Rajiv Gandhi Institute of Technology
Chola Nagar, Bengaluru, Karnataka - 560032

Sign Language Interpreter Using Deep Learning

Under the guidance of : Pushpalatha Dubey

- Bahaduri Prachiti Jagdish - 1RG17CS009
- Mohammed Azim - 1RG17CS030
- Mohammed Touseef - 1RG17CS031
- Prajwal B Mani - 1RG17CS037

ABSTRACT

Speech Impairment is a disability, which affects an individual's ability to communicate using speech and hearing. This brings about the difficulty for both the sign and non - sign language speakers to communicate with each other. With recent advances in deep learning and computer vision, the focus of our project is to create a vision of an end to end Convolutional Neural Network that will be trained on the ASL(American Sign Language) dataset then modeled on robust architectures like GoogLeNet/MobileNet architecture and deploy it on an android application so that it will have more accessibility and provides an ease of use, thus aiding communication between signers and non-signers. It is a challenging and interesting problem that
if solved will bring a leap in social and technological aspects alike.

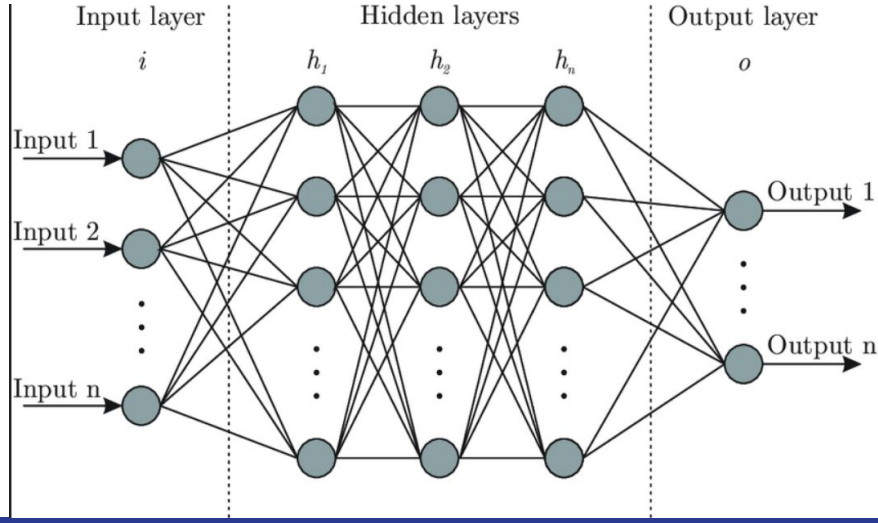
INTRODUCTION

What do you mean by Neural Networks?

COMPONENTS OF A NEURAL NETWORK :

- INPUT LAYER
- HIDDEN LAYER
- OUTPUT LAYER

- A neural network is a series of algorithms that endeavors to recognize underlying relationships in a set of data.
- A neural net consists of thousands or even millions of simple processing nodes that are densely interconnected.



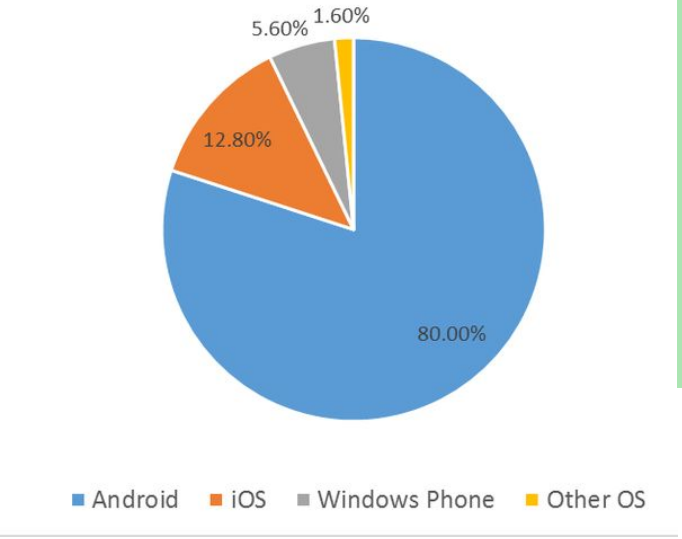
APPLICATIONS :

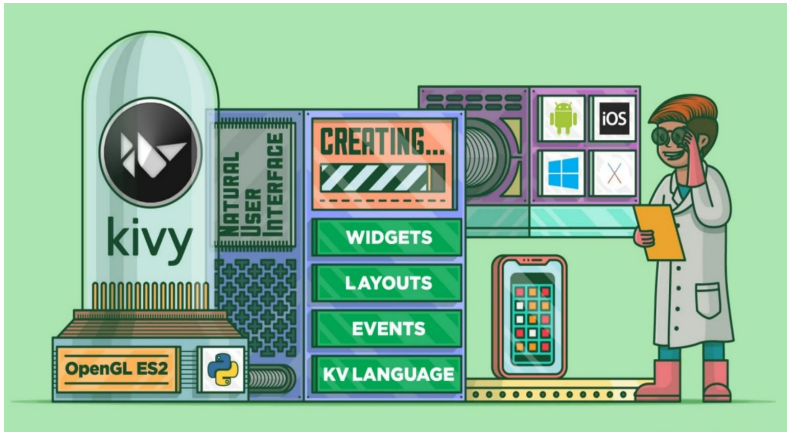
TIME SERIES PREDICTION

HUMAN FACE RECOGNITION

Android

Android is a mobile operating system based on a modified version of the Linux kernel and other open source software, designed primarily for touchscreen mobile devices such as smartphones and tablets

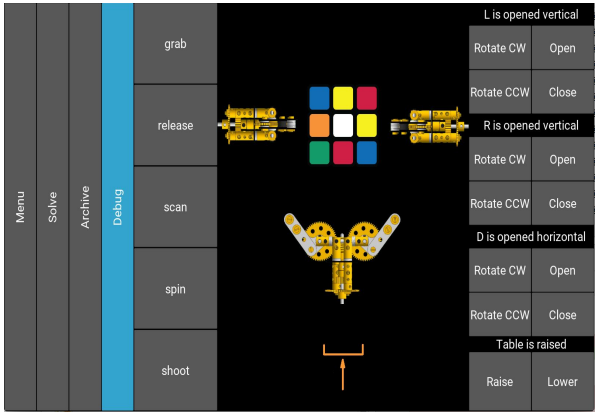


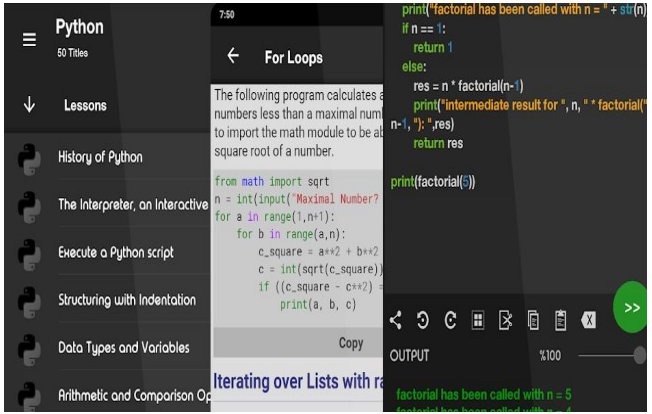


Kivy : *Cross-platform Python Framework for NUI Development*

- Open source Python library for rapid development of applications that make use of innovative user interfaces, such as multi-touch apps.
- **Cross platform** - Kivy runs on Linux, Windows, OS X, Android, iOS, and Raspberry Pi. You can run the same code on all supported platforms.
- **Business Friendly** - Kivy is 100% free to use, under an MIT license (starting from 1.7.2) and LGPL 3 for the previous versions. The toolkit is professionally developed, backed and used. You can use it in a commercial product
- **The framework is stable and has a well documented API, plus a programming guide to help you get started.**

Project examples in Kivy:-





- Meccano Rubik's Shrine
- Learn Python Offline & IDE
- Morse Code Jam
- Chisel

REQUIREMENTS

- Front End : Android (using Kivy)
- Video & image processing : Open CV, scikit-image
- Neural Network Modeling:Tensorflow(tf.lite), Keras, Pytorch
- Architectures : Google LeNet, MobileNet, VGG 19

OpenCV

Keras

LITERATURE SURVEY

Sl no.	Title and Author name	Abstract	Techniques and algorithms used	Drawbacks
1.	Real-time American Sign Language Recognition with Convolutional Neural Networks *Brandon Garcia & Sigberto Alarcon Viesca (Stanford University Stanford, CA)*	They hereby present the development and implementation of an American Sign Language (ASL) fingerspelling translator based on a convolutional neural network. They utilize a pre-trained GoogLeNet architecture trained on the ILSVRC2012 dataset.	a deep learning framework, in order to develop, test, and run our CNNs. Specifically, they used Berkeley Vision and Learning Center's GoogLeNet pre-trained on the 2012 ILSVRC dataset.	space and time constraints initially required them to choose a less-than-optimal batch size value of 4, resulting in the noisy loss.
2.	Deaf Talk using 3D Animated Sign Language Mateen Ahmed, Mujtaba Idrees, Zain ul Abideen, Rafia Mumtaz, Sana Khalique	The proposed system is dubbed as Deaf Talk, and it acts as a sign language interpreter and translator to provide a dual mode of communication between sign language speakers and natural language speakers.	With Microsoft's Kinect in the market, especially for hands,used the depth sensor of Kinect to recognize around thousand phrases from ASL and this recognition is based on hidden Markov model (HMM),also focused on Kinect base hand gestures recognition.	Depth image sensing is also used for recognizing hand gestures but it remained limited to simple hands gestures.

Sl. no.	Title and Author name	Abstract	Techniques and algorithms used	Drawbacks
3.	American Sign Language Recognition using Deep Learning and Computer Vision *Kshitij Bantupalli & Ying Xie*	The fields of motion and gesture recognition using deep learning and computer vision-based techniques. The focus of this work is to create a vision-based application which offers sign language translation to text thus aiding communication between signers and non-signers.	use Inception net, a CNN for recognizing spatial features. then use a RNN to train on temporal features. The dataset used is the American Sign Language Dataset.	The model also suffered from loss of accuracy with the inclusion of faces.
4.	NumPy CNN Android : A Library for Straightforward Implementation of Convolutional Neural Networks for Android Devices *Ahmed Fawzy Gad*	A new open source library called NumPy CNN Android is proposed that minimizes the overhead of building and running convolutional neural networks on Android devices.	AlexNet , VGGNet and GoogLeNet are examples of CNN architectures trained with the ImageNet dataset.	The weakness of the proposed library is its computational time.

Sl. no.	Title and Author name	Abstract	Techniques and algorithms used	Drawbacks
5.	A Deep Learning Based Video Classification System Using Multimodality Correlation Approach *Jungheon Lee, Youngsan Koh and Jihoon Yang*	In this paper, we show that the way of using image and audio modality of video is better than the usual way, and introduce a method of using multimodality efficiently by utilizing correlation.	image feature vector extraction Audio feature vector extraction Normalization of feature vectors	Even though it works better than other video extraction models still the computational speed is more than the image extraction model which have high accuracy
6.	Hand Gesture Recognition Using Deep Learning *Soeb Hussain and Rupal Saxena*	This paper aims at the automatic interpretation of gestures based on computer vision. And proposes a technique which commands computer using six static and eight dynamic hand gestures.	Transfer Learning : VGG16 architecture	Avoiding skin color segmentation, blob detection, skin area cropping and centroid extraction for unidirectional dynamic gestures.

Sl no.	Title and Author name	Abstract	Techniques and algorithms used	Drawbacks
7.	Deep Learning for American Sign Language Fingerspelling Recognition System *Huy B.D Nguyen, Hung Ngoc Do*	A sign language fingerspelling alphabet identification system would be developed by using image processing technique, supervised machine learning and deep learning.	Histograms of Oriented Gradients(HOG) and Local Binary Pattern(LBP), Moore Neighbourhood algorithm, SIFT algorithm	the validation accuracy of the CNN-SVM model is lower than that of HOG-LBP-SVM model, it has a better chance to counter overfitting.
8.	American Sign Language Video Hand Gestures Recognition using Deep Neural Networks *Shivashankara S & Srinath S*	In this paper an effort has been placed to translate / recognize some of the video based hand gestures of American Sign Language (ASL) into human and / or machine readable English text using deep neural networks.	Speeded Up Robust Features(SURF), Zernike Moment(ZM), Discrete Cosine Transform(DCF), Radon Features	Due to the gestures captured in low illumination night time, there is a bit of loss of recognition rate.

Sl no.	Title and Author name	Abstract	Techniques and algorithms used	Drawbacks
9.	American Sign Language Character Recognition using Convolutional Neural Network *Sarfaraz Masood, Manish Chandra Thuwal, Adhyan Srivastava*	This work aims at taking the basic step in bridging the communication gap between normal people and deaf and dumb people using sign language.	Convolutional Neural Network Model, Image Preprocessing and VGG16 Model	The images in the data set were of a varying size and shape.The first step was to read and resize each of the image to the similar size of 224x224 pixel. Only when all of the images in the dataset are of the same size can the images be fed into a neural network for training.
10.	Optimization of Transfer Learning for Sign Language Recognition Targeting Mobile Platform *Dhruv Rathi*	The target of this research is to experiment, iterate and recommend a system that is successful in recognition of American Sign Language(ASL).	Convolutional Neural Network(CNN) on ImageNet,two pre-trained models, Inception V3 model and MobileNets open-source models for comparison purposes.	The two modes tend to learn different features when the same dataset is feed to the model this gives us different accuracies

PROBLEM STATEMENT

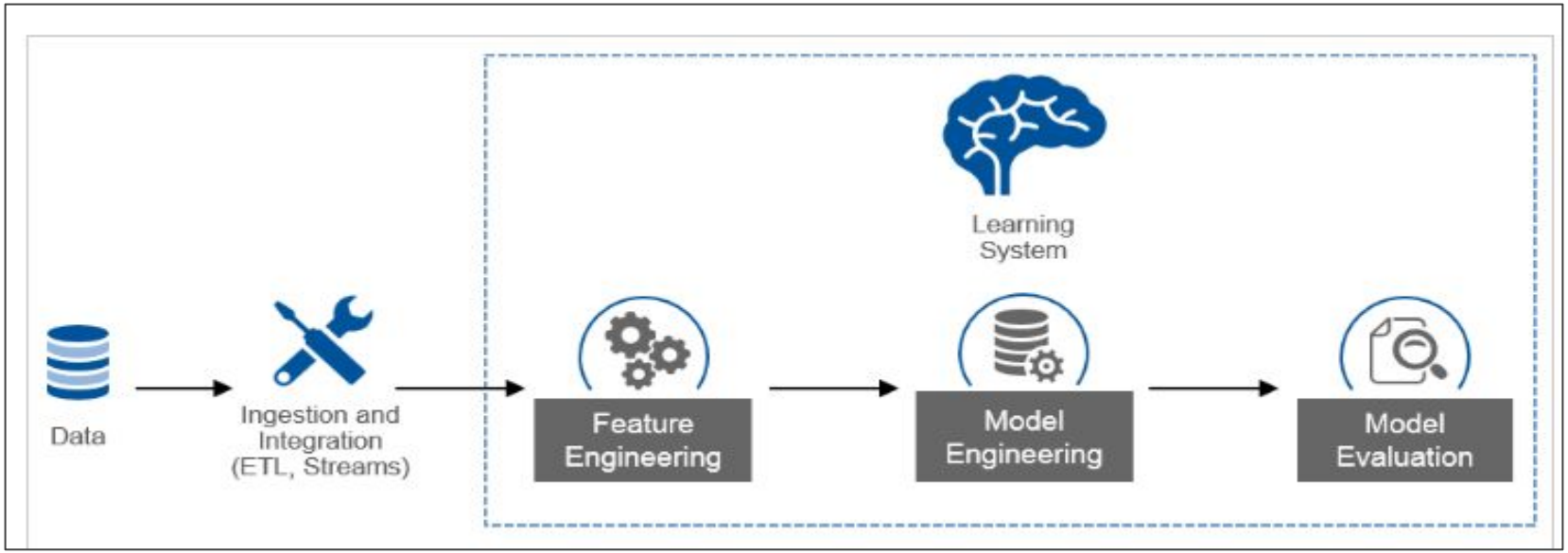
- Sign Language is a virtual gestural language (words expressed in air) which is used to communicate with people having impairments and speech
- From the systematic analysis, difficulty is faced by deaf community towards access to health care leading to communication barrier.
- According to the projection of data it is expected to rise to 900 million by 2050 that is double the current stats.
- Without prior knowledge, it is difficult for the non - signers to understand the signs and communicate effectively.
- The Typical cost of interpreter is 100 dollars/hrs
- And Microsoft kit costs around 4000 dollars for a pair

OBJECTIVES

1. Create a robust android application that aids in communication for deaf people.
2. Use deep learning with transfer learning techniques to build a neural network model that as to ability to learn pattern in video and classify the images.
3. An error correction model for identifying the pattern mismatch and correcting it for the audio or text input format.
4. To build a model that has higher accuracy with less bias and overfitting problems.
5. To enhance the model to classify at least 20 classes or more .

METHODOLOGY

Development life cycle of MLT Projects



DataSet : **American Sign Language video dataset**

ETL(Extract, transform, load):Web Scraping(bs4),OS Module

Feature engineering : Feature engineering is the process of using domain knowledge to extract features from raw data via data mining techniques

Algorithms for feature selection :

- ORB (Oriented FAST and Rotated BRIEF)
- Color Gradient Histogram
- Vantage Point Tree
- KAZE

CONCLUSION

A sign language is a natural mode of communication used by the deaf community. India has large population of speech and hearing impaired but a very small number of certified sign language interpreters are available. Research in hand gesture recognition has gained attention with advancement in the field of computer vision. A Sign Language Interpreter (SLI) decodes and understands the information conveyed by signs. SLI can be a major breakthrough in helping a common people to communicate with the deaf and can help in bridging this communication gap. A SLI can be designed based on video/image processing and deep learning techniques which requires a standard dataset, determination of an optimal feature set, and an appropriate classification technique.


Many thanks.