1

# Multimodal Chatbot

Project submitted to the

SRM University – AP, Andhra Pradesh

for the partial fulfillment of the requirements to award the degree of

**Bachelor of Technology**

In

**Computer Science and Engineering**

**School of Engineering and Sciences**

Submitted by

Velaga Subhash (AP22110010068)

T. Ganesh Manikanta (AP22110010071)

Sai Sri Harshitha Voleti (AP22110010096)

Sajida Begum Mohammad (AP22110010098)

Akhil Chowdary Nalla (AP22110010102)

Yarra Krishna Teja (AP2110010105)



**SRM University–AP**

**Neerukonda, Mangalagiri, Guntur**

**Andhra Pradesh – 522 240**

**[December, 2022}**

**Acknowledgements**

We would like to express our deepest appreciation to all who contributed to the success of our project. Special thanks goes to our project supervisor Dr. Anusha Nalajala madam for their guidance and support which was instrumental in seeing our project through.

**Objective:**

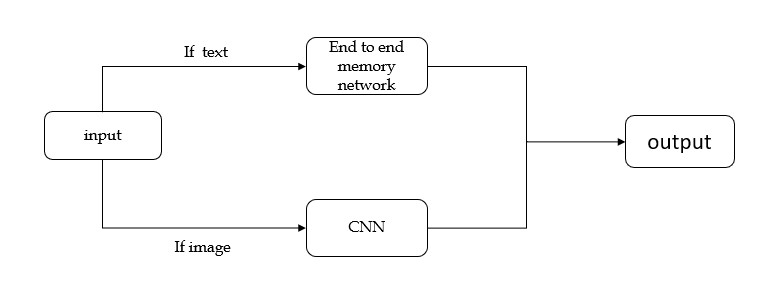
Our goal is to create a multimodal chatbot on covid 19 that assists users in answering their questions on a specific concept. When a user runs into issues during the process, the chatbot can offer solutions or directions based on how it categorizes the user's intent. We want to make our system better so that it can recognize the user's wants by inferring their purpose and then give them pertinent and useful information. By using the Convolutional Neural Networks (CNN) and End–to--End memory networks, this system captures visual and textual characteristics and outputs visual and textual features . We used End-to-End memory networks for textual data and Convolutional Neural Networks for image classifications and for outputs related to images i.e., Gender and age detection and to classify the scan reports of the chest.

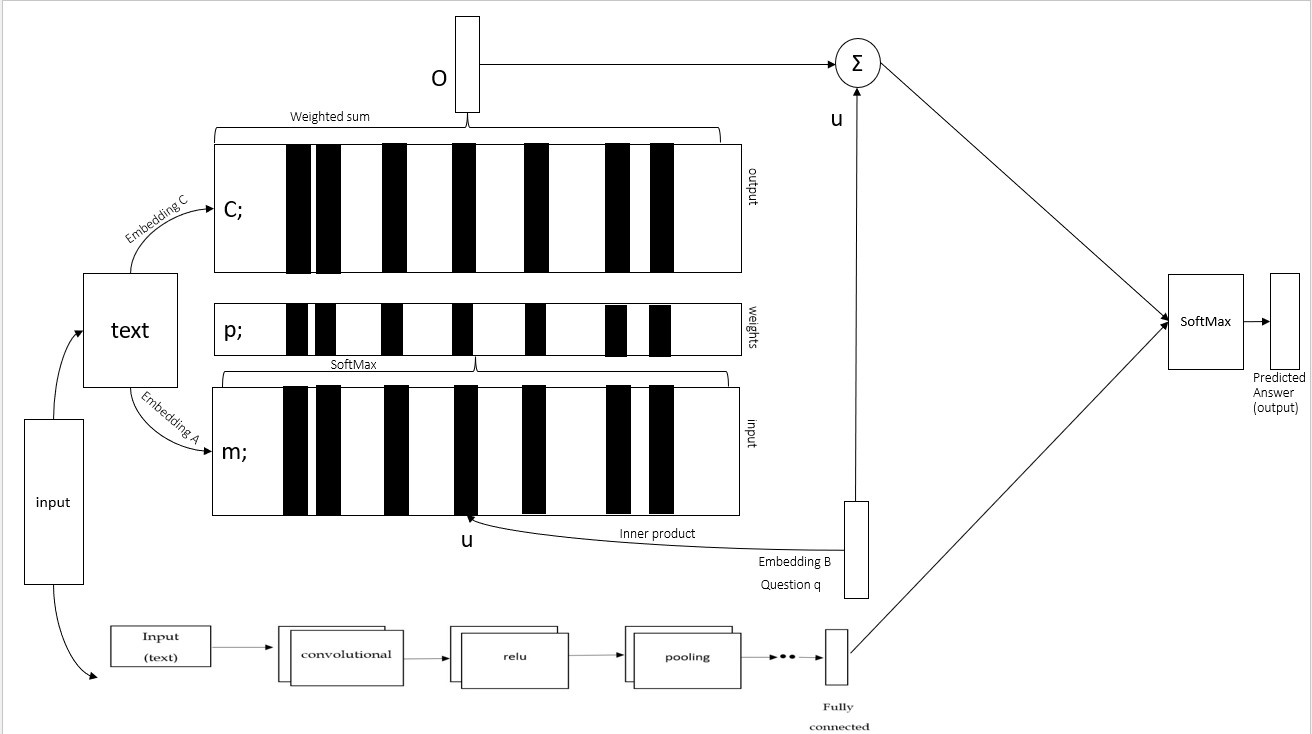
**Architecture:**

Chatbot architecture or working is as follows

It will accept the both textual and image data as input and it will use the appropriate neural network of the given data. End-to-end memory network if the data is text and Convolutional Neural Network if the data is image.

The block diagram will be as follows:





End-to-End Memory Networks:

An end to end network is a type of neural network which uses an iterative attention mode and it can span huge external memory. Unlike the model utilized in another study, this design is trained from start to finish and requires less supervision. It can also be considered an extension of RNN search in scenarios where multiple calculations (hops) are needed for each output symbol.

This model accepts a discrete set of inputs (x1;...;xn), and a query (p), and an output (q) (b). Each of the symbols for the xk, p, and an is taken from a lexicon that has V words. Once all of the x values have been stored up to a predetermined limit in memory, the model is used to generate a continuous representation for the x and p. The continuous representation is then processed in several stages to generate an output. These steps include identifying and eliminating any potential plagiarism and rephrasing the text to make it unique and original. This permits backpropagation of the error signal to the input during training due to numerous memory accesses.

Both the input and the text are transformed into word embeddings to obtain the internal state v. Compute the correspondence between each memory nk in the embedding space by computing the inner product and adding the SoftMax:

**ak = SoftMax(vTnk).**

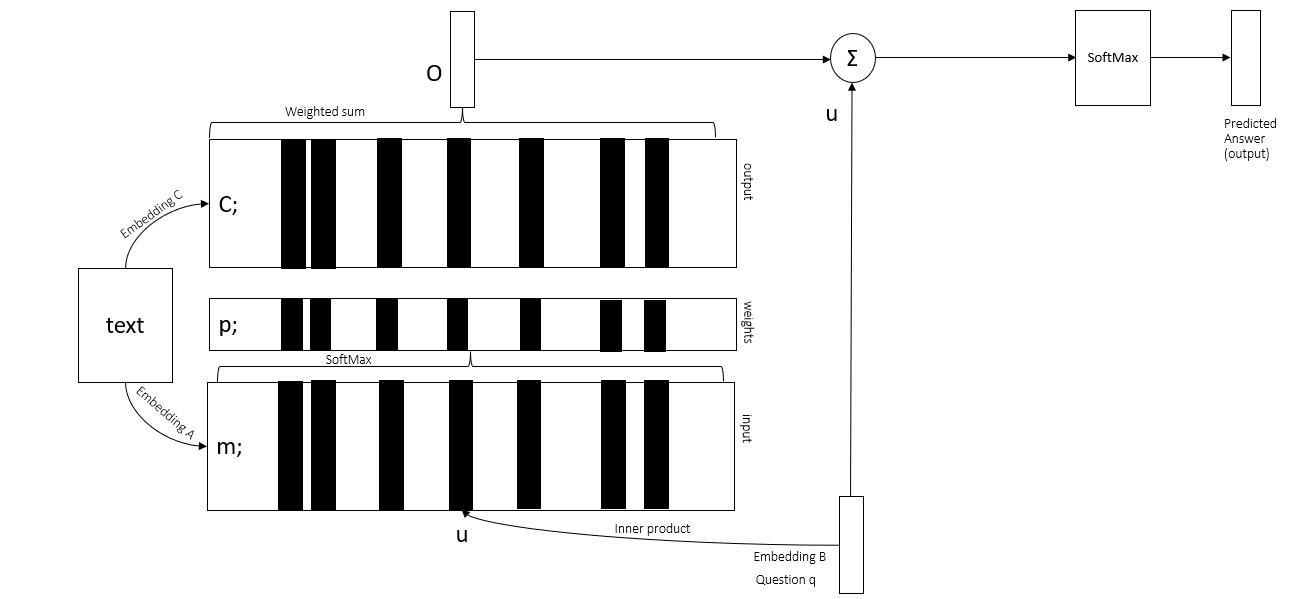
For every xk there is an equivalent output vector bk. The sum of the modified inputs bk weighted by the probability vectors from the inputs gives the response vector from memory:

**q =** 𝚺**akbk.**

The output vector p and the input embedding vector l are combined to form the predicted labels and passed to the final weight matrix U and SoftMax:

**â = SoftMax(U(p + l))**

End to End Memory Network Architecture:



Convolutional Neural Networks:

Convolutional neural networks perform better than other neural networks because they can input speech, audio, or visual data. CNN is used to solve computer vision issues including image recognition and classification. The way a computer perceives an image is as an array of Pixel values. Computers see images as values between 0 and 255, where 0 to 255 represents the current pixel intensity. In the CNN, there are 4 levels ,which are:

* Convolutional layer
* Relu Layer
* Pooling layer
* Fully Connected layer

**Convolutional layer:**

Most of the calculations are done in this layer and it also serves as a basic part of CNN. It applies a filter to an input to produce a feature map that lists the existence of any features that were found there. Here, there will be a template picture and a test image. The test image will match the features in the template image since we have produced them. One picture is transformed into a stack of filtered images in a convolutional layer; the number of filtered images depends on the number of filters.

**Relu layer:**

In this layer, all of the negative values from the filtered photos are eliminated, and their places are filled with zeros.

**Pooling layer:**

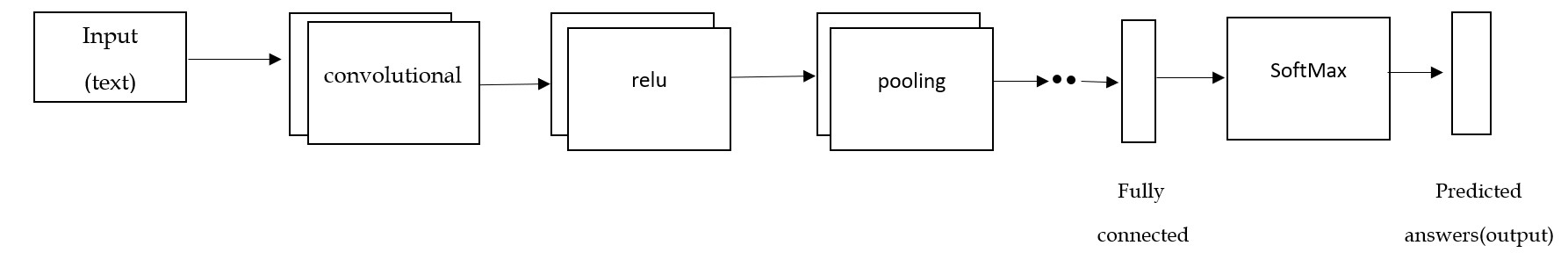
Pooling layers in CNN are used to reduce the spatial size of the input feature map. Used to reduce the number of parameters and reduce overfitting by providing an abstracted formal representation. By combining the output from a cluster of neurons in one layer into a single neuron in the next layer. This is done by applying a pooling operation to a small region of the input feature map, such as taking the maximum or average value of the region. Pooling layers are usually followed by convolutional layers and are used to reduce model complexity and also control overfitting . Then we need to select the Max.Pool & Avg.Pool

* **Max Pooling:** The filter examines the pixels in a designated window and selects the one with the highest value to send to the output array, as it sequentially traverses the input.
* **Avg. Pooling:** The value which is average in the window is calculated as the filter advances over the input, and it is then sent to the output array.

**Fully-Connected layer:**

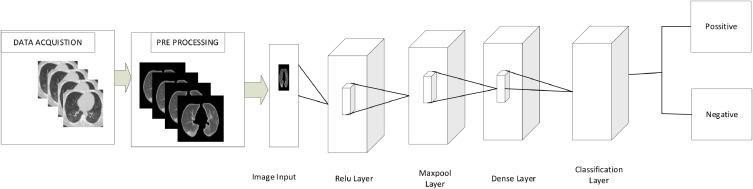
A fully connected layer in a is a layer that lacks convolution or pooling operations. It is the last layer in the network where the outputs from the preceding layers are compressed before being attached to each neuron in the fully connected layer. During training, the weights for each neuron are modified to allow for accurate classification of the features extracted from convolutional layers. To make predictions, SoftMax activation functions are usually employed to generate probabilities ranging from 0 to 1.

The below figure is the architecture of the CNN:



In this study, we create a chatbot that asks a user for their age and gender and, with the use of a CT scan, teaches the chatbot whether or not that user is afflicted by COVID-19. The chatbot uses CNN to provide the output after taking a specific user's picture. Male or female genders may be anticipated, and ages between 0 and 2 years old, 4 to 6 years old, 8 to 12 years old, 15 to 20 years old, 25 to 32 years old, 38 to 43 years old, and 48 to 53 years old may be predicted 60 years to 100 years old. It can be difficult to identify a person's actual age from a picture because of factors including cosmetics, lighting, impediments, and facial expressions.

Another module that detects Covid-19 positive and negative cases uses CT scan images of the lungs to classify them. The following processing steps are used to categorize positive and negative cases. image collecting, data augmentation, model design, iterative model training, model testing, and model performance evaluation. The architecture for detecting the covid positive and negative.

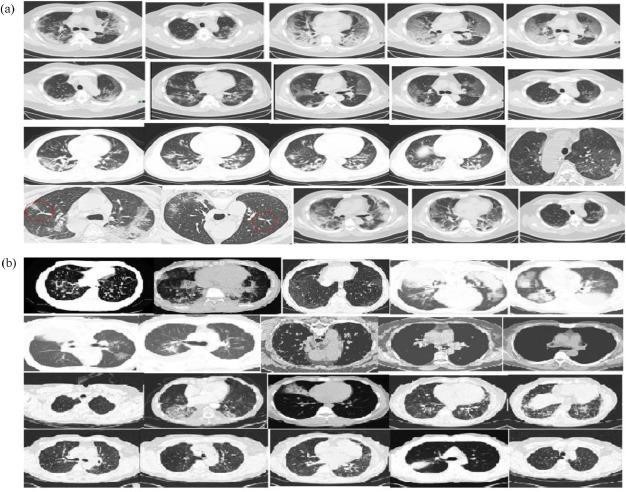


The lung is the primary target of COVID-19 in humans. A chest CT scan can be used to visualize the lung description. Images from the sample collection for both positive and negative situations are shown below. Images from the data set as follows .

**Dataset :**

We used the existing dataset of CT scans of lungs of both covid-19 positive and negative from Kaggle the api link goes below and some of the images of dataset are as follows .

kaggle kernels output chauhantamanna /covid19-detection-through-ct-scan -p /path/to/dest



(a)CT scan images from a sample COVID-19 Positive dataset.

(b)Sample CT scan images from the COVID-19 Negative dataset .

Data preprocessing simply transforms the raw data into a convenient and efficient format and this raw data is called dirty data. It may contain null values, duplicate values and unwanted values. This type of data provides inaccurate results, So this data should be changed accordingly and unwanted values should be removed in a way that the machine understands. Cleaning is done by removing duplicate, null and unwanted values.

**FEATURE EXTRACTION:**

Feature extraction is a process of reducing the unwanted resources and this process helps introduce components that are only needed to describe a large amount of data. Analyzation of complex data in machine learning comes with one major problem that is too many variables involved in the dataset. With too many variables to analyze it may need huge amounts of computation power, memory.

Concluding Remarks:

Covid-19 is very harmful nowadays. We are all working together to combat the COVID-19 pandemic. If properly developed and implemented, chatbots could benefit us by promptly disseminating current information, encouraging desired health-improving activities, and reducing the psychological harm brought on by isolation and anxiety. Despite this possibility, there is reason for caution due to the possibility of amplifying false information and the paucity of prior effectiveness study. Healthcare professionals, businesses, academics, and governments should work together immediately to prepare for potential pandemics.

Future Work :

We will develop a multi-modal chatbot which is used to detect the age and gender of the and we included a Covid19 chatbot which is used to detect weather the covid19 is positive or negative by using the help of CT scans which are provided by the user . Here we have used some approaches like CNN,MemN2N,MemNN,RNN and LSTM to build a chatbot. We are trying to increase the dataset to improve the accuracy of a result. We need to do more for improving the analysis model, and we have planed to integrate a decision support module to help users assess their risk of contracting COVID-19.

References

1. Madhavi M.Supraja P, [**Advances in Engineering Software**](https://www.sciencedirect.com/journal/advances-in-engineering-software) [Volume 173,](https://www.sciencedirect.com/journal/advances-in-engineering-software/vol/173/suppl/C) November2022,103214

<https://www.sciencedirect.com/journal/advances-in-engineering-software>

1. NourchèneOuerhani, AhmedMaalel, HendaBenGhézala, Soulaymen Chouri,2020 <https://doi.org/10.21203/rs.3.rs-33343/v1>
2. B.Abirami,T.S.Subashini,V.Mahavaishnavi,Materialstoday:Proceedings [Volume](https://www.sciencedirect.com/journal/materials-today-proceedings/vol/33/part/P7)

[33, Part 7,](https://www.sciencedirect.com/journal/materials-today-proceedings/vol/33/part/P7) 2020, Pages 4708-4712 <https://doi.org/10.1016/j.matpr.2020.08.350>

1. Memory Networks, JasonWeston, Summit Chopra & Antoine Bordes , Facebook AI Research,770 Broadway, New York, USA.
2. End-to-End Memory Networks, Sainbayar Sukhbaatar, Dept. of Computer Science , Courant Institute, New York University, Arthur Szlam JasonWeston Rob Fergus, Facebook AI Research, New York.
3. An Introduction to Convolutional Neural Networks Keiron O'Shea, Ryan Nash[Submitted on 26 Nov 2015 (v1), last revised 2 Dec 2015 (this version, v2)]
4. Understanding of a Convolutional Neural Network by Saad ALBAWI , Tareq Abed MOHAMMED Department of Computer Engineering Faculty of

Engineering and Architecture Istanbul Kemerburgaz University Istanbul, Turkey. [https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8308186.](https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8308186)

1. Convolutional neural networks: an overview and application in radiology Rikiya Yamashita, Mizuho Nishio, Richard Kinh Gian Do & Kaori Togashi Insights into Imaging volume 9.

https://link.springer.com/article/10.1007/s13244-018-0639-9