Language Recognition from Handwriting Based on Machine Learning and Deep Learning

Jayesh Pranav
UG Scholar
Dept of CSE
Sathyabama Institute of
Science and Technology
Chennai, Tamil Nadu, India
nateonthego007123@gmail.com
P.Ajitha
Associate Professor
Dept of IT
Sathyabama Institute of
Science and Technology
Chennai, Tamil Nadu, India
ajithaponnupillai@gmail.com
R. M. Gomathi
Associate Professor
Dept of IT
Sathyabama Institute of
Science and Technology
Chennai, Tamil Nadu, India
gomathi.it@sathyabama.ac.in
A.Sivasangari
Associate Professor
Dept of IT
Sathyabama Institute of
Science and Technology
Chennai, Tamil Nadu, India
sivasangarikavya@gmail.com
Joel Thomas Zachariah
UG Scholar
Dept of CSE
Sathyabama Institute of
Science and Technology
Chennai, Tamil Nadu, India.
joeltz2000@gmail.com
T.Anandhi
Assistant Professor
Dept of CSE
Sathyabama Institute of
Science and Technology
Chennai, Tamil Nadu, India

anandhitamil2908@gmail.com

Article Info
Page Number: 157-167
Publication Issue:
Vol 71 No. 3s2 (2022)

Abstract— In order to develop the understandings of the machinery mind, there has been a lot of upfold in the development of machine learning. As humans, we learn how to do a task by learning it, and optimize the tasks by learning from the mistakes in the process. Just like the brain's neurons automatically trigger and quickly perform learnt tasks, machines can comprehend to the situations of strengthening developed neurons. Deep learning is just as interesting as the human brain's concept. Usage of different types of architectures for such neural networks collide for different types of problems, like image and sound classification, object recognition, image segmentation, object detection, etc. Following these layers of different commemorations and accuracy that Artificial Intelligence provides, one can come up with answers to many unresolved issues, problems, and tasks, since machine is now with the capability of approaching the task as if it is of a human's approach, but with the ability to drive through the task with the machine's work function. The fact that Machine Learning and Deep Learning had taken the industry up by storm is the very reason for its unending usages in this field. The very fabrics of precision and balance that a task completed under the machine's supervision is where we can entrust tasks that can bring a change for many lives with challenges, ailments and difficulties.
Keywords— Virtual Help, Notes Guidance, HTS Converter, Handwriting Analyser, Information Extraction.

Article History
Article Received: 28 April 2022
Revised: 15 May 2022
Accepted: 20 June 2022
Publication: 21 July 2022

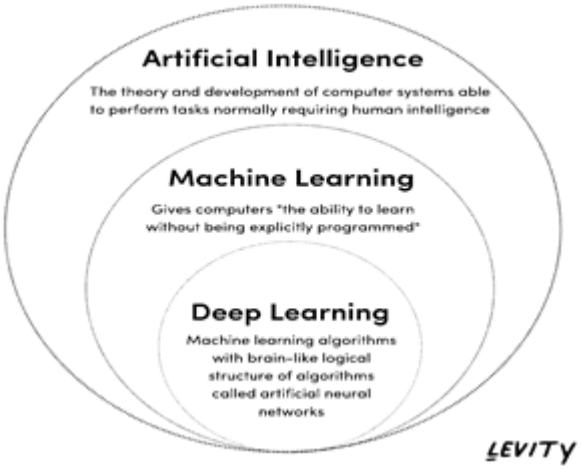
I. INTRODUCTION

Machine Learning has always been a reproductive, dynamic, and excelling approach to artificial intelligence-cum-task reasoning. In 1950, the first test endured by machines to learn themselves, The Turing Test, was considered the apt and satisfactory notion to coin a machine's self-actualization, developing some timed sense of a human cognitive reasoning for a machine. AI being a vast abyss of definitions and modules, its sub-fields are confined to two [1][2]. A figurative way of enhancing the continuations of the writing experience in various applications can be deemed using Natural Language Processing, abbreviated as NLP. It has the age-old subdivision within the classes of separations such as machine translation commonly understood as the language-united translations. Machine translation algorithms defined a significant number of spikes in numerous applications where thorough check of spelling mistakes and grammatical errors were to be corrected. The usage of a different language or a script as whole is totally upon the user/developer who is given the task of developing the language processing model.

II. INITIAL THOUGHTS ON THE PROJECT

A. Selective Momentum of Machine Learning

Since Machine Learning has its niche in a spectre of domains and researches, many approaches have been partially or totally established. Bayesian Network, Clustering, Decision Tree Learning and many more have collectedly made Deep Learning only a part of the approaches [3][4][5]. The following review mainly focuses on the collated usage of deep learning, the basics of forming this proposed application in said and different fields. Additionally, it presents several figures portraying the viable usage of this project across many nuances.



Source: Levity via Google Images

Fig. 1. Layers of Artificial Intelligence

B. The Aim behind this Project

In recent times, many new inventions have been made in order to people with a certain set of disabilities/disorders. We see the development of speech-to-neural transmitter (earpiece) for the deaf, VR-enabled course looker (yet to be developed) for the physically disabled, and so on. One of the main problems that the blind people have to tackle in their daily life is to convert any piece of information that is non-Braille in nature. Since everyone isn't very keen in the aspect of sharing their piece of information in the usable method for the disabled, we seek to find a method that dissembles the correct way of bringing about a valid piece of technology to help this part of community. This search led me to learning and use the knowledge of Deep Learning, ML, and UI/UX, to develop on interface that allows you to bring forth a piece of information (currently in English), either written or typed, to the program as an input, and in return, the machine gives out a voice note of the very information that has been entered, along with numerous help for the set of information provided [6].

III. Understandings of Methods in the Project

We must understand of all the various parameters that has been used for building a perfectly working Machine Learning model. The main parameter to analyse and learn is the types of models and NNs to implement [7]. We know that in order to study data that are in the form of images, audios, or videos, we take the primary help of a Sequential Model and Convolutional Neural Network. We learnt the course of "Machine Learning A-Z™: Hands-On Python & R In Data Science" done via Udemy, where we learnt that for every type of ML model that could be developed and developed new skills to improve our understanding of this challenging field. To get on with the idea, we did use the help of a few research papers from the IEEE website [8], with a few of them present in the Literature Review. Many reviews are presented in literature by many researchers with respect to ecommerce applications in different domain [9][10][11]. This analysis will surely enable the researchers with the idea of deep learning technique in different applications [12][13][14][15]. Different issues also discussed in machine learning applications [16][17][18].The whole concept of bringing forth and developing a model that can decipher the given text into an audio clip can be split into three works – the scanning of the input for words from the manuscript, the identification of the words by the model as a sub-output, and the conversion of the sub-output as the necessary voice

output. In the case of scanning the input, we need a neural network layer that can split the data into parts of necessary schema that needs deciphering. For identification, the help of the Sequential Model along with the guidance of a Recurrent NN is required. For the conversion of the deciphered manuscript as a voice clip, a text-to-speech module is to be implemented that can create an automated voice note of the sub-output clearly.

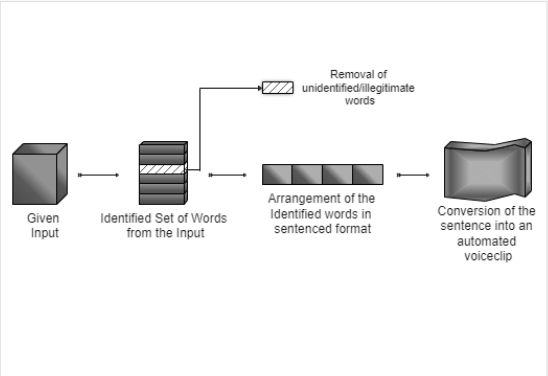


Fig. 2. The Process of Converting Given Manuscript to A Voice Clip

A. Usage of Related Works in the Project

The paper that helped the most to bring this project to its fullest potential was the "Object Recognition Development for Android Mobile Devices with Text-to-Speech Function Created for Visually Impaired People" developed by Andrei Burta, Roland Szabo, and Aurel Gontean. The paper was issued in accordance of developing a speech application that could scan the given picture's input and speak out the name/description of that product. This project was implemented in order to help the old-aged people with identifying some of the essential products while purchasing them in markets. The project basically ran under the guidance of Convolutional NNs and Recurrent NNs. So, we learnt the utmost capabilities of CNN and RNN following the understandings of integrating multi-substantial layers to the model, which was later integrated to a rich and intuitive website UI to be used by people for testing.

B. Selected Interfaces and Methods

- To manifest the application that we dreamt of seeing in action, we decided to implement the base of the application on the cores of the following model – Convolutional Neural Network, that has always been known as the apt Neural Network model to handle and process audio, image, and video clips, Long Short-Term Memory (LSTM) implementation of Recurrent Neural Network (RNN) has proven applicable to compile and transfer information in a larger surface of data layer with better data training, and the Connectionist Temporal Classification (CTC) links with the RNN to compute the loss value. CTC's interference comes up as the task of mining the given array of dataset and finding the apt data node for a certain tested data and give the final text output.

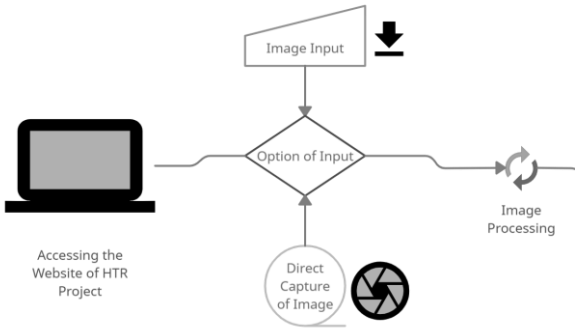


Fig. 3. Accessing the User Image to obtain the Manuscript

- CNN is considered to be one of the core implementations for most of the pattern recognition modules; all the way from voice recognition to image processing. CNNs have the added advantage of controlling the number of Artificial Neural Networks (ANNs) parameters. The failure of impossibility in bringing the core output conversion from the classic ANNs was the reason why the progress of using and propagating models using CNNs were highly appreciated. Obtaining abstract features was another commendable toolwork of CNN, allowing deep layers when the input propagates. Let us say, in the case of image classification, the first sets of layers detect the image's edges, and the second layer then combs through the simpler shapes, and then n-cases of layers for n-such high-level features such as misinterpretable texts, in our case.

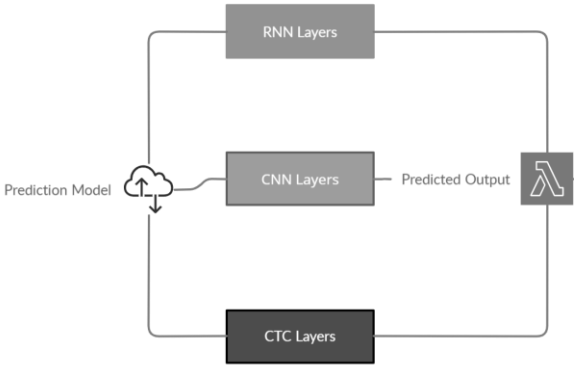


Fig. 4. Layers of the pertained Image-to-Text model

C. Implementation of the Model

As the first block, a Neural Network is built that is completely trained on images of words and sentences from a designated dataset (IAM). Keeping the input layer moderately small for the word/line images, every other layer is compelled with the initial layers, making it feasible for the NN training on the CPU. These are the initial steps to build a base (i.e., a Neural Network) for the conversion that we're aiming in this project.

We use a Neural Network for developing a model for our task. The main layers that are needed in this case are as follows: Convolutional NN layer for depicting the word-line image, Recurrent NN for the feature scaling, and a final layer of CTC (Connectionist Temporal Classification) for length and relevancy domain.

The depiction of the Neural Network is also feasible in a more formal way as a functional equation, mapping a precise image M (considering a matrix) with a length between 0 and L. The necessity of classifying the words on a character level is plausible for the improvement of the model's overall purpose of increasing its accuracy on the "words and sentences" scheme of things. Thus,

$$NN: M \rightarrow (C_1, C_2, \ldots, C_N) \qquad \square\square\square$$

D. Tools/Data Required

- The word "data" is plural, not singular The correct flowing of the project requires the following set of installations and implementations – Ver 3.8 of Python, Ver 1.3 of TensorFlow, and any or latest versions of NumPy and OpenCV.
- These tools are used accordingly in a secure open-source IDE (such as PyCharm in our case) to maintain an ideal and superlative process in the project.
- In order to develop the interface website that collects the data from user and gives the output in form of a voice note, we've employed the services of implementing framework using Bootstrap, and making the interactive UI using HTML & CSS, with the functionality procedures of Js.
- So far, we've thought of either making the model with just a one-time use and destroy interface that uses the input only once to replicate the output rather than saving the whole process. If the latter is in the need, of implementation, we employ either AWS or Firebase for the creation of a nominal database.

IV. Project Implementation

With the implementation of the Python modules in this project, we use about 5 modules that are assigned to the various tasks of the prediction model and text-to-speech conversion. Firstly, we have the "sample pre-processor" module that pre-processes the dataset of images for the model. Which is followed by "data loader" module that reads the obtained samples and loads them in batches to create a sort of iterator-interface to read the whole set of samples, and the actual "model" that predicts the word/sentence from the given input using the samples' pattern. The "T2SConvert" module converts the predicted sentence from the input to an automated voice output for the user as the final product. All these modules are encapsulated into a "main" module for the unanimous functioning of all modules as one program.

TABLE I.

Python Module	**Purpose of Module**		
	Task Withheld	**Input**	**Output**
main.py	Holds all the relevant .py programs	-	-

Python Module	Purpose of Module		
	Task Withheld	**Input**	**Output**
	together		
SamplePreprocessor.py	Accesses the dataset of images from the Neural Network and prepares the data.	DataLoader.py	Model.py
DataLoader.py	Reads all the pre-processed samples given and segregates the common data	main.py	main.py
Model.py	Creates, stores and loads the developing model for the program	-	-
def T2SConvert():	Converts the text module given as an output in the main.py to an automated speech	main.py	WEB SITE

The algorithms used in the project are solely based on the models of CNN, RNN, CTC and LSTM. The flow goes in the manner of deriving the data from the dataset that we use in the program, the IAM Dataset. Following that, we use the pre-processing methods to batch, loss-take and load the

data into the model. The model also has a character error check module which will continually improve the model's accuracy as the error is received via the model running, and keeps the model in its sanity if the model has predicted the input just about right.

Word validations occur predominantly as the pre-processor will get the image from OpenCV and infer the model's run for the specified. This will also check the error ratio and rate for the further simplification and improvement of the model in the future. An inferring function is run alongside the cv2 module to convert the given image in a recognizable manner for the machine (grey-scale) and give the recognized output with the probability percentage of the given phrase's success in the model.

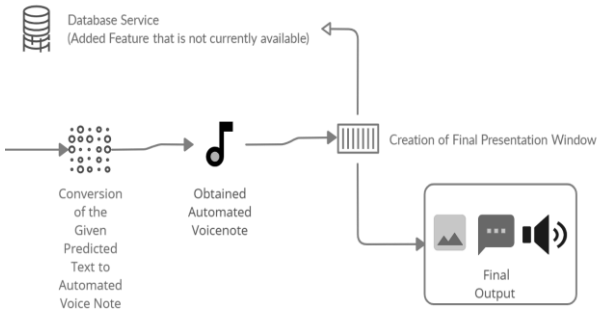


Fig. 5. Conversion process of developing the required

voice note from the predicted phrase

The processed output line is then sent into the next function that converts the given output text phrase into an automated voice note for the user to listen to and gather the information. This concludes the flowchart of the project and defines its full functionality.

A. Working of the Program

As discussed earlier, the usage of CNN and RNN (in form of LSTM) is the primary key role of this project. In order to build the accuracy of the model and to prevent overfitting, we have used the functions of Feature Scaling, Ensembling, k-fold Cross Validation and Regeneralization.

The input is a grey-value image. It is usually either a maximum grey-scale image or a minimalistic B/W, with a size of 128x32. Finding a few images from the dataset out of proportion and size could be common, hence the images are resized which occurs without distortion. The image is later copied to a target image of the same specified size until the image is scaled either between the choices of "128 width" or "32 height". Finally, the normalization of the grey-values of the image is performed, to simplify the Neural Network's performance.

In the case of CNN, the CNN layers fix the output rate length at a sequence of 32. There are 256 features for every entry, which are processed furthermore by the secondary layers of RNN. Even amidst all this, some cases are definitive where high correlation is seen amongst a few sets of handwriting from the input, such as the characters "e" and "l" have a lot in common when it comes in the cursive handwriting, or in the case of duplicate or misspelt characters (classic case of comparing "ll" to "tt").

In the case of the writing of the word “little”, it can be seen that most of the time, the predictions of the images are done in contrast to the placement of that letter in the word, like the letter "i" in comparison to the letter "L" or "t". Since the CTC operation is a position-free entity and seldom goes out of proportion for a handwritten word, the characters “l”, “i”, “t”, “e” and the void label of the text’s classification is a minimalistic task to decode, hence we concatenate the most similar pattern of the words in the sentence to that output line. This creates a best-case scenario for that sentence, thus eliminating all repeated words, letters, and blanks, and with that, we conclude “l---ii--t-t--l-…-e” → “l---i--t-t--l-e” → “little”.

B. Performance Analysis of the Model

Looking at the initial stages of the model’s training aspects, the training data – testing data was split in the ratio of 4:1 (i.e., 80% - 20%). The model was trained by the mean of the loss values of the batch elements. With regards to analysing the preloaded dataset (IAM Sentences version and some self -made dataset), the images of various pre-written sentences is linked and trained along with their actual phrase in the ASCII (or word) format. In the middle of the training process, some broken images are also added to identify the potential outliers in the input of such were present. During the Training Phase, the model is set to find a stable accuracy rate with every count batch it analyses, and the completion of a standard 25 epochs (batch analysis) of the whole training dataset, the accuracy increases up to about 78%.

C. Results and Discussion of the Model Improvement

For improving the recognition accuracy of the model, we followed some of the common regimens – We enabled Data augmentation to be used as to expand the size of the dataset by enabling non-sequential input transformation of the images via Regeneration and Duplication. By increasing the size of the input of the NN layer, if the layer is large enough, we could enable the text-lines completely. We confined the output words to be a proper word from the dictionary is permitted, and if the dictionary doesn’t seem to find the predicted word from the input, then a text correction is done by searching for the most similar one, and the cursive writing style was removed from the input images.

D. Conclusion

We discussed about a Neural Network that was able to recognize text in images, whether handwritten or typed, and converted the text format to a voice note. A Python (.py) implementation using TensorFlow (TF) is provided, along with their roles in the project. Some important parts of the model were introduced separately and inspected for their functions. The final NN model had consisted of 5 Convolutional NN and 2 Recurrent NN layers, which outputs a character-probability matrix. This matrix is either used as a throughput to the model improvement techniques, that were displayed to improve the recognition accuracy of the shown developed model.

ACKNOWLEDGMENT

This paper has been possible by learning the works of the output of the project “Object Recognition Development for Android Mobile Devices with Text-to-Speech Function Created for Visually Impaired People” developed by Andrei Burta, Roland Szabo, and Aurel Gontean. In this paper, their

sole purpose was to enact the task of captioning the shown product to the camera to its nomenclature (name). We had taken the initiative to have changed the approach of reading a product, to reading a script.

REFERENCES

1. Qinge Xiao; Congbo Li; Ying Tang; Xingzheng Chen (2020) contributed to the IEEE Transactions on Automation Science and Engineering on developing an Energy Efficiency modelling that could configure a variably dependent cohesive model development using Machine Learning.
2. Tahir, Ghalib Ahmed; Loo, Chu Kiong (2020) contributed for an open-ended Continual Learning (CL) model that is deemed on Food Recognition by the means of Extreme Learning Machines using the incrementations of class.
3. A thesis on enabling an Ensemble Hierarchical EL (Extreme machine learning) for dereverberation of different speech patterns was conducted by Tassadaq Hussain, Sabato Marco Siniscalchi, Hsiao-Lan Sharon Wang, Yu Tsao, Salerno Valerio Mario, and Wen-Hung Liao via IEEE 2020
4. Fuwei Cui; Qian Cui; and Yongduan Song (2020) made a convoluted survey on Learning-Based approaches for computation and execution of Human-Machine Dialog Systems via IEEE.
5. Hao Tang; Hong Liu; Wei Xiao; and Nicu Sebe (2020) developed a thesis to demonstrate the intuitive connections of Machine Learning and Deep Learning's coding NNs for Image Recognition with Limited Data via IEEE.
6. Guoqiang Zhong; Kang Zhang; and Hongxu; Yuchen Zheng; and Junyu Dong (2019) discussed on the vast sectors of Marginal Deep Architecture, which further followed up on the learning modules and stacking features to develop Deep Learning Models via IEEE
7. Yulia S. Chernyshova, Alexander V. Sheshkus and Vladimir V. Arlazarov came up with the development of a Two-Step CNN Framework that serves the purpose of Text Line Recognition from the images captured via cameras in different places via IEEE.
8. Yucheng Zhou and Zhixian Gao (2019) propelled the idea of executing the Medical Motion Image Recognition with the combined help of Convolutional Neural Network and iOTs (internet of things via IEEE.
9. Kanyadara Saakshara, Kandula Pranathi, R.M. Gomathi, A. Sivasangari, P. Ajitha, T. Anandhi, "Speaker Recognition System using Gaussian Mixture Model", 2020 International Conference on Communication and Signal Processing (ICCSP), pp.1041-1044, July 28 - 30, 2020.
10. R. M. Gomathi, P. Ajitha, G. H. S. Krishna and I. H. Pranay, "Restaurant Recommendation System for User Preference and Services Based on Rating and Amenities," 2019 International Conference on Computational Intelligence in Data Science (ICCIDS), 2019, pp. 1-6, doi: 10.1109/ICCIDS.2019.8862048.
11. Subhashini R , Milani V, "IMPLEMENTING GEOGRAPHICAL INFORMATION SYSTEM TO PROVIDE EVIDENT SUPPORRT FOR CRIME ANALYSIS", Procedia Computer Science, 2015, 48(C), pp. 537–540
12. Harish P, Subhashini R, Priya K, "Intruder detection by extracting semantic content from surveillance videos", Proceeding of the IEEE International Conference on Green Computing, Communication and Electrical Engineering, ICGCCEE 2014, 2014, 6922469

13. Sivasangari, A., Krishna Reddy, B.J., Kiran, A., Ajitha, P.(2020), " Diagnosis of liver disease using machine learning models", ISMAC 2020, 2020, pp. 627–630, 9243375
14. Sivasangari, A., Nivetha, S., Pavithra,, Ajitha, P., Gomathi, R.M. (2020)," Indian Traffic Sign Board Recognition and Driver Alert System Using CNN", 4th International Conference on Computer, Communication and Signal Processing, ICCCSP 2020, 2020, 9315260
15. Ajitha, P., Lavanya Chowdary, J., Joshika, K., Sivasangari, A., Gomathi, R.M., "Third Vision for Women Using Deep Learning Techniques", 4th International Conference on Computer, Communication and Signal Processing, ICCCSP 2020, 2020, 9315196
16. Ajitha, P.Sivasangari, A.Gomathi, R.M.Indira, K."Prediction of customer plan using churn analysis for telecom industry",Recent Advances in Computer Science and Communications,Volume 13, Issue 5, 2020, Pages 926-929.
17. Gowri, S. and Divya, G., 2015, February. Automation of garden tools monitored using mobile application. In International Confernce on Innovation Information in Computing Technologies (pp. 1-6). IEEE.
18. Gowri, S., and J. Jabez. "Novel Methodology of Data Management in Ad Hoc Network Formulated Using Nanosensors for Detection of Industrial Pollutants." In International Conference on Computational Intelligence, Communications, and Business Analytics, pp. 206-216. Springer, Singapore, 2017.