speech api in translation phase 1(ASL to text) .
- Adding more words in the words translation section.
- Apply dynamic sign recognition .

- Applying multiple languages to translate different sign languages.
- Making the avatar more interactive.
- Adding speech to text in letter translation.
- Adding text to speech in sign to text translation.

Chapter Eleven

11. References:

1. Kelly Mercer, ‘Deaf History Month: Assistive Technology’, April 8, 2015[Online].Available:https://101mobility.com/blog/deaf-history-month-assistive-technology/

2. WHO ,’ Deafness and hearing loss’, 1 April 2021 [Online].Available: https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss

3. WHO,’Deafness and hearing loss Rate’,1 April 2021 [Online].Available:https://www.who.int/pbd/deafness/estimates/en/

4. IBM Cloud Education ,’Machine Learning ’ , 15 July 2020[Online] . Available: https://www.ibm.com/cloud/learn/machine-learning

5. Developed by JavaTpoint, ‘Machine learning Life cycle’, 2011-2021[Online].Available:https://www.javatpoint.com/machine-learning-life-cycle?fbclid=IwAR2QTGOYMprZxZY9-AiGxr4gt2YXWIqWUTG1UgdY7s7s-Lvs_94PuS38YFQ

6. 7wData,’Types of Machine Learning Algorithms’,June 5, 2020[Online].Available:https://7wdata.be/visualization/types-of-machine-learning-algorithms-2/

7. Jason Brownlee ,’4 Types of Classification Tasks in Machine Learning’, August 19, 2020[Online].Available: https://machinelearningmastery.com/types-of-classification-in-machine-learning/

8. Developed by JavaTpoint,’Unsupervised Machine Learning’, 2011-2021,[Online].Available:https://www.javatpoint.com/unsupervise

d-machine-learning?fbclid=IwAR2CqCwMWpdpa8HeAAyukqH2PX4nhMPchzU31H-tegNIEjSgp5IHcyZEj_U

9. Surya Priy,'Different Types of Clustering Algorithm',08 Jun, 2021 [Online].Available:https://www.geeksforgeeks.org/different-types-clustering-algorithm/

10. Wikipedia,'Reinforcement learning ',13 July 2021[Online].Available: https://en.wikipedia.org/wiki/Reinforcement_learning

11. JavaTpoint,'Association Rule Learning',2011-2021 [Online].Available: https://www.javatpoint.com/association-rule-learning

12. MARSHALL HARGRAVE,'Deep Learning',May 17, 2021[Online]. Available:https://www.investopedia.com/terms/d/deep-learning.asp

13. FAIZAN SHAIKH,'Deep Learning vs. Machine Learning',April 8, 2017[Online].Available:https://www.analyticsvidhya.com/blog/2017/04/comparison-between-deep-learning-machine-learning/?fbclid=IwAR2vFG4GBuduSRd7IlVl81AWC3l6xvvIHv26mWeMjCy_wwlFsgGNovOKkbk

14. DataFlair,'How Deep Learning Works with Different Neuron Layers ',2021[Online].Available:https://data-flair.training/blogs/how-deep-learning-works/

15. JAKE FRANKENFIELD,'Artificial Neural Network (ANN)',Aug 28, 2020[Online].Available:https://www.investopedia.com/terms/a/artificial-neural-networks-ann.asp

16. Aston Zhang,Zack C. Lipton,Mu Li,Alex J. Smola,'Dive into Deep Learning',May2019[Online].Available:https://d2l.ai/chapter_multilayer-perceptrons/backprop.html

17. Hamza Mahmood,'Activation Functions in Neural Networks',Dec 31,2018[Online].Available:https://towardsdatascience.com/activation-functions-in-neural-networks-83ff7f46a6bd

18. Sumit Saha,'A Comprehensive Guide to Convolutional Neural Networks',Dec15,2018[Online].Available:https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53

19. Prabhu,'Understanding of Convolutional Neural Network (CNN) — Deep Learning',Mar 4,2018[Online].Available: https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148

20. Jiwon Jeong,'The Most Intuitive and Easiest Guide for Convolutional Neural Network',Jan 24,2019[Online].Available: https://towardsdatascience.com/the-most-intuitive-and-easiest-guide-for-convolutional-neural-network-3607be47480

21. Lilian Weng,'Object Detection for Dummies Part 3: R-CNN Family ',Dec 31, 2017[Online].Available: https://lilianweng.github.io/lil-log/2017/12/31/object-recognition-for-dummies-part-3.html

22. The MathWorks, Inc,'Getting Started with R-CNN, Fast R-CNN, and Faster R-CNN',1994-2021[Online].Available: https://www.mathworks.com/help/vision/ug/getting-started-with-r-cnn-fast-r-cnn-and-faster-r-cnn.html

23. Aston Zhang,Zack C. Lipton,Mu Li,Alex J. Smola,'Dive into Deep Learning',May2019[Online].Available:https://d2l.ai/chapter_computer-vision/rcnn.html?fbclid=IwAR1Vm-p_5XmcPel7j2Tpt6KRluOlTKilLPsHs-dK5olTi5GazYTRyefpvCQ

24. Rohith Gandhi,’R-CNN, Fast R-CNN, Faster R-CNN, YOLO — Object Detection Algorithms’,Jul 9 ,2018[Online].Available: https://towardsdatascience.com/r-cnn-fast-r-cnn-faster-r-cnn-yolo-object-detection-algorithms-36d53571365e
25. Udemy,’Introduction to Animation’,April,2017[Online].Available: https://www.udemy.com/course/introduction-to-animation/
26. ALYSSA MAIO,’What is Animation? Definition and Types of Animation’,NOVEMBER 18, 2020[Online].Available: https://www.studiobinder.com/blog/what-is-animation-definition/
27. Adrien-Luc Sanders,’Introduction to Vector Animation’,December 21, 2020[Online].Available:https://www.lifewire.com/introduction-to-vector-animation-140890
28. Justin Beegel ,’What Is 3D Animation ?’,2020[Online].Available: https://infographicworld.com/what-is-3d-animation/
29. Renderforest Staff,’What is 3D Animation? A Beginner’s Guide’,10 May,2021[Online].Available:https://www.renderforest.com/blog/3d-animation
30. Aldric Chang,’The Process of 3D Animation’,[Online].Available: https://www.media-freaks.com/the-process-of-3d-animation/
31. stada media,’What Are The Uses Of 3D Animation?’,2021[Online]. Available:https://stadamedia.co.uk/what-are-the-uses-of-3d-animation/
32. ADAM HAYES,’Augmented Reality’,Dec 2, 2020[Online].Available: https://www.investopedia.com/terms/a/augmented-reality.asp
33. Dave Johnson,’What is augmented reality? Here's what you need to know about the 3D technology’,Dec 5, 2020[Online].Available: https://www.businessinsider.com/what-is-augmented-reality

34. CORY MITCHELL,’Virtual Reality’,Sep 29, 2020[Online].Available: https://www.investopedia.com/terms/v/virtual-reality.asp

35. Ivy Wigmore,’mixed reality (hybrid reality, extended reality)’,1999 - 2021[Online].Available: https://whatis.techtarget.com/definition/mixed-reality

36. Interaction Design Foundation ,’Augmented Reality’,2020[Online] .Available:https://www.interaction-design.org/literature/topics/augmented-reality

37. Sidharth Jain,’Benefits of Augmented and Virtual Reality for Constructing Top Mobile App’,March7,2021[Online].Available: https://www.europeanbusinessreview.com/benefits-of-augmented-and-virtual-reality-for-constructing-top-mobile-app/

38. Designing Digitally,’WHAT IS AUGMENTED REALITY AND HOW DOES IT BENEFIT EMPLOYEES?’,Jun 17,2021[Online].Available: https://www.designingdigitally.com/blog/2021/06/what-augmented-reality-and-how-does-it-benefit-employees

39. Nadia Kovach ,’What is Augmented Reality (AR) and How does it work’,*June 8, 2017*[Online].Available: https://thinkmobiles.com/blog/what-is-augmented-reality/

40. Vuforia,’Getting Started with Vuforia Engine in Unity ’,2011-2021 [Online].Available:https://library.vuforia.com/articles/Training/getting-started-with-vuforia-in-unity.html#about

41. Vuforia,’Vuforia Engine Overview’,2011-2021 [Online].Available:https://library.vuforia.com/features/overview.html

42. Google Developers,’ARCore overview’,May 27,2021[Online] .Available:https://developers.google.com/ar/discover

43.IBM Cloud Education,’IBM Cloud services’,[Online]. Available:https://www.ibm.com/watson/in-en/products-services/

44.IBM Cloud Education,’IBM Cloud’,[Online]. Available:https://www.ibm.com/cloud

45.Sai Vennam,’Cloud Computing’, August 18 ,2020[Online]. Available:https://www.ibm.com/cloud/learn/cloud-computing

46.IBM Cloud Education,’Watson Speech to Text’,june 6 2021[Online].Available:https://cloud.ibm.com/apidocs/speech-to-text-data?code=unity

47.IBM Cloud Education,’Watson Speech to Text’,July 2020[Online]. Available:https://www.ibm.com/cloud/watson-speech-to-text/features

48.J Simpson,’Speech-to-Text APIs’,March 28, 2019[Online]. Available:https://nordicapis.com/5-best-speech-to-text-apis/

49.NIHdeveloper,’Cochlear Implants’,February 2016[Online]. Available:https://www.nidcd.nih.gov/health/cochlear-implants?fbclid=IwAR23yyK0ahUqKaQtoc2rSVideMkfDR7oErKpCaLGCloyhXjGqD09LNoP9qw

50. Kirsten Nunez,’Cochlear Implant’,February 27, 2020[Online].Available:https://www.healthline.com/health/cochlear-implant#surgical-procedure

51.Hearing Systems developer,’The History of Hearing Aids’,February 19,2019[Online].Available:https://hearingsystemsinc.com/the-history-of-hearing-aids/

52.WebMD developer,’Hearing Aid Basics’, 2005 - 2021[Online]. Available:https://www.webmd.com/healthy-aging/hearing-aids#1

53. FAD developers,’Benefits of hearing aids’,January16, 2018[Online]. Available:https://www.fda.gov/medical-devices/hearing-aids/benefits-and-safety-issues

54. Debbie Clason,’Hearing aids’,February 15, 2018[Online]. Available:https://www.healthyhearing.com/report/52837-The-pros-and-cons-of-small-hearing-aids

55. University of Washington,’SignALoud’,April 12, 2016[Online].Available:https://www.washington.edu/news/2016/04/12/uw-undergraduate-team-wins-10000-lemelson-mit-student-prize-for-gloves-that-translate-sign-language/

56. Matthew Chin,’Wearable-tech glove’,June 29, 2020[Online]. Available:Wearable-tech glove translates sign language into speech in real time

57. Back4App developer,’2020’,[Online]. Available: https://blog.back4app.com/firebase-storage/

58. Cloud flare,’Cloud definition’,2021[Online]. Available: https://www.cloudflare.com/learning/cloud/what-is-the-cloud/

59. PIXX0,’Female Character’,January 07, 2020[Online]. Available:https://blendswap.com/blend/24118

60. Melanie Griffin,’Blender: Character Modeling’,Feb 11, 2020[Online]. Available:https://all3dp.com/2/blender-2-8-character-modeling-simply-explained/

61. AlexLemminG,’RigifyToUnity’, 2021[Online].Available: https://github.com/AlexLemminG/Rigify-To-Unity

62. Unity developer,Animator Controller,2021-07-16[Online]. Available:https://docs.unity3d.com/Manual/class-AnimatorController.html

63. ASL University ,
https://www.lifeprint.com/asl101/pages-layout/facialexpressions.htm

64. Antonio Dom `enech L.,ASL Recognition with MediaPipe and Recurrent Neural Networks, 28. July 2020,[Online] Available: https://upcommons.upc.edu/bitstream/handle/2117/343984/ASL%20recognition%20in%20real%20time%20with%20RNN%20-%20Antonio%20Dom%C3%A8nech.pdf?sequence=1&isAllowed=y.