Impact Factor (SJIF): 5.301

International Journal of Advance Research in Engineering, Science & Technology

e-ISSN: 2393-9877, p-ISSN: 2394-2444
Volume 6, Issue 6, June-2019

PIXTALK :Picture Exchange Communication for Autism Spectrum Disorder in Children using Machine Learning

Aishwarya Gentyal , Sakshi Jaiswal, Pratikkumar Mohite, Mansi Shah
Savitribai Phule Pune University, India.
Pune Institute of Computer Technology (PICT)

***Abstract*—The aim of project is to develop an Android application specifically for Autism Spectrum Disorder (ASD) in children. Autism is defined as deficit behavior that has less social interaction and development. The children facing autism sense a lack of not being able to communicate well. The proposed system will use the concept of picture exchange communication. The proposed system will use Instructional System Design methodologies that help to create mobile based learning. Machine learning concepts will be used for the purpose of picture generation. The stages that will be followed for the proposed system design are analysis, design, development, implementation and evaluation. There will be a user interface developed with accordance to HCI principles. The application will be developed keeping in mind to be a less complex for simple usage. For the future expansion there could be multiple languages added and transfigured for the picture exchange.**

Keywords - Data mining, NLP, Autism, Picture exchange communication, Mobile based learning.

I. Introduction

Autism is considered as a mental disorder [4]. An autistic child do not tend to speak although the child is competent to spoke. The kid seems to be normal to a stranger, however, only after spending a good amount of time the problems can be noticed. Autistic child, sometimes get panicked to have conversation and feel shy to speak. That’s why it is very important to provide the proper training, treatment, attention and care to help them cure. Generally, social interaction of an autistic children is very low because they are more likely to be alone.

The autistics have difficulty using stereotyped phrases including intonation in communication. The repetitive, restricted behaviours and interest of autistic children are evident when they repeat the action like spinning objects. This system will involve design, development and evaluation of an android based app using Picture Exchange Communication System for the use of Autistic Spectrum Disorder (ASD) children [1]. This application will be developed for autistic children, which aim to function as an “Assistant” to teachers, mentors, caregivers and parents. This system will provide an interface for caregivers and autistic children as well. PixTalk app is based on HCI and usability convention which provide user friendly environment, interactive functions and voice at the background behind every picture which might help the kid to speak such as alphabet[5].

II. Motivation

In India out of every 10,000 children, 23 children (0.0023%) are suffering from autism. More than 70 million children overall world live with an autism spectrum disorder. An exclusive motivator for this is to guarantee a better and inexpensive mediation for the autistic children. There is a lack of economical intervention thus the application will help towards eradicating need for frequent therapist visits. The children diagnosed with Autism often struggle with the barrier of communication in the society. It mainly concerns with Picture Exchange Communication to bridge the communication gap between caregivers and autistic children.

III. Related study

3.1. Autism Spectrum Disorder (ASD)

An autistic child do not tend to speak although the child is competent to spoke. The kid seems to be normal to a stranger, however, only after spending a good amount of time the problem(s) can be noticed. Kids, affected by autism, sometimes get panicked to have conversation and feel shy to speak. This is thus very important to provide the autistic kids with proper training, treatment, attention and care to help them cure. The term “autism” was derived from Greek word “autos” which means “self”.

Picture Exchange Communication System (PECS) truly helped people grow spoken language, reduce tantrums & strange manners and improved socializing. The Autism app followed certain phases along with progress which are considered to be supportive for teachers as well as for parents. These phases are as follows:

How to Communicate (Phase I): This phase emphasized on difficulties with common association that exists in autistic kids. Large number of children having autism experienced trouble with shared communication [2]. Although impairments in communication within inhabitants differ greatly, pupils study to conversation using particular images for actions.

Discrimination between symbols (Phase II): Students learned to make their picks from dissimilar objects. For instance, students were asked what is their favourite food and to identify apple and orange from two different pictorial cards. Through emblematic composition, the autistic kids were capable to composite actual life story with imaginary component, with knowledge that story was not real [3].

Sentence Structure (Phase III): Learners were trained to make easy sentences pursued by images of different things being appealed.

Answering (Phase IV): PECS was used to response the query, “What do you want?”

3.2. GAN Model

Generative adversarial networks (GANs) provide a way to learn deep representations without extensively annotated training data. They achieve this through deriving back propagation signals through a competitive process involving a pair of networks. The representations that can be learned by GANs may be used in a variety of applications, including image synthesis, semantic image editing, style transfer, image super-resolution and classification.

The networks that represent the generator and discriminator are typically implemented by multi-layer networks consisting of convolutional and/or fully-connected layers. The generator and discriminator networks must be differentiable, though it is not necessary for them to be directly invertible. If one considers the generator network as mapping from some representation space, called a latent space, to the space of the data (we shall focus on images), then we may express this more formally as $G : G(z) \rightarrow R|x|$, where $z \in R|z|$ is a sample from the latent space, $x \in R|x|$ is an image and $|\cdot|$ denotes the number of dimensions.

In a basic GAN, the discriminator network, D, may be similarly characterized as a function that maps from image data to a probability that the image is from the real data distribution, rather than the generator distribution: $D : D(x) \rightarrow (0,1)$. For a fixed generator, G, the discriminator, D, may be trained to classify images as either being from the training data (real, close to 1) or from a fixed generator (fake, close to 0). When the discriminator is optimal, it may be frozen, and the generator, G, may continue to be trained so as to lower the accuracy of the discriminator. If the generator distribution is able to match the real data distribution perfectly then the discriminator will be maximally confused, predicting 0.5 for all inputs. In a basic GAN, the discriminator network, D, may be similarly characterized as a function that maps from image data to a probability that the image is from the real data distribution, rather than the generator distribution: $D : D(x) \rightarrow (0,1)$. For a fixed generator, G, the discriminator, D, may be trained to classify images as either being from the training data (real, close to 1) or from a fixed generator (fake, close to 0). When the discriminator is optimal, it may be frozen, and the generator, G, may continue to be trained so as to lower the accuracy of the discriminator. If the generator distribution is able to match the real data distribution perfectly then the discriminator will be maximally confused, predicting 0.5 for all inputs.

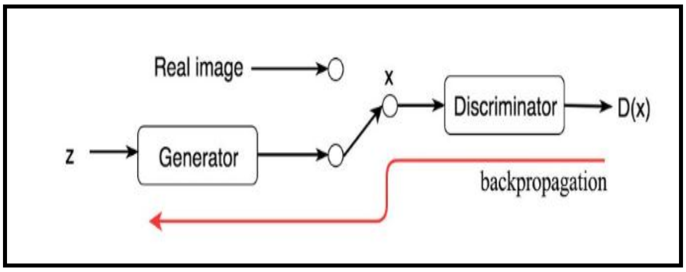


Fig 1. GAN Backpropagation

3.3. GAN Architectures

A. Fully Connected GANs

The first GAN architectures used fully connected neural networks for both the generator and discriminator. This type of architecture was applied to relatively simple image datasets, namely MNIST (hand written digits), CIFAR-10 (natural images) and the Toronto Face Dataset (TFD).

B. Convolutional GANs

Going from fully-connected to convolutional neural networks is a natural extension, given that CNNs are extremely well suited to image data. Early experiments conducted on CIFAR-10 suggested that it was more difficult to train generator and discriminator networks using CNNs with the same level of capacity and representational power as the ones used for supervised learning. The Laplacian pyramid of adversarial networks (LAPGAN) offered one solution to this problem, by decomposing the generation process using multiple scales: a ground truth image is itself decomposed into a Laplacian pyramid, and a conditional, convolutional GAN is trained to produce each layer given the one above. Additionally, Radford et al. proposed a family of network architectures called DCGAN (for “deep convolutional GAN”) which allows training a pair of deep convolutional generator and discriminator networks. DCGANs make use of strided and fractionally-strided convolutions which allow the spatial down-sampling and up-sampling operators to be learned during training. These operators handle the change in sampling rates and locations, a key requirement in mapping from image space to possibly lower dimensional latent space, and from image space to a discriminator. Further details of the DCGAN architecture and training are presented in Section IV-B. As an extension to synthesizing images in 2D, Wu et al. presented GANs that were able to synthesize 3D data samples using volumetric convolutions. Wu et al. synthesized novel objects including chairs, table and cars; in addition, they also presented a method to map from 2D image images to 3D versions of objects portrayed in those images.

C. Conditional GANs

Mirza et al. extended the (2D) GAN framework to the conditional setting by making both the generator and the discriminator networks class-conditional (Fig. 3). Conditional GANs have the advantage of being able to provide better representations for multi-modal data generation. A parallel can be drawn between conditional GANs and InfoGAN , which decomposes the noise source into an incompressible source and a “latent code”, attempting to discover latent factors of variation by maximizing the mutual information between the latent code and the generator’s output. This latent code can be used to discover object classes in a purely unsupervised fashion, although it is not strictly necessary that the latent code be categorical. The representations learned by InfoGAN appear to be semantically meaningful, dealing with complex intertangled factors in image appearance, including variations in pose, lighting and emotional content of facial images.

IV. PROPOSED SYSTEM

The proposed system is Generative Adversarial Network based system, which uses input in form of voice and generate an image card. Our aim is to create a system which helps Autistic Spectrum Disorder (ASD) children [1] to learn new words and things easily and fast. In PixTalk App, autistic kid can learn single words without any middleware. They can play different quiz to develop their IQ. GAN is main module used at care giver side for optimise searching of pictures. The whole system is based on Picture Exchange Communication System (PECS) and GAN. In GAN model, new image will generate from description and existing images of dataset even if that image is not present in dataset.

In previous system AutiPECS [1], PECS(Picture Exchange Communication System) and Instruction System Design (ISD) techniques for creating mobile based learning. Limitations of previously developed system is that it is available in only Malay language and can’t generate a new picture card if that is not present in dataset. GAN model: In GAN model, two neural networks are used for generating of new images from noise and discriminating image is fake or real. First neural network is generator which takes text features as input and generate a new picture and second neural network is discriminator which discriminate a generated image is fake or real using existing dataset.

Steps a GAN model architecture takes:

1. Voice Input is given to NLP model.
2. NLP Model
    a. Conversion of Speech to Text

- b. Pre-processing of text
- c. Feature extraction from processed text

3. NLP processed text is given to text encoder for attribute extraction from trained dataset
4. Word features are given to GAN Layer generation for new image generation
5. Generated image is given to image discriminator for return probabilities, a number between 0 and 1, with 1 representing a prediction of authenticity and 0 representing fake.
6. Discriminated image is loaded into image encoder and take another image from COCO dataset for best fit Matching score
7. Image get displayed

In the following figure[6], the two models which are learned during the training process for a GAN are the discriminator (D) and the generator (G). These are typically implemented with neural networks, but they could be implemented by any form of differentiable system that maps data from one space to another; see text for details.

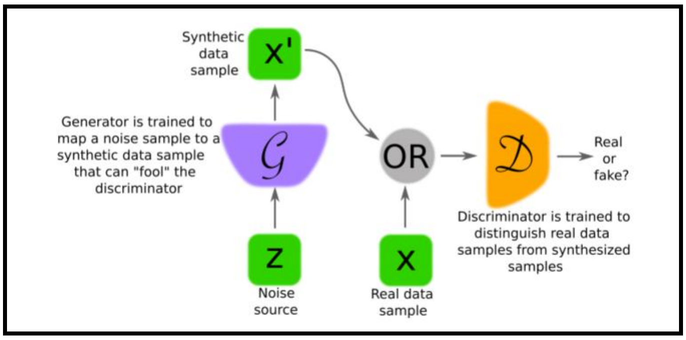

Synthetic data sample
X'
Generator is trained to map a noise sample to a synthetic data sample that can "fool" the discriminator
$\mathcal{G}$
OR
$\mathcal{D}$
Real or fake?
Discriminator is trained to distinguish real data samples from synthesized samples
Z
Noise source
X
Real data sample

Fig 2. GAN Generator and Discriminator

The following figure shows the GAN model architecture of the proposed system.

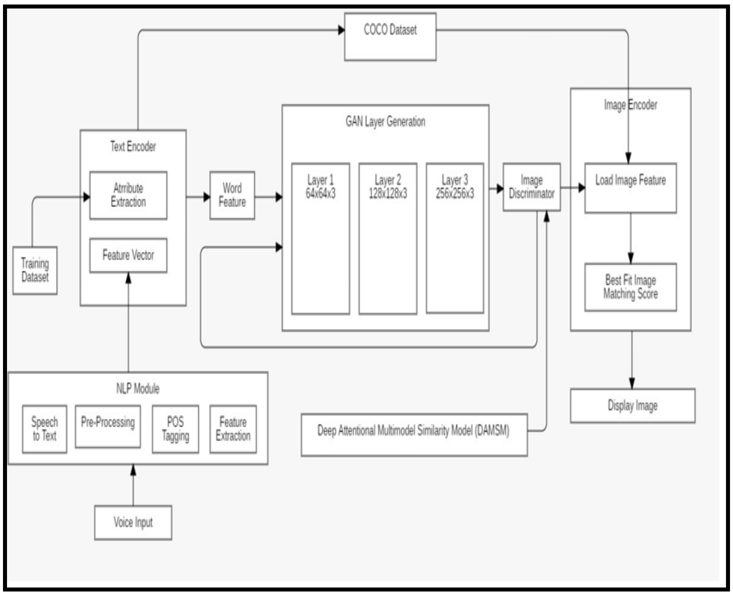


Fig 3. System Architecture

System Design:

Caregivers:
Communication: Voice will be recorded and converted into pictures. Audio cues are provided.
Image card generation :

GAN model will be used to generate images taking a voice input and plotting image card using class attributes with maximum similarity.

Autistic Children:

- We follow the phases of PECS. There are four options for increasing child's reading, visualisation and listening skills.
- Audio cues are associated with every image.

a) Single word learning:

- Pictures are tapped and audio is played.
- Subcategories provided like animals, food, emotions etc.

b) Counting made fun:

- Kids learn to count numbers.
- They can select cards to match with the numbers.
- Cards include three different objects and number cards from 1 to 9.

c) Differentiate:

- User selects and matches the object; sensing the similar and dissimilarity between object cards play into action.

d) QnA:

- Simple English questions are asked. Children can select a correct option.
- Scores are allotted based on answers given by children.
- Enhances learning and thinking capacity.

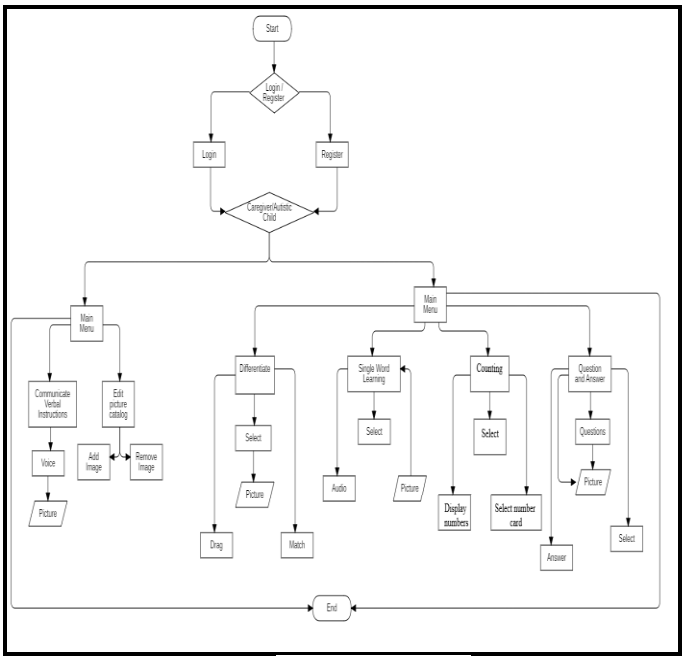


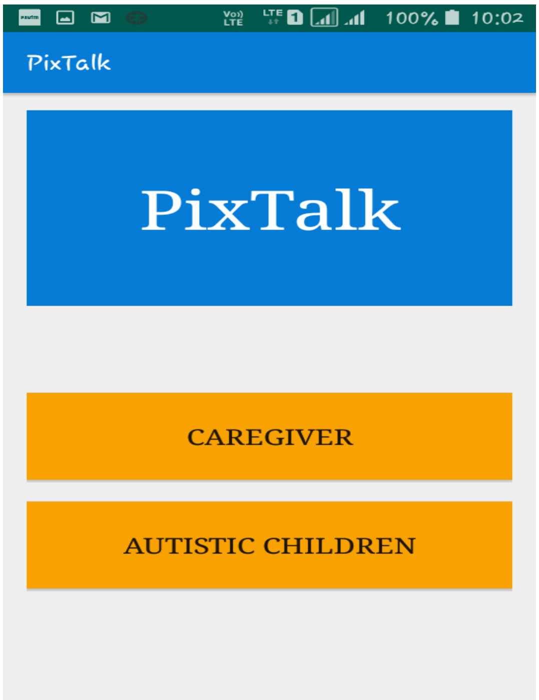
Fig 4. Activity Diagram

V. ALGORITHM

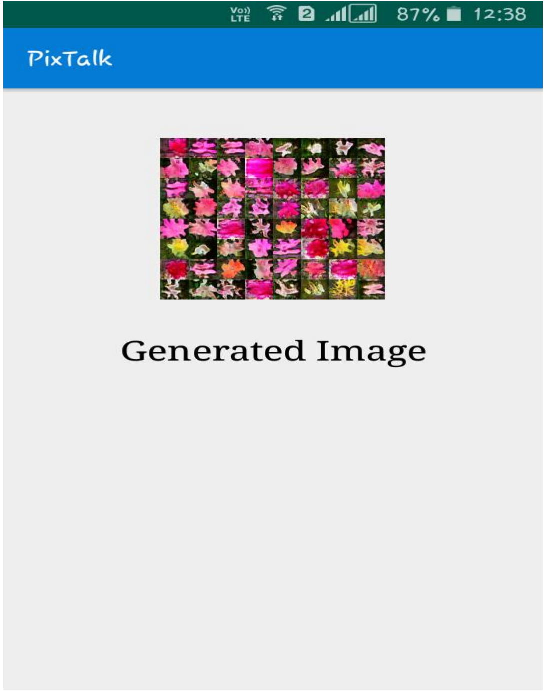
```
Minibatch stochastic gradient descent training of generative adversarial nets:
The number of steps to apply to the discriminator, k, is a hyperparameter. We used k=1, the least expensive option, in our experiments.
For number of training iterations do
Step 1: for k steps do
        •Sample minibatch of m noise samples {z $^{(l),\ldots,}$ z $^{(m)}$ } from noise prior $p_g$ (z).
        • Sample minibatch of m examples {x$^{(l),\ldots,}$ x$^{(m)\}}$ from data generating distribution $P_{data}$(x).
        • Update the discriminator by ascending its stochastic gradient:
        $\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^{m} \left[\log D\left(\boldsymbol{x}^{(i)}\right) + \log\left(1 - D\left(G\left(\boldsymbol{z}^{(i)}\right)\right)\right)\right]$
Step 2: end for
        • Sample minibatch of m noise samples {z $^{(l),\ldots,}$ z $^{(m)}$ } from noise prior $p_g$ (z).
        • Update the generator by descending its stochastic gradient:
        $\nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^{m} \log\left(1 - D\left(G\left(\boldsymbol{z}^{(i)}\right)\right)\right)$
Step 3:  end for
        The gradient-based updates can use any standard gradient-based learning rule.
```

VI. RESULTS

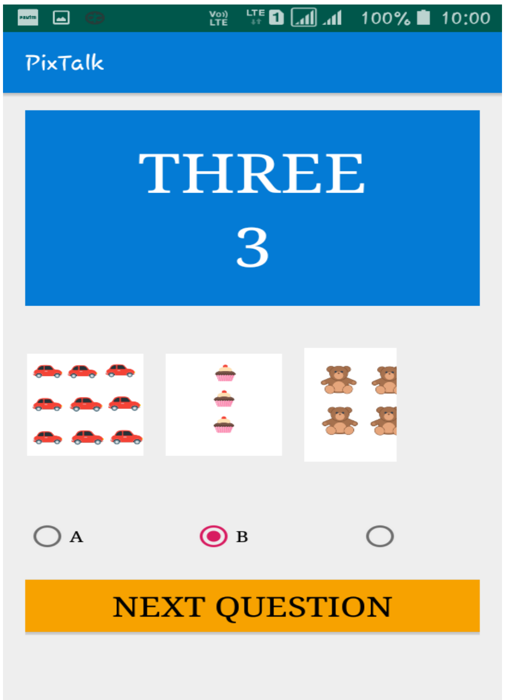
This is the first menu page of the application the user will select between two option as Caregiver or Autistic child.



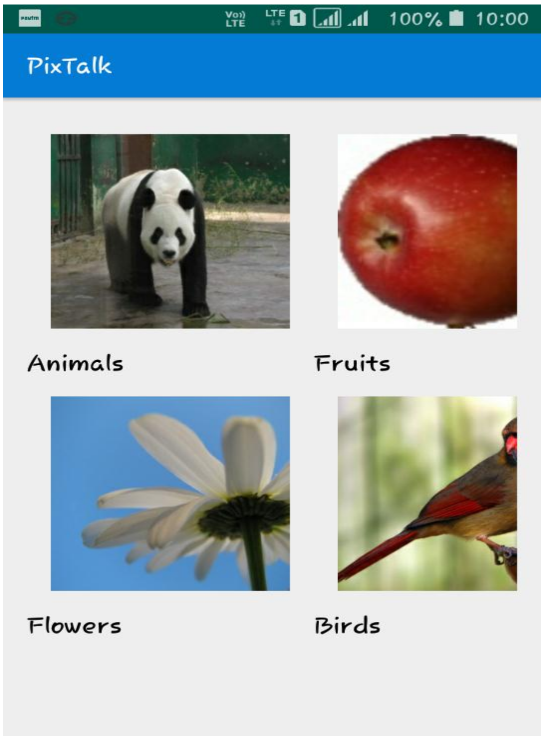
The Image card generated at caregiver side. The image is of 8x8 grid



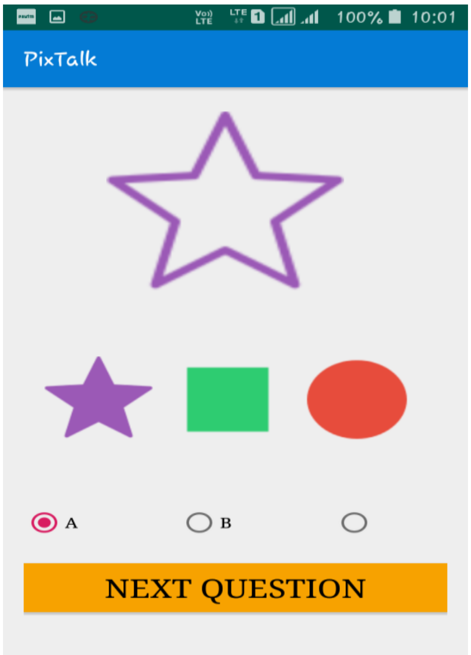
The counting pebbles activity having different levels.



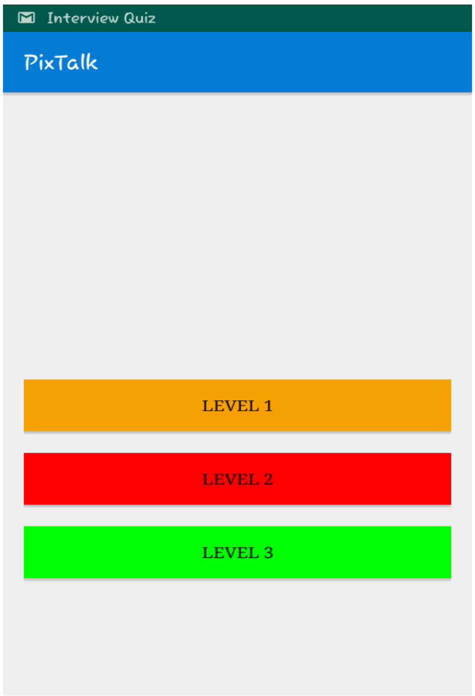
The single word leaning having four categories and voice cue associated with image.



In differentiate game child has to select filled shapes for outliner image question.



The Question and answer section having 3 levels to increase learnability of autistic child.



About Generative Adversarial Network

We validate that a GAN can perfectly reproduce a simple image from dataset. GAN model can train model in 5k iteration. After training new image can generate with 84% of accuracy. In Gan model new image is generated in three different layes with different resolution. At each layer clearity of image got improved. In proposed system, GAN model can generate new flashcards of different categories like food, emotions, fruits, shapes and colors etc.

VII. Conclusion

The proposed system eliminates the limitation of static picture retrieving from database. New picture is generated from existing dataset without availability of picture in dataset. Proposed system is significantly contributions, in terms of technology and psychometric training, to advance active learning of the autistic kids. It is to facilitate to enable communication and instinctive among those children who suffer from Autism. The four important phases of the PECS have been covered. Communication, discrimination between symbols, sentence structure formation and answering phases have a dedicated separate module. The conversion of voice input from the caregiver to flashcard images is done using GAN Model (Generative Adversarial Networks) [6] with 84% accuracy. The application can incorporate more deep voice recognition techniques to increase the functionality of the application and its applicable areas. Future research and work will include adding new language support, preferably Hindi and Marathi. Also, widening the scope of the flashcard dataset will be done.

References

[1] Ahmad Sofian Shaminan, Rabiatu Adawiyah Adzani, Sabariah Sharif,Nung Kion Lee, “AutiPECS: Mobile Based Learning of Picture Exchange Communication Intervention for Caregivers of Autistic Childre”, 2017.

[2] N. D. Londhe, M. K. Ahirval, P. Londha ,”Machine learning paradigm for speech recognition of Indian Dialect” , International Conference on Communication and Signal Processing, April 6-8, 2016.

[3] Mary Randolph,” Autism: A Systems Biology Disease”, IEEE, 2017.

[4] Inge Gavat, Diana Militaru, “Deep Learning in Acoustic Modeling for Automatic Speech Recognition and Understanding”, IEEE, 2015.

[5] Devashish Shankar,Sujay Narumanchi,Ananya H.A., Pramod Kompalli, Krishnendu Chaudhury, “Deep Learning based Large Scale Visual Recommendation and Search for E-Commerce”,IEEE, 2017.

[6] Tao Xu, Pengchuan Zhang, “AttnGAN: Fine-Grained Text to Image Generation with Attentional Generative Adversarial Networks”, IEEE, 2016.

[7] Shankar, Devashish, et al. "Deep learning based large scale visual recommendation and search for e-commerce." arXiv preprint arXiv:1703.02344 (2017).

[8] Xu, Tao, et al. "Attngan: Fine-grained text to image generation with attentional generative adversarial networks." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018..

[9] Gavat, Inge, and Diana Militaru. "Deep learning in acoustic modeling for Automatic Speech Recognition and Understanding-an overview." 2015 International Conference on Speech Technology and Human-Computer Dialogue (SpeD). IEEE.