Interactive System for Toddlers using Doodle Recognition

Gagandeep Konana Nagaraj00-0002-2632-4797, Atharv R Belagali†, Rashmi M† and Ram Mohana Reddy Guddeti†

[1*]Department of Information Technology, National Institute of Technology Karnataka, Surathkal, Mangalore, 575025, Karnataka, India.

Contributing authors: gagandeep.konana@gmail.com; atharv910@gmail.com; nm.rashmi@gmail.com; profgrmreddy@nitk.edu.in;
†These authors contributed equally to this work.

Abstract

Typing using the keyboard or using a mouse is hard for small children. In this paper, we proposed an interactive system to improve the learning ability of a toddler. The proposed doodle recognition system provides an attractive and efficient way to interact toddlers with computer systems by following the Human-Computer Interaction guidelines and deep learning. The most common practice that toddlers develop is scribbling random images, so we decided to use this skill to provide a gateway for the toddlers to interact thus and learning with computers by using our proposed simple interface. Toddlers come across developmental milestones in their phase of life. Developmental milestones are milestones achieved by toddlers for instance saying their first word and putting their first step forward etc..Toddlers move about more in the second year and are more aware of themselves and their environment. Their drive to learn about new things and people is also growing. We have researched and studies show that visual learning is far more efficient when compared to listening and chanting the words taught to us. So in this paper, we present a way to bestow knowledge while children do what they are good at:(scribbling). When the toddler (user) starts to scribble or draw something on the screen, whiteboard, or paper; the application goes into input mode, and as soon as the drawing is stopped the image on the screen or whiteboard is processed by the trained CNN model and the action is carried out based on the output of the model. The environment is set up by the parents or day care workers, Users analysis was done on four questions and based on the age groups and background.

Keywords: Convolutional Neural Network (CNN), Doodle recognition, Human-Computer Interaction (HCI), Toddlers, Learning ability.

1 Introduction

Human-Computer Interaction (HCI) is a study that mainly focuses on human interaction with computers. Humans play the main role of a user in the study, and in this paper, we considered toddlers as users. HCI makes the usability and interaction with the computers for the user easier. The main objective of the HCI study is to make easy use of navigation through the technology for the users.

The early stages of life are very crucial for a successful human life. Jean Piaget [1], a child development theorist, found that in the first two years of life (Sensorimotor stage), children generally learn through their movement and their senses. When they cross the 12 months, there are two sub-stages in toddler cognitive (thinking) development. The first one, usually occurs between 12 –18 months. In this period, toddlers learn through the process of trial and error. They are continuously trying to implement ideas that ponder in their head. Communicating ideas with them is very vital. Doing so helps toddlers process the information they gather and see that one can respect their thoughts. The second stage is typically from 18 – 24 months and is the "Beginnings of Thought" stage. Here, the toddlers learn symbolic thoughts and begin to pretend play. The concept that one thing can represent another occurs here. For instance, a block can be a telephone, or a paper plate could be a pie [1].

Our paper focuses on these two sub-stages, where a child can interact with the system through scribbles or doodle drawing on a whiteboard or drawing board (connected to the computer) and learn simultaneously. We followed many HCI design rules such as the Fitts' law, Hicks' law, and Don Norman's principles. Further, we proposed a CNN-based deep learning model to make the interactive system more efficient by real-time recognition of doodles. Sketches, unlike real photographs, are highly abstracted because they lack the rich elements of genuine photographs, such as diverse colours, backgrounds, and environmental information. Doodles or sketches are still meaningful enough to encompass a sufficient amount of significance despite all of these shortages and being drawn with only a few strokes [2].

The most well-known image identification and classification technique is the Convolutional Neural Network (CNN), often known as a ConvNet. Convolution is a mathematical linear action between matrices that gives it its name. CNN is a type of neural network that uses convolution rather than standard matrix multiplication in at least one layer [3]. A CNN's main building block is the convolutional layer. The parameters of the layer are a series of learnable filters (or kernels) [4].

Although existing works provide good outcomes with respect to doodle recognition, there is not much work in the platform, which solves toddlers' interaction with computers. So in this paper, we discussed a potential solution for the problem by using the limited skill of doodling that toddlers possess. To the best of our knowledge, the key contributions of this work are:

- Providing a way for toddlers to interact with the system and make sure their scribbling does not go wasted.
- Providing an alternative way of learning for young minds instead of traditional charts and books.
- Providing a platform for parents to keep track of the learning behavior of their kids.

In this new era of rapid development, product design must address not only practical and functional needs, but also emotional and psychological design needs, and in this proposed model, it must address the needs of a toddler[5]. The design and implementation of an effective approach to educating toddlers through doodles are novel in this study. Our major goal is to create a platform for toddlers to use their drawings and doodles to engage and learn with computers.

The remaining portions of the paper are structured in the following manner: The Human-Computer Interaction principles are discussed in Section 2, which provides an overview of how the suggested model adheres to these concepts. In Section 3, we address a background investigation as well as some related works that inspired the work that is being presented. The many phases that are involved in the proposed interactive system are broken down and explained in Section 4. In Section 5, additional information is provided regarding the tests that were conducted and the results that were achieved, as well as an analysis of the user experience. In the final part of the paper, that is, Section 6, we shall conclude the work and discuss some potential future paths.

2 HCI Design Guidelines

The proposed doodle recognition system is efficient in terms of usability and real-time execution. As taking the cursor to the destination and clicking the button to get results takes a lot of time and effort for a toddler, and it is less accurate as they are still at the starting of the learning phase; therefore, the proposed work solves this problem by giving out the output/ action as soon as the user finishes drawing the doodle on the board, which makes it easy and fast. The proposed system attempted to solve this problem by giving minimal buttons and a big rectangular tab or whiteboard for doodling, thus reducing the effort and time. The proposed system follows Don Norman's Principles of interaction design:

- The structure of tasks is simplified.

- Visibility: User gets to know as soon as he/she looks at the user interface about the options and their access.
- Feedback: User gets feedback after every action performed in the application to acknowledge the input.
- Affordance: As the entire system is based on drawings, it makes easier for the users (toddlers) to use drawings to interact and get the desired results as toddlers tend to have the image stored in mind rather than the word itself.
- Designed for error: As small children are meant to make errors too often; the proposed system is designed based on this assumption where the error is handled accordingly.
- If the similar-looking things do not output the same action, toddlers tend to get confused, so the images are processed carefully and made sure that the same output is produced every time a similar drawing is drawn.

Our application has been designed by considering the famous Neilsen's Heuristic principles like we have maintained a very simple and minimum design to avoid any confusions or complications which would increase our user base in a developing country like India. As the application is very simple and easy to maintain, error prevention and detection becomes very easy. Our application is consistent and also links with the real world. There is very high user freedom with respect to the features and design which makes our application more user friendly. We have tried to avoid any complicated language by substituting words with animations as much as possible to enhance our user base. We have considered all the possible user cases with respect to our application to avoid errors during run time. Thus we have made sure that all the basic HCI guidelines are incorporated which would make our application more attractive and simple thus increasing the user base.

3 Related Work

Machine learning based technique for doodle recognition is a unique topic where not much focus is given. But it is a vital area for the reasons stated in Section I. Let us look at the papers that we surveyed to make this work possible:

A multi-class classifier to classify doodles from Google's Quick, Draw! is introduced in [6]. This work classifies doodles into 345 distinct categories. In addition, the authors tested and compared various KNN and CNN versions for the classification of doodles. In [7], the authors performed image classification based on many diverse images containing two types of animals: cat and dog. Four distinct CNN structures with four different combinations of classifiers and activation functions are compared on CPUs. The images used were real and not doodles in [7]. The work in [8] used a publicly accessible dataset from [9] with 20,000 sketches divided into 250 groups to suggest the usage of convolutional neural networks (CNNs) to improve performance and boost recognition accuracy on sketches done by various people. In addition, using a ResNet technique, they investigated the influence of numerous hyperparameters on overall performance.

In [2], authors proposed a unique sketch recognition technique based on CNN. The proposed model included 21 layers and was tweaked automatically to determine the best-optimized model. To determine the efficacy of the suggested model, it was tested on Quick, Draw! dataset. According to the results, the proposed model's accuracy was assessed to be as high as 89.53%.

The study in [10] performed a statistical survey of three of the categories in the Quick, Draw! dataset: mountain, book, and whale. A Classification Neural Network was trained to obtain a classification score for the study of the drawing quality.

In [11], authors looks at the HCI requirements of toddlers in the context of Computer Assisted Language Learning (CALL) resources. The context of deployment is explained in this work. It explains several elements that must be considered when creating content for this particular learner group.

Authors propose a structure for children-oriented HCI based on children-oriented HCI design concepts in [12]. The framework incorporates a physical user interface, gesture interaction, Avatar-based HCI, high-level semantics interactive technologies, and natural child-computer interaction. They verify the structure using a children's video game.

In [13], the author aims to investigate the coordinated development of children's art skills and creativity using human-computer interaction technology, and examines the human-computer interaction for children development trend, the principles of human-computer interaction design for children, the connotation of children's art creativity, and art skills and creation. This paper performed a questionnaire survey of art teachers in five kindergartens based on the relationship of force to better understand the current status of the application of human-computer interaction technology in children's art education.

4 Proposed Methodology

The basic idea is to build a model that takes the input of a doodle drawn by the toddler and gives output accordingly, enabling the toddler to interact and learn with the computer. The basic architecture of the proposed model is as shown in Fig. 1. It mainly has five subsystems. Beginning with a subsystem for drawing doodles with three distinct options. This is followed by a subsystem for detecting the doodle that was drawn. Before passing the detected doodle to the deep learning-based prediction subsystem, the detected doodle is preprocessed by the preprocessing subsystem. A detailed description of each subsystem is given below.

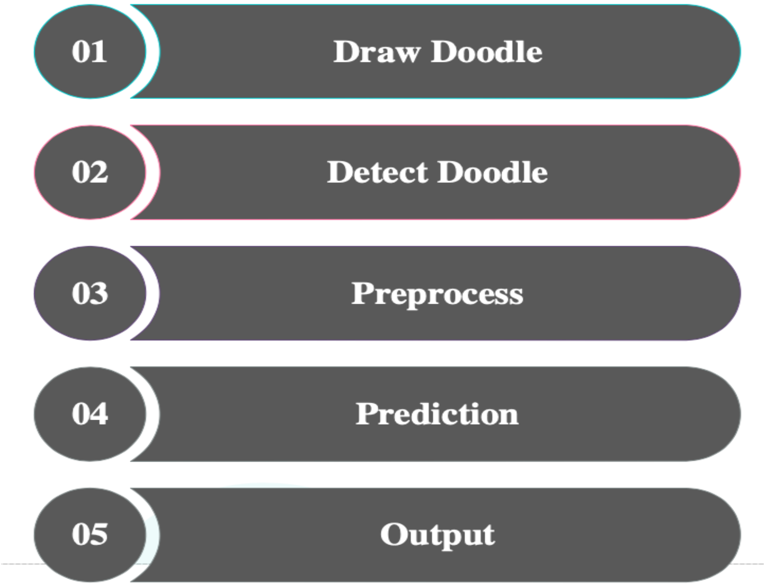


Fig. 1: Basic Architecture.

4.1 Draw Doodle

Users can choose the medium on which the doodle will be drawn. The input options available for the users are:

- Whiteboard: Users can use whiteboard with camera facing the board.Model will detect a rectangle in the camera frame and extract any doodle present in the frame to process it and predict the doodle.
- Drawing board: Users can use the drawing board connected to the laptop as an input method. The model will take the input of the doodle once the pen or stylus is lifted off from the drawing board.
- Gesture (hand movements). Users can use gesture as a form of input. the model will detect blue pointer on the camera and follow it till the "space" key is pressed on keyboard. this Doodle will be used as an input. This method of input did not produce good result as toddlers tend to have random hand movements and it requires manual interaction to press the "space" key.

4.2 Detect Doodle

We used OpenCV to capture the doodle. In the case of a whiteboard, the doodle is captured as soon as the doodling is finished and the whole rectangular frame of the whiteboard is visible. In the case of gesture, the user has to press the spacebar to indicate the completion of the drawing, and the doodle is captured on the key trigger of “Space”.

4.3 Preprocess

The drawing is perspective corrected to match the doodle in the training dataset. Images of dimension 28×28 are preprocessed in such a way that the values of pixel vary from 0 to 1. Now, this 28×28 is ready to be passed in the model.

4.4 Prediction

In paper [6], different model for doodle recognition on the same dataset is evaluated and CNN has outperformed other models. So we used the Convolution neural network (CNN) for Doodle recognition. The basic layout of the model proposed is as shown in Fig. 2:

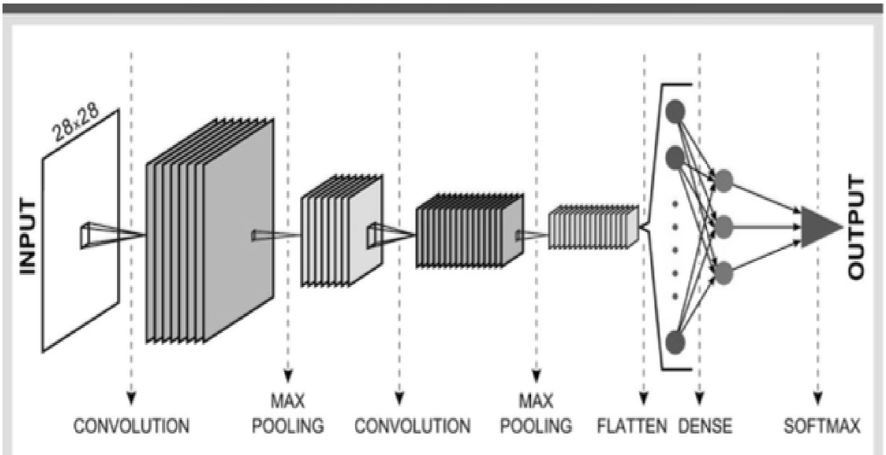


Fig. 2: Basic layout of the deep learning model

Conv2D is the first layer, with a kernel size of (5, 5), an input size of (28, 28, 1), and a ReLU activation function. Second layer is a MaxPooling2D with pool size = (2, 2), strides = (2, 2), and output dimension is equal to the first layer (padding = same). Conv2D is the third layer, with a kernel size of 5,5 and ReLU as the activation function. The fourth layer is a MaxPooling2D with pool size equal to (2, 2), strides equal to (2, 2), and preserving the dimension. After that, the array is flattened. The fifth layer is a dense layer with a ReLU activation function. The dropout rate of 0.6 was used and passed to the output layer with the softmax activation function.

4.5 Output

There are two modes that determine the output of the model. If the interaction mode is selected, the doodle-related operations are carried out, like opening paint application if the brush is drawn or opening the cake game if the cake is drawn. If interaction mode is off, it is in the learning model; the doodle is detected, and the actual images of the doodle are shown with its spelling and pronunciation. So, for example, if an apple is drawn, the real-time images of the apple are shown with apple's spelling and pronunciation spoken out loud.

5 Experimental Results and Analysis

5.1 Dataset

The Quick, Draw! dataset was used for the evaluation of the proposed model. The Quick, Draw! dataset is a Google dataset including 50 million drawings classified into 345 categories and collected from Quick, Draw! users [14]. We chose ten classes and had over 120,000 doodles for each of them. The images are grayscale and 28×28 pixels in size. Because the entire dataset is large (73 GB), we only used a portion of the dataset, with a total of 959364 doodles. More classes of doodles can be downloaded from Quick, Draw! to extend the dataset as per the requirement. Fig. 3 shows the sample doodles from four classes, namely (from left to right): a) star, b) baseball, c) apple, d) bow tie, e) clock, f) airplane, g) car.

Fig. 3: Sample doodles from the data set.

5.2 Training Phase

The data is split in the ratio of 9:1 for training and testing. The model is trained only on the training dataset, which consists of a total of 863427 doodles. The number of epochs is set to 30 with batch size =32 during training. Model is trained with learning rate of 0.001 and Adam optimizer.

5.3 Testing Phase

Testing on the dataset: The model predicts on testing data, these values are compared to their original classes. The performance on train and test set is evaluated to optimize the bias and variance of the model.

Testing on our doodle: Doodle, for which the class is to be predicted, is drawn on a whiteboard. The laptop camera detects the doodle automatically and corrects the perspective error. Then, this captured image is processed to match the data on which we are training our model.

Fig. 4: Capturing the doodle.

In Fig. 4, the rectangle around the whiteboard is detected, doodle inscribed in the frame is extracted.

In Fig. 5, The image extracted from the frame is perspective corrected to be fed into the model.

In Fig. 7, the basic workflow of the model is visualized. The preprocessed star doodle is fed into the network, and the model outputs the confidence percentage of each class, and the maximum out of that is chosen to continue the operation forward.

Fig. 6 is a Saliency Map that shows how much each pixel contributes to the result is plotted. A saliency map is basically an image in computer vision that depicts the distinctive quality of each pixel [15].

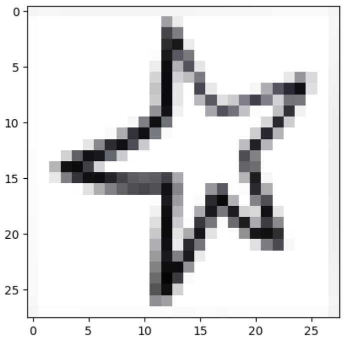


Fig. 5: Preprocessed images of star.

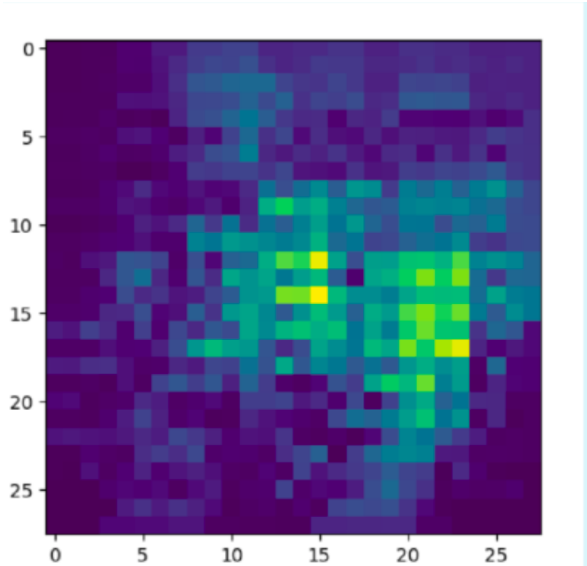
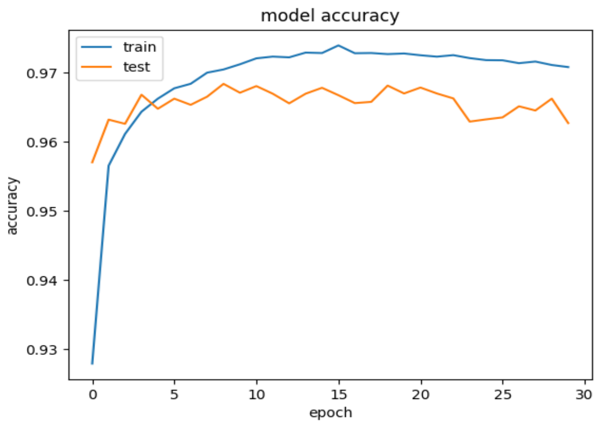


Fig. 6: Saliency map.



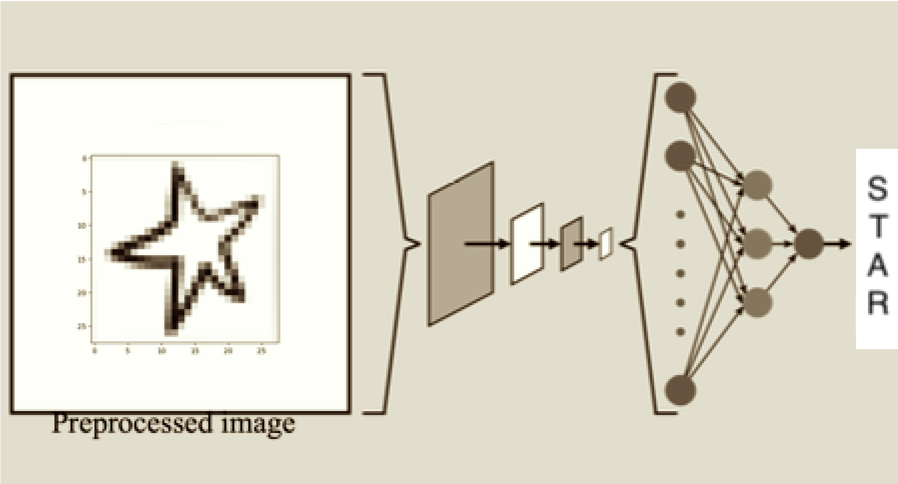


Fig. 7: Workflow of the network.

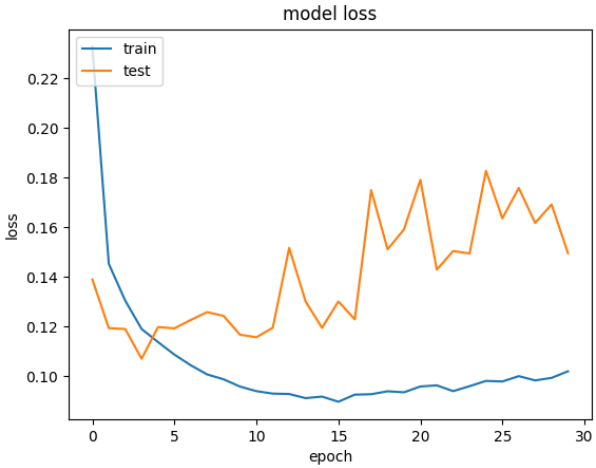


Fig. 9: Model loss vs epoch.

The Figs. 8 and 9 show the model accuracy and loss versus epoch for the dataset size of 40000 images for each class, respectively. There is not much improvement in training accuracy after epochs = 15, in contrast, the test accuracy fluctuates throughout. Also, there is no significant decrease in model loss after epochs = 15.

Table 3 shows the performance of the proposed model with two experimental settings with different dataset sizes. From this, it can be inferred that the accuracy of the model in training and the testing phase increased when the

Table 1: Performance of the proposed model.

Sl No.	***Data size***	***Epochs***	***Training Accuracy***	***Testing Accuracy***	***Avg Precision***	***Avg Recall***	***Avg F1-Score***
1	40000	15	97.24%	96.27%	0.965	0.965	0.9678
2	120000	15	98.2%	97.8%	0.96	0.96	0.958

[a]Datasize represents number of Doodles per class

Table 2: Performance comparison with different model.

Sl No.	***Data size***	***Model***	***Training Accuracy***	***Testing Accuracy***
1	120000	Linear SVM	75%	71%
2	120000	Linear SVM + KMeans	82%	78%
3	120000	CNN	98.2%	97.8%

[a]Datasize represents number of Doodles per class

size of the data was increased. Also, Using the same data, we have performed experiments using different machine learning algorithms. The comparison of the proposed model using CNN with other machine-learning algorithms is given in Table 2. It clearly shows that the proposed model using CNN stands superior.

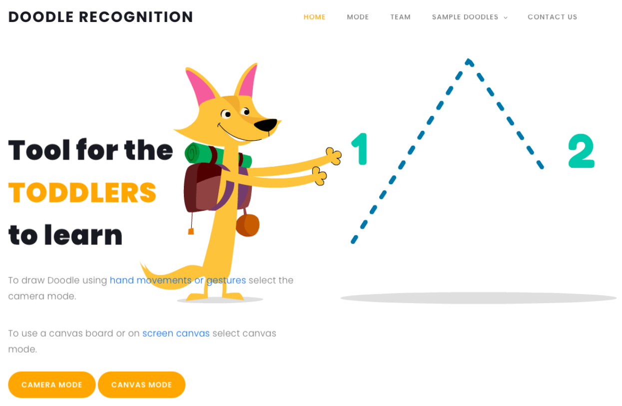


Fig. 10: Home page.

We evaluated the proposed interactive system in real-time with 17 parents and 10 daycare centers to get suggestions and experience analysis. We asked them to use the system for 3 days to answer questions based on the usability experience. The replies we received are encouraging.

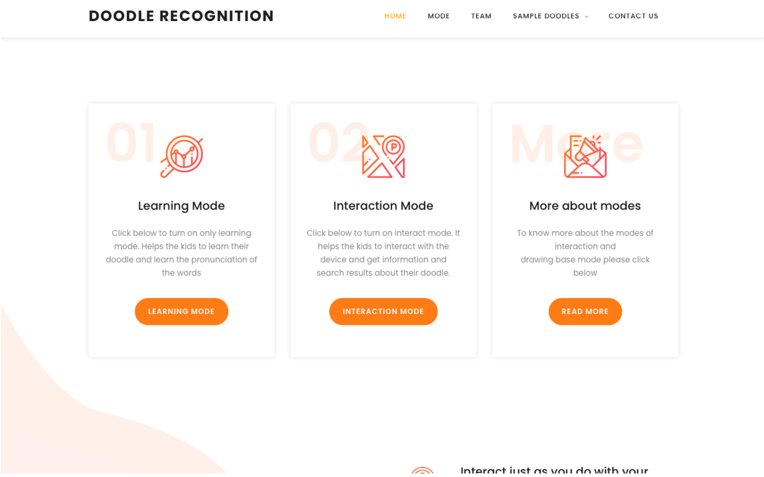


Fig. 11: Mode selection.

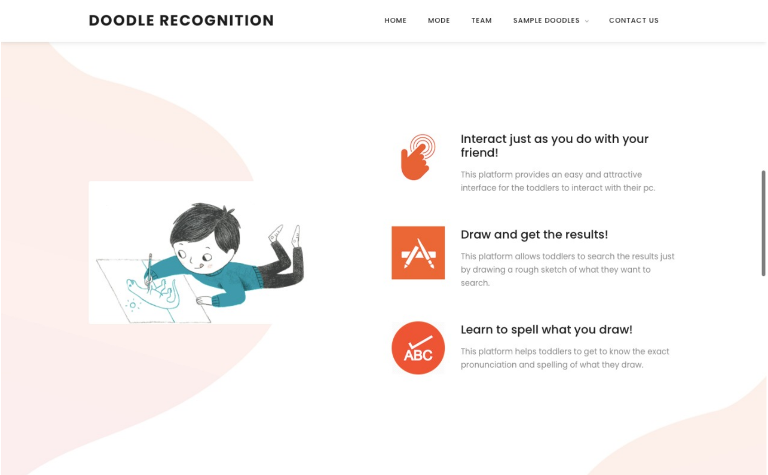


Fig. 12: More information about the model usage.

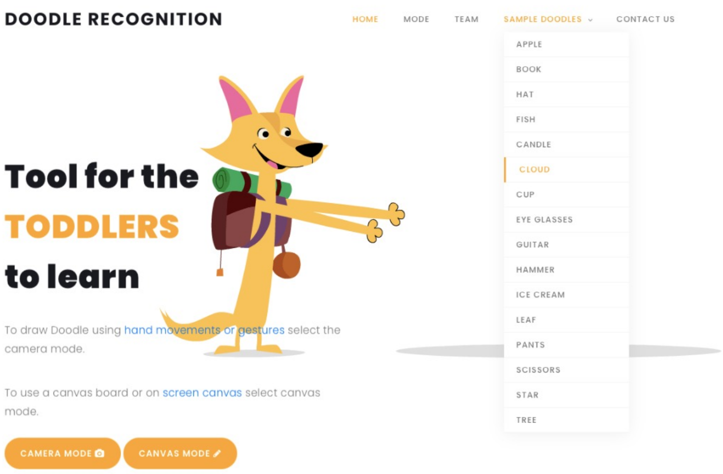


Fig. 13: Dropdown to select sample doodles to display.

Fig. 10 is the home page of the user interface provided for the test with parents and daycare centers. The menu bar on the home page has Home, Mode, Team, Sample Doodles, and Contact us to help the users navigate easily. There are two buttons that lets one to select the whiteboard (camera mode) or drawing board connected to the computer. Once the user selects one of the two buttons, they are navigated to the screen shown in Fig. 11, which lets the user select learning mode or interact mode; the explanation about the modes is clearly given in "Read more". More information about the working and usage of the model is provided as shown in Fig. 12. Sample Doodles can be accessed by selecting one of the doodle shown in Drop down like in the Fig. 13

5.4 Analysis

In the analysis, we have taken Feedback from 10 daycare centers and 17 parents/guardians regarding our model and asked them to rate our model based on the questions posed by us as displayed below with a graph for better comprehension.

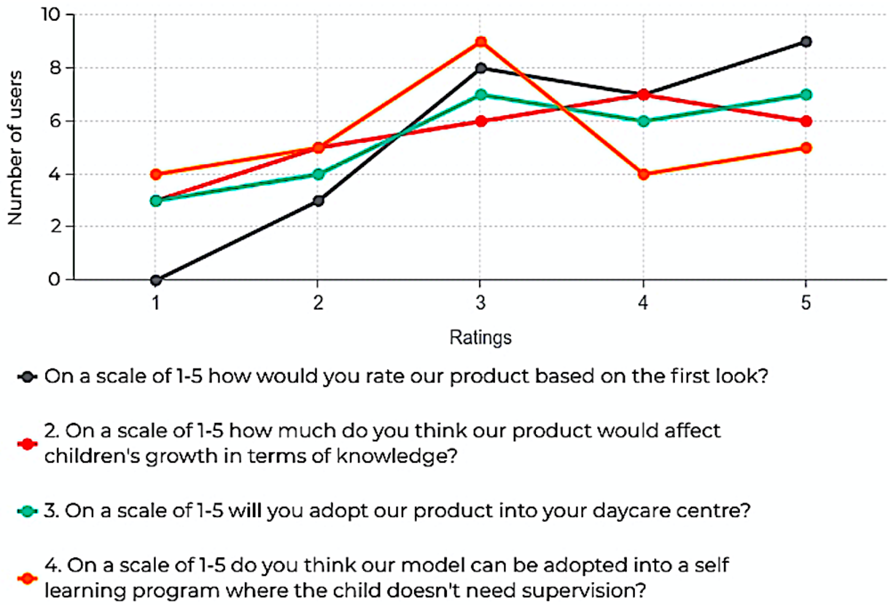


Fig. 14: Graph of ratings vs users.

In Fig. 14, the legend represents the question for the respective coloured line in the graph. The intent for asking these questions are as follows.

Question 1: We are taking reviews from customers and see in any way we can improve upon our current product and make it more user-friendly, reliant, and resourceful.

Question 2: This is a question which needs some time to be answered as it requires analysis of the toddler for a while. After a certain time 't' we can reference our users to an average toddler and measure improvements in their knowledge department and see how resourceful our product is.

Question 3: As discussed in the paper previously, the research has established visual learning is a faster way or perhaps a smarter way to establish knowledge in the minds of people. So, we think if the daycare does adapt our model, they can perhaps spare more time for other activities.

Question 4: This can be a little tricky as the product has a potential to act as a parental or a guardian figure for the supervision of the toddler. This can be theoretically achieved by monitoring the behavior of the toddler and seeing what he/she responds to. For instance lets take, he/she responds to a rhyme. We can use the rhyme to keep the toddler's attention on the product and not make other hassle or depending on the behavior might soothe the toddler and make him perform an activity like putting him/her to sleep.

The average ratings of the questions are as follows:

The pie graph in Fig. 15, was plotted to the response of the question: "Do you think the proposed interactive system has the potential to act as

Table 3: Average ratings for the questions

Sl No.	*Question*	*Average Ratings*
1	On a scale of 1-5 how would you rate our product based on the first look?	3.8
2	On a scale of 1-5 how much do you think our product would affect children's growth in terms of knowledge	3.3
3	On a scale of 1-5 will you adopt our product into your daycare centre?	3.4
4	On a scale of 1-5 do you think our model can be adopted into a self learning program where the child doesn't need supervision?	3.0

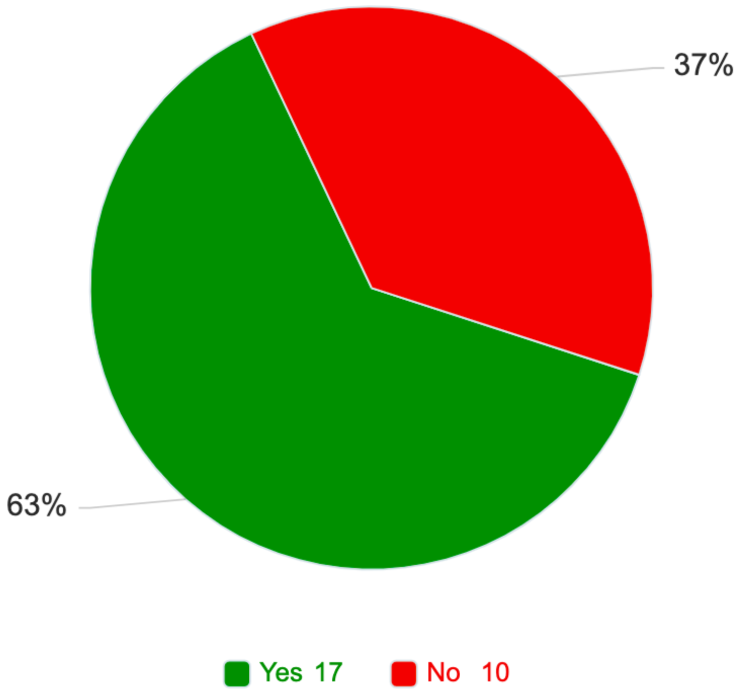


Fig. 15: Pie graph.

a secondary parent to teach their kids at the early stages of growth (yes or no¿'17 out of 27 users said yes, and 10 said no.

Time analysis was done on 5 trials and results are tabulated in Table 4. The activities that come under the time frame are: Opening up the website, Selecting the mode (interact or learning), Toddler doodling, Doodle recognition, Output.

Table 4: Time analysis

Mode	***Trail 1***	***Trail 2***	***Trail 3***	***Trail 4***	***Trail 5***
Learning mode	17	16	19	20	21
Interact mode	19	24	22	23	21

[a]Time recorded in Seconds

As the toddlers tend to have excess random movements, the gesture mode where the doodle input is taken from the hand movement of the toddler gave poor results.

5.4.1 Age Based User Analysis

Parents, guardians, and day care staff use the user interface in the form of a website to configure the learning and interaction environment for toddlers by choosing the appropriate modes before toddlers begin doodling. The analysis made below are based on the survey on group of people who use the first parse of the proposed model. We will be considering 3 parameters to classify our test users on the following basis:

1. Age which will be divided into three groups:

- Age group 1: Users between 15 to 30 years
- Age group 2: Users between 30 to 55 years
- Age group 3: Users above 55 years

2. On the level of computer literacy:

- Computer literate
- Computer non-literate

3. Based on the geographical factor:

- Urban
- Rural

Age group 1, Computer literate, Urban:

This group considered the interface to be really intuitive and encountered no issues when using it. Most of them found the features to be well distinguished and efficient. Some of them suggested to include more features and make it more professional while others were happy because even their care taker could make use of this as most of them were working and have very little time to spend with their child.

Age group 1, Computer illiterate, Urban:

Even though these people had some difficulties using the interface, particularly the interactive mode, the majority of them were sure that our proposed interface will become easier with time. They were delighted with the concept and found it to be quite useful.

Age group 1, Computer illiterate, Rural:

These people were very keen and excited to use our interface, despite encountering considerable difficulty while interacting with the computer. Since the majority of them did not have personal computers at home, it was quite difficult for them to explain the features and their usability. One of the most positive factors found in this age group was their inquisitiveness. They inquired extensively and were eager to spend more time with the interface. They found the interactive system to be quite intriguing, and the majority of them considered it to be highly beneficial for them along with the toddlers for their education. Overall we had a very positive response from them despite their unfamiliarity with the technology.

Age group 1, Computer literate, Rural:

These individuals shared the same level of enthusiasm as the illiterate individuals and discovered that the interface was simple and straightforward to utilise. They independently investigated all of the features that were offered, with very little assistance from us in the process. They considered the majority of professional websites to be really cluttered and difficult to navigate, therefore they appreciated how simple the layout was. Even though they were not as comfortable as urban people but we found them to be very fast learners and hence feel that this group of people can make significant use of our application.

Age group 2, Computer literate, Urban:

These categories of people represented the working class and contained the greatest number of people with professional backgrounds among all the groups. They absolutely had no problem in understanding the application as expected. Some complained that the features were very less and also the interface was too simple while others were happy about this as even their caretakers or their parents could use it. They focused more on the technical part of the application and also gave various feedback with respect to the design,algorithms used and what more features can be included.

Age group 2, Computer illiterate, Urban:

This was the group where we had a lot of mixed feedback and opinions. Some of them who had a personal computer at their home found this useful while people who did not have a personal computer found very little practical use of this application. Some of them believed that technology should be incorporated in their child's education while some were strictly against it. Most of them still preferred the traditional way of learning more impactful than this.

Age group 2, Computer literate, Rural:

We found very few people in this category and thus it was difficult for us to judge our application's effectiveness with respect to this category.

Age group 2, Computer illiterate, Rural:

These people found it very hard to understand the usability and purpose of this application. But most of them were very patient as they were very much interested in exploring new ways to educate their child and make them ready for the technological world in the future. Although these people had a lot of difficulty with the interface, most of them assured us that they will put all the effort required to get used to this application and make efficient use of it in educating their children.

Age group 3, Computer literate, Urban:

These groups of people found the application to be very useful and a productive way of spending their time with their grandchildren. They appreciated the purpose behind this application and also, urged us to concentrate our efforts more on the development of educational technology. Though most of them were very comfortable with the interface while others promised to spend more time with it. This group provided us with feedback that was overwhelmingly favourable and hopefully they will make effective use of it.

Age group 3, Computer illiterate, Urban:

The responses that we received from this group were highly diverse. While some people were enthusiastic about the idea of educational technology and urged us to focus our efforts in this direction, others were opposed to the thought using electronic gadgets as a medium of education. Most of them still believed that the classical method of learning in schools is the most effective one and technology cannot replace it. They were not very interested in exploring the application and as a result, we received a great deal of unfavourable comments from these individuals.

Age group 3, Computer literate, Rural:

We could not find people who fit in this group and hence the analysis of our interface with respect to this group was not possible.

Age group 3, Computer illiterate, Rural:

This age group shown the lowest level of interest compared to the others. They were adamantly opposed to the concept of integrating technology into the classroom because they were under the impression that attending classes, writing on a blackboard, and reading books are the only ways of obtaining education. In addition to this, they were of the opinion that the use of electronic gadgets by their grandkids would have a devastating effect.

6 Conclusion and Future Work

We have successfully implemented a program for toddlers which can be used unsupervised by any parental figure as it is user-friendly. We have also successfully followed the HCI guidelines which are presented in the introductory sub-section. We believe this tool will help the new generation of toddlers in learning fast as technology is developing and expanding its arena. This can be a very powerful tool that the toddler can use to enhance its knowledge as research has shown that interaction is one of the most proficient ways a person tends to learn and practically apply in real-world situations Toddlers successfully interacted with the system with their doodling. CNN model for doodle recognition worked with 97.8% accuracy. When the dataset is large enough, the error in the train and test data sets will be nearly identical. As a result, we can conclude that no over fitting occurred.

We conclude from the usability survey that parents and daycare centers will be willing to utilize the model as a tool to teach toddlers with some UI improvements. Usability survey based on the group of people revealed that there needs to be little work done on educating the use of the work and how this is more efficient than the traditional way.

The concept can be extended to a child's progress tracker. Automated quizzes can be conducted to evaluate a child's drawings. It can also be extended to intelligent supervision systems like AI nanny cameras. Interactive systems like robots can also incorporate the proposed work to enable better interaction with the toddlers on a physical scale. There is also a scope of extending the concept to determine the toddler's mood through the doodles drawn to better take care of their health and well being.

References

[1] Cherry K. The 4 Stages of Cognitive Development. Stages of Cognitive Development Guide. 2021 March;.

[2] Kabakus AT. A Novel Sketch Recognition Model based on Convolutional Neural Networks. In: 2020 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA). IEEE; 2020. p. 1–6.

[3] Goodfellow I, Bengio Y, Courville A. Deep Learning. MIT Press; 2016. http://www.deeplearningbook.org.

[4] Geron A. Hands-on machine learning with scikit-learn, keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems. 2nd ed. O'Reilly Media; 2019.

[5] Liu Y. Human-Computer Interface Design Based on Design Psychology. In: 2020 International Conference on Intelligent Computing and Human-Computer Interaction (ICHCI); 2020. p. 5–9.

[6] Guo K, WoMa J, Xu E.: Quick, Draw! Doodle Recognition. Stanford University.

[7] Chauhan K, Ram S. Image classification with deep learning and comparison between different convolutional neural network structures using tensorflow and keras. International Journal of Advance Engineering and Research Development. 2018;5(02):533–538.

[8] Lu W, Tran E. Free-hand Sketch Recognition Classification. CS 231N Project Report. 2017;.

[9] Eitz M, Hays J, Alexa M. How Do Humans Sketch Objects? ACM Trans Graph. 2012 Jul;31(4). https://doi.org/10.1145/2185520.2185540.

[10] Fernandez-Fernandez R, Victores J, Estevez D, Balaguer C. Quick, Stat!: A Statistical Analysis of the Quick, Draw! Dataset. ArXiv. 2019;abs/1907.06417.

[11] Ward M. HCI Requirements for Young Primary School CALL Learners. In: Zaphiris P, Ioannou A, editors. Learning and Collaboration Technologies. Technology-Rich Environments for Learning and Collaboration. Cham: Springer International Publishing; 2014. p. 339–348.

[12] Fang Z, Gu Z, Xu J. A Design for Children-Oriented Human-Computer Interaction. In: Wang Y, Li T, editors. Practical Applications of Intelligent Systems. Berlin, Heidelberg: Springer Berlin Heidelberg; 2012. p. 71–78.

[13] Sun Y. Coordinated Development of Children's Art Skills and Creativity Based on Human-Computer Interaction Technology. In: Al-Turjman F, Rasheed J, editors. Forthcoming Networks and Sustainability in the IoT Era. Cham: Springer International Publishing; 2022. p. 50–57.

[14] GoogleCreativeLab.: The Quick, Draw! Dataset. GitHub. https://github.com/googlecreativelab/quickdraw-dataset.

[15] Kadir T, Brady M. Saliency, Scale and Image Description. International Journal of Computer Vision. 2001 Nov;45(2):83–105. https://doi.org/10.1023/A:1012460413855.