Educating autistic children with human emotions and sign language using deep learning

Section A – Ethics Application

	I submitted my ethics application and my application has been approved. I include my ethics certificate in the appendix as evidence
✓	I submitted my ethics application and my application is currently under review.
	I have not submitted my ethics application.

Section B – Project Proposal

- **Problem Statement**

Autism spectrum disorder (ASD) is a neurological sickness characterized by difficulty in communication and social interaction. According to WHO, about one in 160 children are suffering from ASD, and the degree of symptoms shown by each child varies. Hence this disease is called as a spectrum disorder. Despite having various symptoms, the main hindrance to their communication is their inability to identify human emotions and difficulty in handling spoken language. so, by this research, I am going to address these two problems. Michele Ricamato, Speech and Language Pathologist from West Chicago, suggests that sign language is a perfect option for augmenting communication of autistic children. So by this research, the goal is to enhance autistic children's communication by educating them with human emotions like surprise, angry, disgust, fear, happy, neutral, sad and sign language both visually and vocally by building a real-time prototype system using artificial neural network. This paper focuses on elementary-aged autistic people from age 5 to 13 because it is the best period to introduce new mode of communication.

- **Intended users and their requirements**

As mentioned earlier, since autism is a spectrum disorder, based on the degree of behavioral patterns, it is classified into 3 levels, Level1, Level2, and Level3. This research project aims at Level2 category of autistic children with no motor deficits because they are less sensitive than Level 3 but require substantial support than Level1 for speech and social skill therapy. Moreover, these children may or may not communicate verbally and showcase more difficulty in social skills. Hence these are the perfect category to educate emotions and sign language.

Moreover, this system can act as a platform to support autistic children and their families when they are needed even though the pandemic restricts access to certain facilities and certain people. Hence this system will be beneficial to people involved in autistic children's development such as therapists and tutors from special school academies as well. In this way, the special academic institutions can still be able to continue their educational programs even if pandemic forces them to shut down the schools and public gatherings. Besides, in the future government can fund this project by developing it on a large scale. Consequently, this will help to introduce new educational strategies and improve economic status by encouraging companies for developing valuable systems like this.

As per the Center for Autism Research in The Children's Hospital of Philadelphia, various forms of Augmentative and Alternative Communication (AAC) such as the Picture Exchange Communication System (PECS) and sign languages like American Sign Language (ASL) have a greater impact on autistic people for enhancing communication. But due to the need for special equipment like picture cards in PECS, sign language is more feasible. Hence in this project, the researcher aim at educating ASL

alphabets and digits signs. Moreover, human emotions are educated as a part of pre-language skills so that they can understand others' feelings while having a conversation and can respond accordingly.

Furthermore, due to the social isolation of the Covid 19 pandemic and the lack of needed therapy access, the chances of increased behavioral problems in autistic children are quite higher. Therefore, the need of a real-time system that uses the visual(emoji) and vocal mode (voice recording) of teaching is very essential and this type of teaching can attract their attention.

- **Systems requirements, project deliverables and final project outcome**

By this experiment, I am going to build a real-time prototype system that can translate human emotions, ASL alphabets, and ASL digits to corresponding emoji pictures and voice recordings. In addition to that, the system will also incorporate some functionalities like recording, screen capture, and test option to check if the child can able to select the correct emotion or signs.

Due to the behavioral problems of our intended users, this system is going to use by them along with their parents or primary carers. Also, the teaching part will be mostly carried out with special school teachers or speech therapists and by this experiment, I will be developing only the interface that is visible to the intended user. A sample of the final interface is shown below:

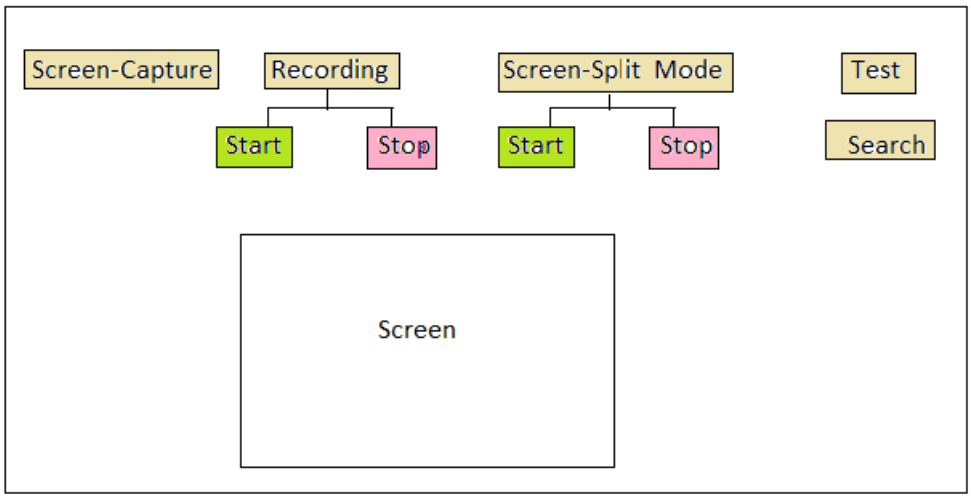

Fig1: Sample Output Screen

- **Screen Capture and Recording Option:**

Unlike normal children, autistic child takes too much time to learn even a single sign or emotion, so it is not feasible for the teaching staff to educate them all the time. These two options solved those problems so that later the recorded video (along with audio) or image can be used by parents or carers to educate the child.

- **Start/Stop Screen-Split Mode:**

The system is going to work in two modes, Normal video call mode, and screen-split mode. Initially while running the system, it will be in the video call model. This is to make the child calm and comfortable before teaching. At that time there will be only one screen showing the tutor's video. On screen-split mode, the screen will be split into two, in one the tutor will show signs or emotions, and on the second one corresponding emoji pic will be shown. Along with that, a voice recording will also be played.

- **Test:**

By this, the tutor can check whether the children understood what he/she taught. Once this is on, a screen will appear that shows emotion or sign, and 2-3 checkboxes below that, the child needs to identify the correct one and check it with the help of parents or carers. After selecting, a message will be shown for correct as well as for wrong answers as a matter of encouragement.

- **Search:**

This functionality is to keep track of the history of the sessions taken. That is, the user can get the information regarding what training was carried out, how much time was it, how many marks got, etc. on a specific date.

Since developing an entire system with all requirements is time-consuming, the functionalities will be prioritized and implemented according to the given time limit. so, in the end this research will yield a prototype system. The experiment will be carried out in two environments, for model creation and training the researcher will use Google Colab due to the availability of GPU (Graphics processing unit) facility whereas for using webcam and other functionality Jupyter notebook will be used. The dataset for training the models is available in Kaggle. There will be 3 datasets and in total 3 model need to build. The link to the dataset will be given in Appendix. The user interface will be build using Python's tkinter GUI (Graphical User Interface) package.

- **Primary Research Plan**

This section covers the set of tasks with the proposed timeline needed to complete the whole research project systematically. There may be slight changes in the timeline while doing the actual implementation.

Task 1: Performing Primary research by finalizing questions, creation of questionnaire using Online survey site and contacting the intended participants. Also, Research on OpenCV, deep learning models with their different methodologies for real-time implementation.

Task 2: Data collection and uploading it to google drive

Task 3: Data pre-processing-This includes data-splitting, data augmentation, the addition of callbacks like early stopping, tensorboard visualization, model checkpoints, etc.

Task 4: Building CNN (Convolutional Neural Network) deep learning model from scratch for human emotions and sign digits.

Task 5: Training the models with pre-processed images.

By the end of this task, the researcher gets the model weights needed for classifying emotions and ASL sign digits and can be downloaded as a ".h5" file.

Task 6: Building and training model for ASL alphabet using transfer learning technique. This is because the corresponding dataset is quite larger and building CNN from scratch and training it will take a lot of time and produce less accuracy. By the end of this task, the researcher gets the model weights needed for classifying ASL sign alphabets and can be downloaded as a ".h5" file.

After completing task 6, there will be three "h5" format files for three models

Task 7: Testing the models with test dataset and OpenCV with webcam. By this, it is possible to check if correct label is shown for emotions and signs. For testing it with OpenCV, a plain screen with webcam on will be generated and detected label will be shown above.

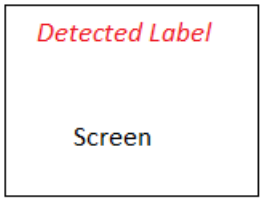

Fig2: Sample Test Screen

Evaluate the performance of the models by calculating the error rate and accuracy.

Task 8: Mapping the labels of the model to the corresponding emojis and voice recordings. This can be tested by webcam similar to the above step

Task 9: Implementation of screen capture, recording, screen-split mode, and test option. Integrating everything in the Tkinter user interface. Only by this task end, an interface like shown in Fig 1 will be created.

Task 10: Real-time recognition of facial expressions and sign languages with a webcam using OpenCV library.

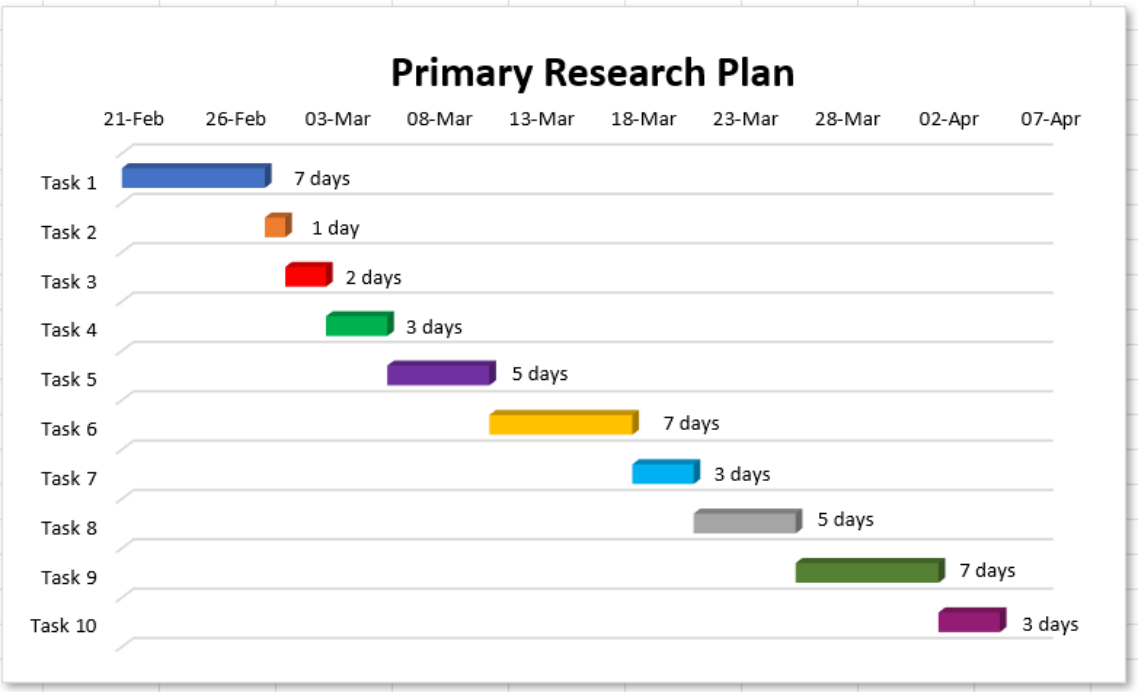

Fig3: Primary Research Plan

Risk Assessment

Description Of Risk	Risk Category	Impact	Probability
Misclassification of emotions and signs	Quality Risk	Low to Medium	70%

Table 1: Risk Assessment

Testing Steps:

- ✓ Run the system
- ✓ Check whether all buttons are present as per the diagram in Fig1 and check if the webcam is working properly.
- ✓ Click on screen-split start mode and check if the screen is split into two sections
- ✓ Show some emotion or sign and check if the correct emoji and voice recording is generated.
- ✓ Click on the screen capture button and check if a screenshot picture is generated in the corresponding location
- ✓ Click on the start recording button and wait for a few minutes and stop it.
- ✓ Check if the "avi" file is generated in the corresponding location.
- ✓ Click on the Test button and select the correct emotion or sign shown by selecting the checkbox. Then select the wrong one for another emotion or sign. Check in both cases if correct message is shown.
- ✓ Click on the search button and enter a date, check if the details of the corresponding class are shown.

The location for recording and screen capture images will be given in later stage of the project.

- **Mini Literature Review**

In the paper "Emotion Analysis from Facial Expressions Using Convolutional Neural Networks", the authors Irmak, M. C., Taş, M. B. H., Turan, S., & Haşıloğlu, A. (2021). have utilized the benefits of deep learning in their work to construct an emotion analysis tool that can identify seven different human emotions. For that, the authors have implemented a CNN model consisting of an input layer, an output layer, and six hidden layers using the TensorFlow and Keras library. In their experiment, they took the FER-2013 dataset consisting of image matrices and normalized it before giving it to the model. To get better accuracy and to reduce the issue of overfitting the authors have also incorporated batch normalization, dropout layers into their model. For the testing purpose, they have used the initially separated test image set and got an accuracy of 70%. The authors are claiming that their proposed system shows higher accuracy than other studies conducted with the same dataset. However, the authors have not utilized this benefit for real-time application. From this. I understood the requirement of incorporating the model into real-time applications.

Another facial emotion recognition approach is done by Joseph, J. L., & Mathew, S. P. (2021) in their paper "Facial Expression Recognition for the Blind Using Deep Learning". In this study, the authors intended to build an android application to assist blind people by capturing the facial expression of confront person and converting it to speech. Differently from the above work, here the authors have implemented two deep learning models with two different datasets using a proposed CNN architecture and another with a transfer learning approach and got an accuracy of 67.18% and 75.55% respectively. Later the model with higher accuracy got deployed as an android application. For CNN architecture, the authors have created a 10-set convolutional layer model that can detect seven emotions same as the first work whereas the other model is built using transfer learning with a pre-trained model which is capable of detecting four human emotions. For reducing overfitting, certain data augmentation techniques were also carried out. This work showcases the possibility of converting the identified labels of the model to speech and other functionalities.

An approach of recognizing sign languages using machine learning has been done by the authors Anupama, H. S., Usha, B. A., Madhushankar, S., Vivek, V., & Kulkarni, Y. (2021). in their project "Automated Sign Language Interpreter Using Data Gloves". By this, the authors aim to help the deaf and the dumb to communicate with the hearing community without an interpreter. In their project,

the authors have developed a system that can translate English alphabets and a few phrase signs with the help of sensor-based gloves and a KNN machine learning algorithm. After recognizing the sign, it is then displayed on a PC and then convert to speech with the help of Google's text-to-speech converter. Even though the developed system performs well, it is not user-friendly. That is, for the system to work the glove need to worn by the signer all the time which in turn may lead to a certain discomfort. Moreover, wearing such gloves for a long time may lead to overheating to the hands.

Thus in this research, to help autistic children a combination of human emotions and sign recognition is going to build using CNN and transfer learning techniques along with some other functionalities that are missing in the above studies.

- **Possible Risk**

The accuracy of the system depends on the images and its quality used for building the models and it requires deep training of the model with high processing power computers. Since Colab is going to use for model training, there is a limit to using GPU facility, and chances of disconnections are higher, hence there will be some wrong prediction of signs and emotions while doing real-time testing especially for sign digit dataset.

Bibliography

1. Irmak, M. C., Taş, M. B. H., Turan, S., & Haşıloğlu, A. (2021). Emotion Analysis from Facial Expressions Using Convolutional Neural Networks. *2021 6th International Conference on Computer Science and Engineering (UBMK)*, 570–574. https://doi.org/10.1109/UBMK52708.2021.9558917

2. Joseph, J. L., & Mathew, S. P. (2021). Facial Expression Recognition for the Blind Using Deep Learning. *2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON)*, 1–5. https://doi.org/10.1109/GUCON50781.2021.9574035

3. Anupama, H. S., Usha, B. A., Madhushankar, S., Vivek, V., & Kulkarni, Y. (2021). Automated Sign Language Interpreter Using Data Gloves. *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*, 472–476. https://doi.org/10.1109/ICAIS50930.2021.9395749

Appendix

Link to dataset

- ASL alphabets: https://www.kaggle.com/grassknoted/asl-alphabet
- ASL Sign digits: https://www.kaggle.com/rayeed045/american-sign-language-digit-dataset
- Human emotions: https://www.kaggle.com/msambare/fer2013