Handwritten Text Recognition

Submitted in partial fulfilment of the requirements of the degree of

BACHELOR OF COMPUTER ENGINEERING

by

Avishkar Dalvi, 20202002

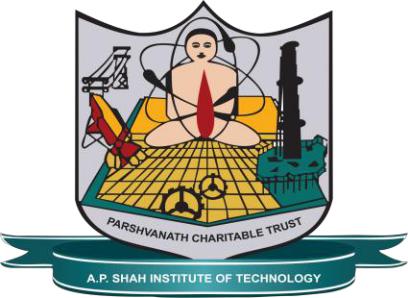
Riddhi Narkar, 19102003

Soumyojyoti Dutta, 19102014

Vedang Gore, 19102065

Guide:

Prof. Deepali Kayande



Department of Computer Engineering

1. P. SHAH INSTITUTE OF TECHNOLOGY, THANE

(2021-2022)



A. P. SHAH INSTITUTE OF TECHNOLOGY

CERTIFICATE

This is to certify that the Mini Project 2B entitled **“Handwritten Text Recognition”** is a bonafide work of **“Avishkar Dalvi (20202002), Riddhi Narkar (19102003)**, **Soumyojyoti Dutta (19102014), and Vedang Gore (19102065)”** submitted to the University of Mumbai in partial fulfilment of the requirement for the award of the degree of **Bachelor of Engineering** in **Computer Engineering.**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Prof. Deepali Kayande

Project Guide

\_\_\_\_\_\_\_\_\_\_\_\_\_

Prof. Sachin Malave

Head of Department

A. P. SHAH INSTITUTE OF TECHNOLOGY

Project Report Approval for TE

This Mini project report entitled ***Handwritten Text Recognition*** by ***Avishkar Dalvi , Riddhi Narkar, Soumyojyoti Dutta, and Vedang Gore*** is approved for the degree of ***Bachelor of Engineering*** in ***Computer Engineering***, ***2021-22***.

Examiner Name Signature

1.\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

2.\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Date:

Place: Thane

**Declaration**

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Avishkar Dalvi 20202002

Riddhi Narkar 19102003

Soumyojyoti Dutta 19102014

Vedang Gore 19102065

Date:

**Abstract**

Handwritten Text Recognition (HTR) is a very important field of research in the domain of Machine Learning, Image processing, and Artificial Intelligence. This is because it often involves decoding human written letters and words in a specific language. This project involves recognition of individual words. We are using ANN (artificial neural network) as it mimics the biological functioning of the brain, and hence is very efficient according to research for this problem statement. We used a Kaggle dataset with 4L transcripted images of names written in capital letters. This is a comparative study by using 4 different activation functions and by keeping the model, algorithm and dataset fixed.

**CONTENTS**

1 Introduction ........................………………………………………………………………9

2 Literature Survey ………………...………………………....…….………………………12

1. Problem Statement.………………………...…...………………………………….……..15
2. Objective and Scope....….………………………………………………..……………….16
3. Experimental Setup.…...…………………………………………..………………………16

5.1 Hardware Requirement

5.2 Software Requirement

1. System Design……………………………………………………………………………17

6.1 UML Diagram

6.2 Algorithm/Process (with Expected input and outputs)

1. Implementation.….…………...…………………………………………………………20

7.1 Code

7.2 Screenshots

1. Result………....………………...……………………….………………….……………23
2. Conclusion………………………………………………………...………….………….24

References 25

Annexture 1- Project Planning (Using Gantt chart) ……..………………………………..27

**LIST OF FIGURES**

1.1 Unistrokes character set by Goldberg and Richardson 10

1.2 Levels of HTR (simplest at the top and complex and the bottom) 11

6.1 UML diagram 17

7.2.1 Sigmoid Activation function 20

7.2.2 Elu Activation function 22

7.2.3 Selu Activation function 23

7.2.4 Relu Activation function 25

**LIST OF TABLES**

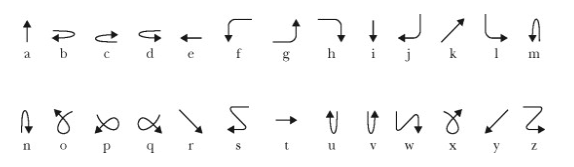
2.1 Research papers with their authors and abstract…………………………………… 13

8.1 Comparative study of different activation functions and their accuracies…………………………………………………………………………… 23

1. **Introduction**

Handwritten text recognition (HTR) involves an artificially intelligent system to interpret human written words into a machine readable format, for example, Unicode, or editable text. There are two versions of this technology, the first is an real time recognition, and the second is recognition based on a scanned image of handwritten text. The real time recognition involves an input device, such as a touchscreen tablet or a digital drawing and writing tablet and the recognition runs live when the user writes something using a digital device like a digital pen or a stylus on the tablet. This focuses a lot more on the type of curves and shapes of the material being written on the device to interpret it. Some of the first industry examples of such technologies was the handwritten notes to Unicode characters conversion feature of Apple’s Newton which was launched in 1992. Although it was discontinued after some time, but today, a lot of devices and software have been launched which try to detect and convert handwritten text in real time with significantly more accuracy.

In 1993, Goldberg and Richardson from Xerox PARC published a paper about *Unistrokes* (Goldberg & Richardson, 1993). Goldberg and Richardson wanted to design a character set that could be entered in an eyes-free manner. Hence, they came up with a very simplified version of all 26 English alphabets and called it Unistrokes. Unistrokes was built to optimize the recognition accuracy and text entry speed. Here, each character is drawn using a single stroke. But to put it to practice, one needs to learn this alphabet set. It is illustrated in Fig 1.1. Hence, this was not possible to be used in systems where raw data, in the form of normal English writing could be processed. This would require users to change their normal style of writing, if it were to be used in handwritten text recognition softwares.



*Fig 1.1 Unistrokes character set by Goldberg and Richardson*

Handwritten text recognition and OCR, or Optical Character Recognition are very similar. One of them is a special case of the other, and the other can be used as an umbrella term. So OCR is an umbrella term which involves recognizing any component of human language in the written form. Here human language, suggests any language which humans use to communicate, like for example English. Handwritten text recognition involves Optical Character Recognition, or is a ‘special’ case under OCR where we only deal with handwritten language components.

Handwritten text recognition (HTR) could be again further classified into many different levels. The most basic level is where we deal with recognition of individual characters of a language, in which sense, we call it as handwritten character recognition or HCR. For example, recognizing A, B, g, m, etc. (lowercase and uppercase both included ). The next level consists of individual words. For example, tree, computer, Rahul, etc.

Further down the line comes whole sentences. For example, Let’s play chess, I like to drink milk at night, etc. This level is the most complex and also the most rewarding, in the sense that a software which is able to recognize at this level can have myriad use cases in our day to day lives. This complexity is illustrated in the Fig 1.2.

*Fig 1.2 Levels of HTR (simplest at the top and complex and the bottom)*

In this research project, we worked on handwritten text recognition of English language upto the level of recognizing words.

1. **Literature Survey**

The research space in this domain reflects a lot of attention to the real use cases of this problem statement. The real and actual use case of this problem statement lies in the most complex stage of this problem. The level where the software is able to take any input from the user with any style of handwriting and is able to convert that into Unicode with an acceptable accuracy. Doing is much of a challenge in its own sense, and not a lot of research regarding this exists in the public domain.

A lot of research about this involves finding out the best ML model which can solve this problem, or by using existing model to recognize characters from a different language, which in itself is a much broader domain. Some research also involves applying this concept to historical documents. The challenges when this concept is being used for historical documents, the image processing part needs to be of a significant importance in the overall development.

A lot of research suggest that ANN is a very good option to get started with this problem. Artificial Neural Networks, or ANN mimics the neurons in our brain. It is literally based on the idea of how real, biological neurons behave. ANNs have evolved into a broad family of techniques that have advanced the state of the art across multiple domains. The simplest types have one or more static components, including number of units, number of layers, unit weights and topology. Dynamic types allow one or more of these to evolve via learning. The latter are much more complicated, but can shorten learning periods and produce better results. Some types allow/require learning to be "supervised" by the operator, while others operate independently. Some types operate purely in hardware, while others are purely software and run on general purpose computers.

For this specific problem statement, CRNN, or Convolutional Recurrent Neural Networks is something a lot of research papers suggested.

|  |  |  |  |
| --- | --- | --- | --- |
| **Research papers** | **Year Published** | **Authors** | **Abstract** |
| Handwritten Optical Character Recognition (OCR): A Comprehensive Systematic Literature Review (SLR) | 2020 | Jamshed Menon,  Maira Sami,  Rizwan Ahmed Khan,  Mueen Uddin | During last decade, researchers have used AI/ML tools to automatically analyze handwritten and printed documents in order to convert them into electronic format. The objective of this review paper is to summarize research that has been conducted on character recognition of handwritten documents and to provide research directions. |
| Handwritten Text Recognition in Historical Documents | 2018 | Harad Scheidl,  Robert Sablatnig | HTR is an automatic way to transcribe documents by a computer. The proposed system is based on ANN, i.e. Artificial Neural Networks. |

|  |  |  |  |
| --- | --- | --- | --- |
| Handwritten Text Recognition System based on Neural Network | 2016 | Ahmed Mahi Obaid, IIHazem M. El Bakry,  IIIM.A. Eldosuky, IVA.I. Shehab | Handwritten text recognition is still an open research issue in the domain of Optical Character Recognition (OCR). This paper proposes an efficient approach towards the development of handwritten text recognition systems. 3-layer Artificial Neural Network (ANN) is utilized in this Paper using supervised learning approach. |
| Handwritten Character Recognition in English: A Survey | 2015 | Monica Patel,  Shital Thakkar | This paper presents a comprehensive review of Handwritten Character Recognition (HCR) in English language. The handwritten character recognition has been applied in variety of applications like Banking sectors, Health care industries and many such organizations where handwritten documents are dealt with. |

*Table 2.1 Research papers with their authors and abstract*

1. **Problem Statement**

Writing has been an ancient form of art as well as the main medium for written communication throughout history. It is still prevalent in the society and is considered indigenous to a lot of practices we modern humans do. With the advent of digital media, however, it seems challenging to keep up with writing instead of switching to typing. Typing, although considered as more convenient, we still have a lot of data which is present in hard copies and written form. Not only this, but a lot of sectors involve a mixture of digital and handwritten media to do their job. For example, students and working professionals often need to switch between these 2 media.

A lot of people often face slight inconveniences, when they need a text format of a piece of writing. The only possible way out here is to manually type out the information on a digital medium to use it in the digital form. Our mini project tries to research on the methods which do this efficiently.

For this project, we aim to research about which activation function when used with CRNN for this project gives the best accuracy. We used ‘sigmoid’, ‘selu’, ‘elu’, and ‘relu’ activation functions. We used a dataset from Kaggle, which includes data in the form of words in capital letters and are names and surnames of French students. Due to this, we would be working in the 2nd level of handwritten text recognition, that is, recognizing words.

1. **Objective and Scope**

The objective of this research project is to fix the ML model and algorithm and dataset and change activation functions to get a comparison between these different activation functions. The functions we considered are ‘sigmoid’, ‘selu’, ‘elu’, and ‘relu’.

The dataset we used was sourced from Kaggle and contains transcriptions of 400,000 handwritten names. It is a 1GB data set consisting of images. It has data distributed into train, test and valid sets and contains 3 csv files for labelled data.

This project is to developed on validating handwritten words. So it is on the second level of handwritten text recognition.

1. **Experimental setup**
   1. **Hardware requirements**

1) 8 /16 GB RAM

2) NVIDIA 1050 TI GPU

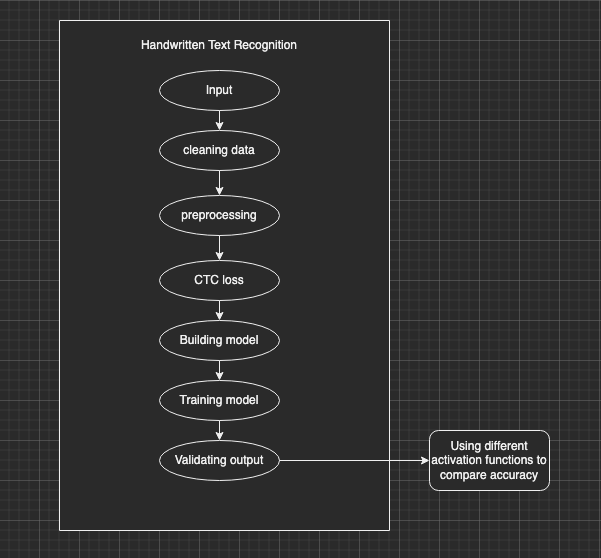
* 1. **Software requirements**

1) Jupyter Notebook

2) VS Code / Jupyter Notebook

3) Google Colaboratory

1. **System design**
   1. **UML Diagram**

****

*Fig 6.1 UML diagram*

* 1. **Algorithm/Process (with Expected input and output)**

1. INPUT

We take input in the form of images.

1. CLEANING DATA

A lot of these images may be null, or unreadable. In this stage, we remove such images.

1. PREPROCESSING

We need to make all the data uniform so that the model can train on it. Here we set the dimensions and crop the image if it is larger, or pad it with white pixels if it is smaller.

1. BUILDING MODEL

Here is where we add the layers and build the model. We built a CRNN model. (Convolution Recurrent Neural Network)

1. TRAINING MODEL

Here we train model on the training set.

1. PREDICTIONS

Based on the model, we predict the result.

**Expected input:**

****

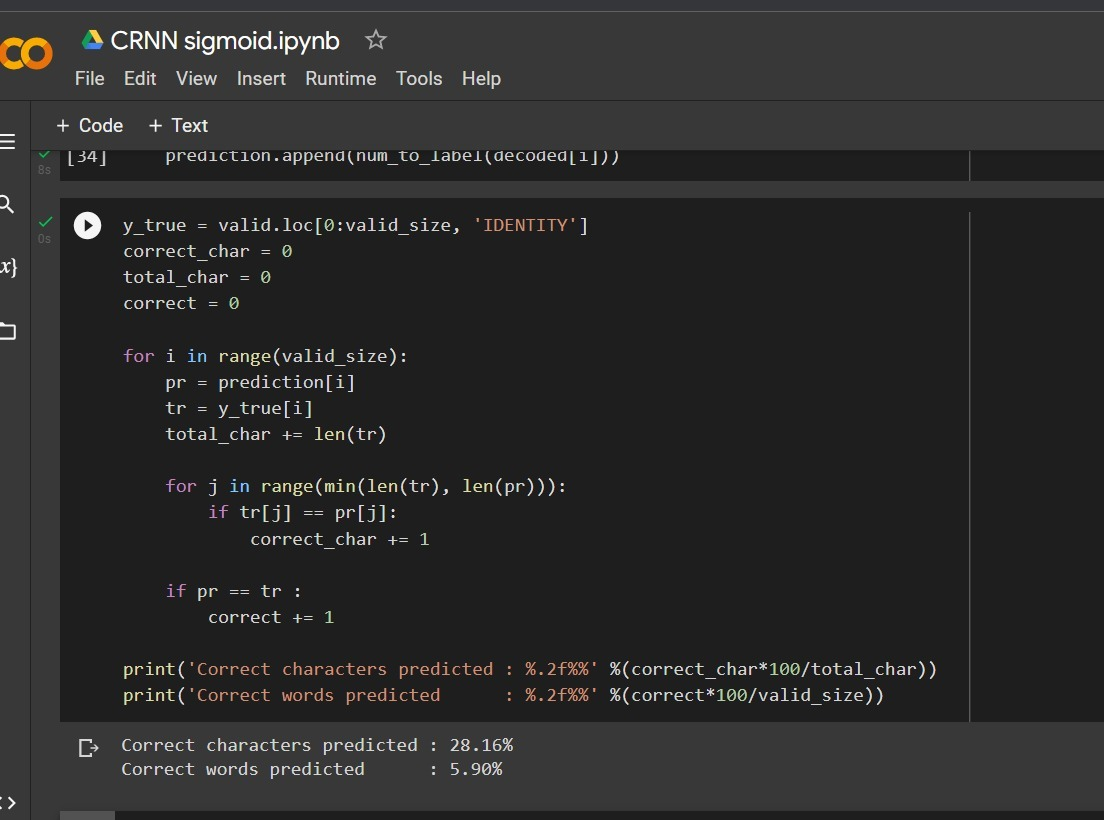
**Expected output:** SIMON

1. **Implementation**
   1. **Code**
2. [Sigmoid](https://colab.research.google.com/drive/1QSHQYjvbRFdMiI-mMoW70kvQxjvKhHdk?authuser=2)
3. [Elu](https://colab.research.google.com/drive/1zKVEyBiQEPsyGgSWB0JzzYdeoiPUhil4?authuser=2#scrollTo=nxs7-fQQev6U)
4. [Selu](https://colab.research.google.com/drive/1AU02VZUz0cGewFU_fQyYUQdYqjtGzp3y?authuser=2)
5. [Relu](https://colab.research.google.com/drive/1WiXzeO0wcijAzhOw_OW-vAs3bQhcZ7CF?authuser=2#scrollTo=q7k8awvLgi2k)
   1. **Screenshots**
6. **Sigmoid Activation Function**

It is a non-linear, continuous, differentiable and monotonic function which outputs between the range of 0 and 1. It accepts real numbers as inputs. It’s a safe function to use, as its output is defined in a bounded interval.

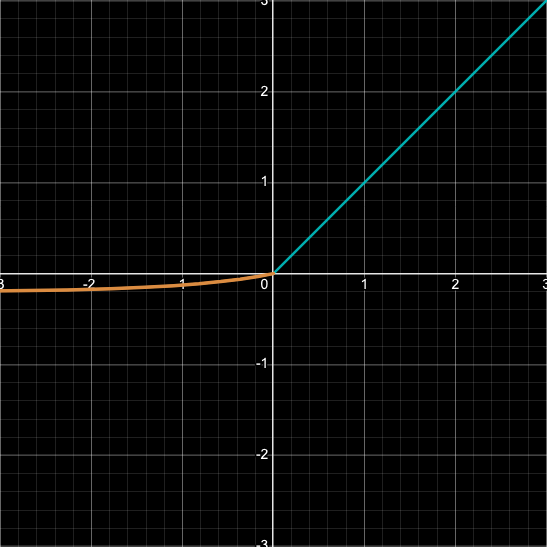


*Fig 7.2.1 Sigmoid Activation Function*

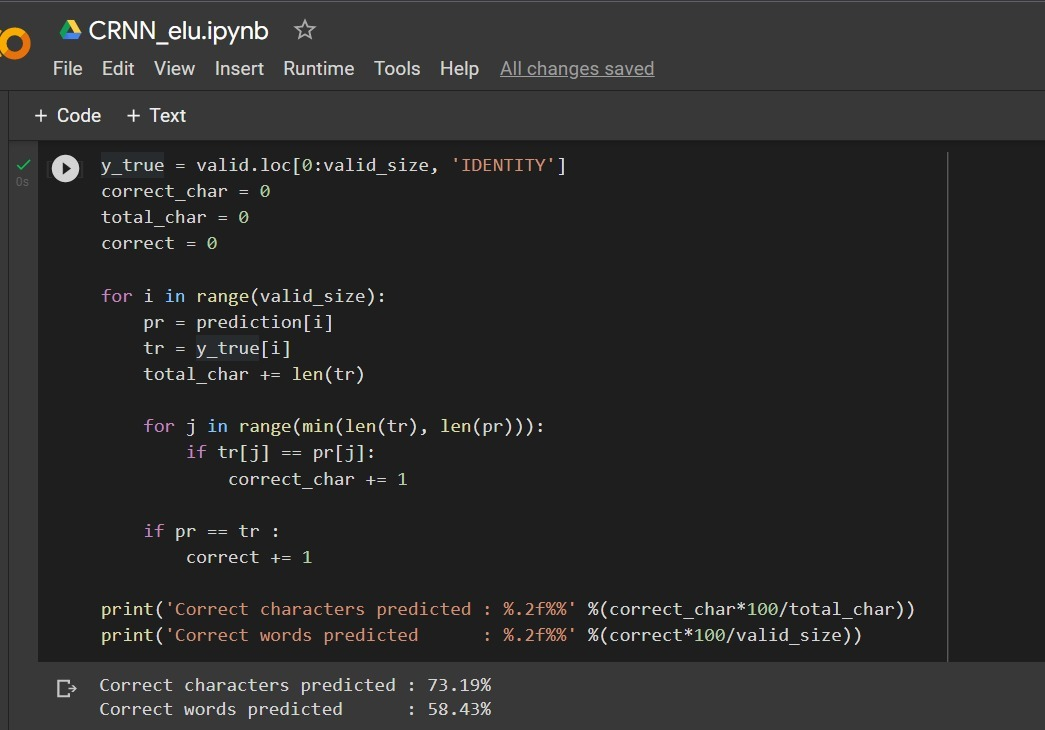
****

1. **Elu Activation Function**

Elu stands for Exponential Linear Unit. It is similar to relu, except for when the input is negative. For this, the value of the constant alpha, generally a value between 0.1 or 0.3 is considered.



*Fig 7.2.2 Elu Activation Function with alpha = 0.2*

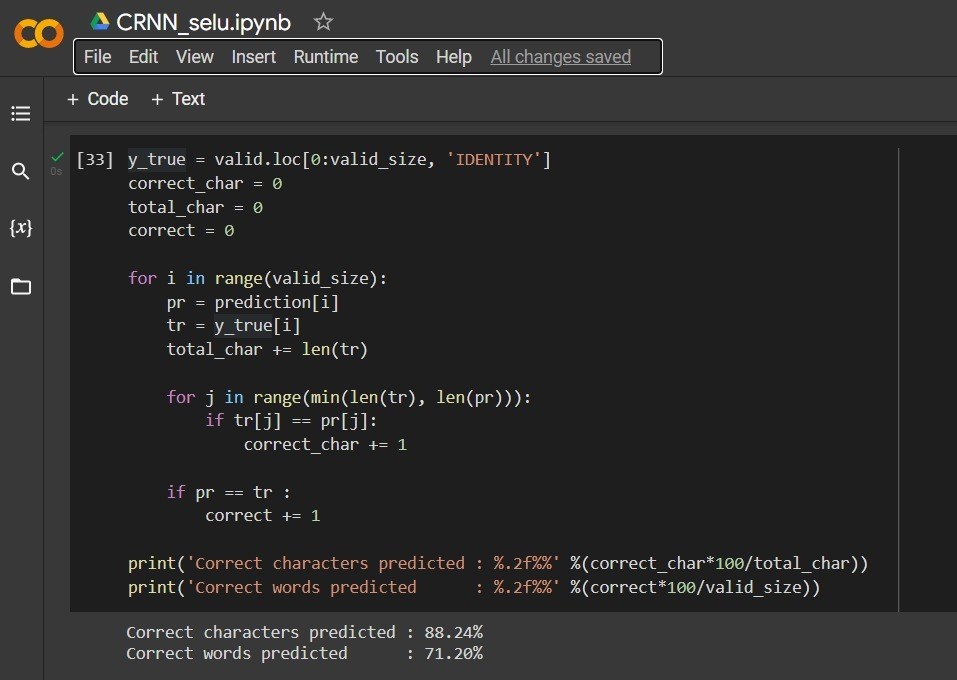


1. **Selu Activation Function**

It stands for scaled exponential linear units. They can self-normalize themselves. The constants used here, lambda and alpha have positive values defined for them. Here we are just assuming the values upto 4th place after the decimal.



*Fig 7.2.3 Selu Activation Function*

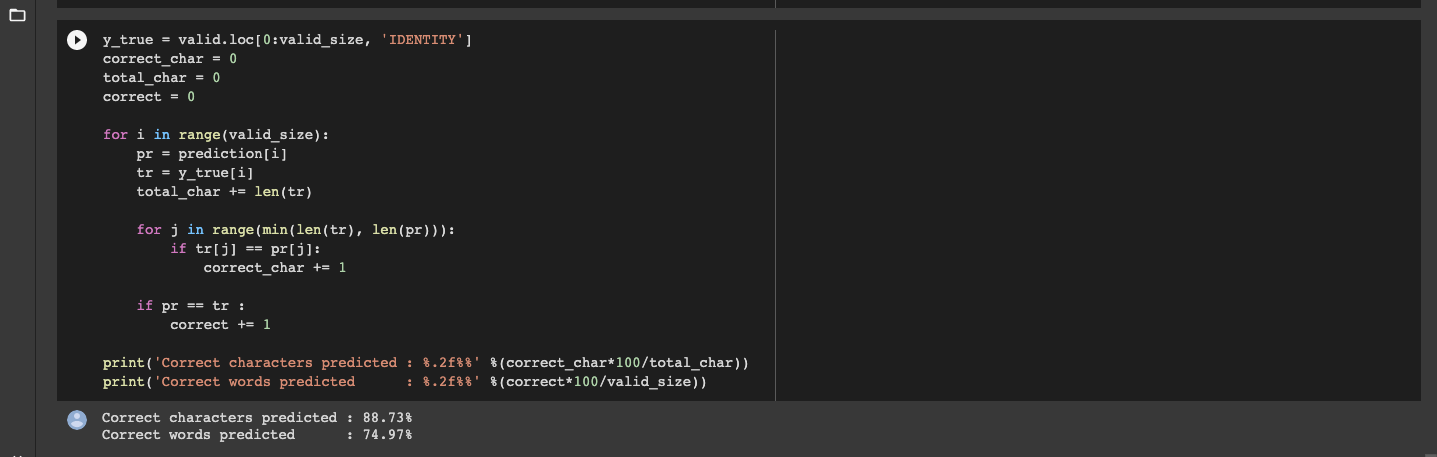


1. **Relu Activation Function**

Relu stands for Rectified Linear Units. It is non-linear, are gives a similar, but a better performance than sigmoid.

**

*Fig 7.2.4 Relu Activation Function*

****

1. **Result**

We thus, implemented for different activation functions on the same dataset and same model. So the constant parameters were the dataset and the model, and the only changing parameter was the activation function. The 4 activation functions were sigmoid, selu, elu and relu.

Sigmoid, Selu, Elu, and Relu all 4 of these activation functions gave us different accuracies which are summarized with the help of a table below.

|  |  |  |
| --- | --- | --- |
| **Sr. No.** | **Activation Function** | **Accuracy** |
| 1) | Sigmoid | 5.90% |
| 2) | Elu | 58.43% |
| 3) | Selu | 71.20% |
| 4) | Relu | 74.97% |

*Table 8.1 Comparitive study of different activation functions and their accuracies*

As demonstrated in this table, Relu has the highest accuracy, i.e. 74.97%, thus surprassing all others. The lowest is that of sigmoid at 5.90%. Elu stands at 58.43% and Selu at 71.20%.

1. **Conclusion**

On the basis of this comparative study, we thus conclude that for a fixed dataset consisting of 4,00,000 transcripted images of uppercase handwriting, and using the CRNN model with CTC loss, we were able to achieve the highest accuracy of 74.97% with the activation function of relu, opposed to three other activation functions - selu, sigmoid and elu.

Thus, relu can be a preferred activation function choice when trying to solve a similar problem under the domain of handwritten text recognition.

**References:**

[1] ﻿ Fathma Siddique, Shadman Sakib, Md. Abu Bakr Siddique, “﻿Recognition of Handwritten Digit using Convolutional Neural Network in Python with Tensorflow and Comparison of Performance for Various Hidden Layers”, ﻿International University of Business Agriculture and Technology, Dhaka 1230, Bangladesh.

[2] Harald Scheidl, 2018, “Handwritten Text Recognition in Historical Documents,” Technische Universität Wien.

1. Jamshed Memon1, Maira Sami2, Rizwan Ahmed Khan3 and Mueen Uddin4, 2020, “﻿Handwritten Optical Character Recognition (OCR): A Comprehensive Systematic Literature Review (SLR)”, 1School of Computing, Quest International University Perak, Ipoh 30250, Malaysia, 2Department of Computer Science, Shaheed Zulfiqar Ali Bhutto Institute of Science and Technology, Karachi 75600, Pakistan, 3Faculty of IT, Barrett Hodgson University, Karachi 74900, Pakistan, 4Department of Software Engineering, Faculty of Science and Technology, Ilma University, Karachi 75190, Pakistan.
2. ﻿Shivangkumar R. Patel, ﻿Ms. Jasmine Jha, “﻿Handwritten Character Recognition using Machine Learning Approach - A Survey”, ﻿ from L.J.I.E.T. PG Department Ahmedabad, Gujarat – India, ﻿presented in International Conference on Electrical, Electronics, Signals, Communication and Optimization (EESCO) – 2015.
3. ﻿Rohan Vaidya, Darshan Trivedi, Sagar Satra, “﻿Handwritten Character Recognition Using Deep-Learning” ﻿Dwarkadas J Sanghvi College of Engineering, Mumbai, India, ﻿Proceedings of the 2nd International Conference on Inventive Communication and Computational Technologies (ICICCT 2018) IEEE Xplore Compliant - Part Number: CFP18BAC-ART; ISBN:978-1-5386-1974-2.
4. ﻿Monica Patel, Shital P. Thakkar, ﻿“ ﻿Handwritten Character Recognition in English: A Survey”, Department of Electronics and Communication, Dharmsinh Desai University, Nadiad, Gujarat, India, ﻿International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 2, February 2015.
5. ﻿ ﻿Ahmed Mahi Obaid1, Hazem M. El Bakry2, M.A. Eldosuky3, A.I. Shehab4, “﻿Handwritten Text Recognition System Based on Neural Network”, Dept. Information Systems1, 2, 4, Dept. Computer Science3 Faculty of Computer Science and Information Systems, Mansoura University1, 2, 3, 4, Mansoura, Egypt, International Journal of Advanced Research in Computer Science & Technology (IJARCST 2016) 72 Vol. 4, Issue 1 (Jan. - Mar. 2016)

**Annexture 1- Project Planning (Using Gantt chart)**

