A Critical Review on Next Word Prediction

Kolambi Narula
Department of Computer Science, Mithibai College, Mumbai, Maharashtra, India
University of Mumbai, Mumbai, Maharashtra, India
9112kolambiinarula@gmail.com

Abstract: *One area of natural language processing that can assist with next word prediction is next word prediction, which is also known as language modeling. It is one application of machine learning. Previous researchers had discussed it using other models, including federated text models and recurrent neural networks. To make the prediction, each researcher utilized their own model. LSTMs use a built-in mechanism for selectively choosing which information to preserve in the hidden state and which to discard, which helps to prevent overfitting. Activation functions, such as ReLU and softmax, are also used in next word prediction to introduce non-linearity into the model and generate a probability distribution over the vocabulary for predicting the next word. Combining techniques such as pre-training, advanced architectures, and large datasets can further improve the performance of next word prediction using LSTM. This paper summarizes the different approaches taken to achieve the above-stated aim. The primary focus is on the next word prediction to get the best result. It chose to make the model using Long Short Term Memory (LSTM), a sequential activation model, and a 2000 epoch for the training. For the libraries, I have used TensorFlow, pickle, Keras, NumPy, and OS This model can be used to predict the next word.*

Keywords: Language Modeling (LM), Natural Language Processing (NLP), Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), Neural Language Model (NLM), Natural Language Generation (NLG), Next Word Prediction (NWP), Coupled Input and Forget Gate (CIFG), Federated Learning (FL)

I. INTRODUCTION

Next-word prediction is a cutting-edge NLP technique that leverages deep learning to predict the most likely next word in a sentence based on its context. The development of models such as RNNs, transformer networks, and language models has enabled next word prediction to achieve high accuracy. These models are trained on massive amounts of text data and use statistical language models and complex neural network architectures to capture the relationships between words and their context. With the continuous advancements in NLP and deep learning, next word prediction will become an increasingly important component in the development of intelligent systems that can understand and generate human language. It also involves the use of various techniques, such as n-gram models, Markov models, and semantic analysis to predict the next word. These models consider the probabilities of words appearing in different contexts and use them to generate predictions. Additionally, advanced models also incorporate contextual information, such as the surrounding words and sentence structure, to make predictions. The use of next word prediction can improve user experience in NLP applications, as it allows for faster and more accurate text entry. It also has potential applications in areas such as content generation, question answering, and personalized recommendation systems. As NLP and deep learning continue to advance, next word prediction is poised to play a crucial role in the development of more intelligent and human-like language systems.

There are different techniques used for next word prediction, including:

- N-gram models, which predict the next word based on the previous n-1 words
- Recurrent Neural Networks (RNNs) which are a type of neural network that can process sequential data. They are trained to predict the next word based on the previous words in the sequence.
- Long Short-Term Memory (LSTM) networks, which are a type of RNN that can better handle long-term dependencies in the data.

The models are typically trained on a large corpus of text data and then fine-tuned with specific data for a particular application. Once trained, the model can be used to generate predictions for new sentences or phrases, and can also be used to generate new text by sampling from the probability distribution of words predicted by the model.

II. LITERATURE REVIEW

[2.1] A likelihood distribution over word sequences may serve as an application of mathematics language models. When given a sequence of this type, let's say one of length m, the language model employs context to distinguish between terms and sentences that sound similar and assigns a likelihood P(w1,w2… Wm) to the entire series. For instance, in English, the words "recognize speech" and "destroy a nice beach" seem similar but mean very different things. A challenge at the heart of both language understanding and natural language processing is known as language modeling (LM). In addition to writing in code linguistic complexity like grammatical structure, models that can effectively put distributions over sentences also distill a significant amount of information about the knowledge that some corpora may contain. Since they enabled researchers to investigate a number of tasks that the robust conditional independence assumptions area fail to achieve, deep learning and recurrent neural networks (RNNs) have significantly advanced language modeling analysis. Even while simpler models, such as N-grams, only employ a limited history of the previous words to predict the next word, they are nevertheless an important component of high-quality, low mental confusion models. Models of language. Indeed, the most recent research on large-scale language models has demonstrated that RNNs function well in conjunction with N-grams because they each have unique characteristics that enhance N-gram models but exhibit the weakest performance when considered alone.

Information can persist in recurrent neural networks because they contain networks with loops in them.

[2.2] This technique uses letter-to-letter prediction, which means it anticipates a character to build a word, and as LSTM stands for Long Short Time Memory, it will comprehend the previous text and forecast the words that may be helpful for the user to frame phrases. It will aid end-users to outline significant paragraph components and enable users to concentrate on the subject rather than wasting time on what to type next because composing an essay and structuring a lengthy paragraph take time. They anticipate employing LSTM to develop or imitate auto-complete functionalities. We will be experimenting with this problem using LSTM by using the Default Nietzsche text file, also called as our training data, to train a model. Most software employs various methods, including NLP and standard neural networks, to achieve this task. Predicting the word that will follow next is a task known as next word prediction or language modelling. It is one of the core duties of NLP and has numerous applications.

[2.3] The language model (LM) of a recognition system is the essential part of automatic voice recognition and contains the syntactic and semantic restrictions of a given natural language. While today's recognition pass mostly uses backing-off models, the restoring step has benefited greatly with the introduction of feed-forward neural network LMs. Even with feed-forward neural network LMs, the n-gram assumption causes the modeling to be inaccurate.

[2.4] In today's real-time social media environment, mass electronic dialogue and communication is a typical occurrence. For informal talks, the native language is typically used in a transcripted version.

The biggest issue with word prediction in any regional language is that some of our machines only comprehend ASCII values. ASCII is the most widely used text file format on computers and the Internet, and it uses a 7-bit binary representation for each alphabetic, numeric, or special character. Some Assamese characters are not properly displayed on the screen when we use Unicode for the Assamese language while implementing on platforms like Python. In order to train our model, we have phonetically transcribed a set of Assamese words using the Google Indic keyboard and recorded the results in a text file.

[2.5]In India, Hindi is a widely used language. Thus, the vast majority of Indians will find next word prediction in Hindi to be useful. The sheer number of mantras and symbols in Hindi makes it extremely challenging to process. As a result of the spelling issue, this may produce irrelevant results. As a result, word-level language processing produces better outcomes and requires less effort. It is possible to create meaningful writing that is unintelligible to humans using Natural Language Generation (NLG), a methodical and important methodology. Data is gathered for the purpose of creating the text from a variety of sources or by taking user input. The field of NLP has seen a significant transformation recently.

Prior to now, shallow machine learning models, which required a lot of manual labor and had handcrafted features, were used in NLP techniques. Neural networks have outperformed more conventional machine learning models in terms of success because word embeddings are becoming more and more common.

In many parts of the world, English is widely spoken. Since text is generated at the character and word levels, researchers have used a variety of statistical and machine learning methods.Word2vec technique, a continuous bag of words, and other statistical models that assist in text production by establishing a vocabulary have all been utilized for English-language text generation. These methods have the drawback of not guaranteeing that the text they produce is syntactically valid.

[2.6]How to design computers that develop automatically with use is an issue that is addressed by machine learning. By instructing the machines to use the created model, experience can be gained in this area. The system will learn by pattern recognition thanks to that training. The concept of machine learning is distinct from traditional programming, since it makes use of a new paradigm for programming. Rules and data are the inputs in conventional programming, and the result is the solution. Machine learning, on the other hand, uses data and responses as inputs and produces rules as the final product.

Deep learning is a subset of machine learning that determines classification and prediction using several layers of neural networks. Uses for categorization include things like spam detection, cancer prediction, sentiment analysis, and more. Another instance of it that we utilize on a regular basis but may not be aware of is when we type. Nearly every day, whether we're merely browsing the internet or sending messages, we utilize devices like our smartphones and computers, and we undoubtedly also use them for typing. We might notice that it suggests the following term based on what we type while performing those tasks. The word that will likely follow our texts is predicted in this section, which is why it is referred to as next word prediction. We wrote more quickly and efficiently as a result, which was helpful.

Natural language processing (NLP) faces a serious challenge with next word prediction (NWP). The process of mining text is also known as language modelling. It has been utilized by many researchers and programmers alike.

[2.7]Our lives now revolve around using a personal predictive text input system, generally known as T9. For those who have physical restrictions, therapists frequently suggest word prediction as a way to increase typing speed. Even while some studies imply that word suggestions affect writing abilities, increased speed could not be cited as one of its benefits when utilizing a normal keyboard. Users should not review predicted words while typing; doing so might cause word prediction to fail, since it would slow down the user's typing speed more than the word suggestion would. In order to determine if word prediction software would increase typing speed when a user needed to glance away from this paper, a study was conducted. The results of the trial indicate that word prediction increased typing speed for seven out of 10 individuals. The outcome proposes that users may find word prediction useful. However, while creating the next term, mistakes are frequently seen. This research aims to analyze current next-word prediction techniques based on input text and evaluate them in a Ukrainian language context. Lack of Ukrainian language corpora and models supporting Ukrainian language might be a barrier (for example, GPT, BERT, and GPT2 for English). The goal of the effort is to create a predictive system for text input in Ukrainian that should increase the following word's accuracy.

[2.8]Grasp Long Short-Term Memory Recurrent Neural Networks (LSTM-RNN) and its development since the early 1990s can help the interested reader gain a deeper understanding of the subject. A significantly different notation and a much more condensed formulation of the derivations are used in today's articles on LSTM-RNN. Machine learning focuses on creating algorithms that automatically get better with use. The learning algorithm should ideally get better the more it is used. The learning algorithm's job is to build a classifier function from the provided training data. The effectiveness of this built-in classifier is then evaluated by using data that had never been seen before. Artificial neural networks (ANN) loosely mimic the fundamental operations of organic learning systems. Feed-forward neural networks are the most prevalent form of conventional neural network. One input layer, one output layer, and at least one intermediate hidden layer are used to organize the sets of neurons in this system. Static classification tasks are the only ones that feed-forward neural networks can handle. They can only offer a static mapping between input and output as a result.

[2.9]With several deep learning (DL) architectures now being created, deep neural networks have been effectively used in a variety of developing sectors to address complicated real-world challenges. The DL architectures employ activation functions (AFs) to carry out various computations between the hidden layers and the output layers of any specific DL architecture in order to attain these cutting-edge results. This study provides an overview of the current activation functions (AFs) utilized in deep learning applications and emphasizes recent developments in the field. This paper's originality comes from its compilation of the bulk of the AFs used in DL and its discussion of current trends in their utilization in real-world deep learning deployments in comparison to cutting-edge research findings. The fact that this study will be the first to compare trends in AF applications in practice to research findings from literature reported in deep learning research to date makes it topical given that most research papers on AF emphasize comparable works and outcomes.

[2.10]As of 2019, G board, also referred to as the Google keyboard, is a virtual keyboard for touchscreen mobile devices that handles more than 600 different languages. G board offers autocorrect feature, word completion, then next prediction functions in addition to processing noisy signals from input modalities including tapping and word-gesture typing. Reliable and quick mobile input techniques are more crucial as consumers progressively switch to mobile devices. Next-word predictions are a technology that makes text entering easier. Language models (LMs) can forecast the most likely next word or phrase based on a tiny quantity of user-generated previous text. Predictions were produced before using a word n-gram finite state transducer (FST). Created using a finite state transducer (FST) for word n-grams.

III. CONCEPTS

In this paper, they have used Language Modeling, Natural Language Processing, Recurrent Neural Network, Long Short-Term Memory, Neural Language Model, Bi-LSTM, activation function, Federated learning.

3.1 Language Modeling

Language modeling is the task of predicting the next word in a sequence of words given some context, based on a large corpus of text data. It is a fundamental task in NLP and used for various applications such as speech recognition, machine translation, text generation, and more. It involves training a model to understand the patterns and dependencies between words, and using that knowledge to generate coherent and contextually relevant sequences of text.

3.2 Natural Language Processing

Natural Language Processing (NLP) is a field of computer science and artificial intelligence that deals with the interaction between computers and human languages. It involves the development of algorithms and models to enable computers to understand, interpret, and generate human language. NLP covers a wide range of tasks such as sentiment analysis, machine translation, named entity recognition, text classification, and more. The goal of NLP is to bridge the gap between human communication and computer understanding, making it possible for machines to process, analyze, and generate human-like text.

3.3 Recurrent Neural Network (RNN)

It is a type of deep learning model used for processing sequential data, such as time series, text, and speech. Unlike traditional neural networks, which are feedforward networks, RNNs have a memory mechanism that allows them to keep track of previous inputs and use that information in the current prediction. The network uses the same set of weights for all inputs, making it possible to process sequences of varying lengths. RNNs are widely used for natural language processing tasks such as language translation, text generation, and sentiment analysis. The Long Short-Term Memory (LSTM) is a popular variant of RNN, designed to overcome the vanishing gradient problem that can occur in traditional RNNs.

3.4 Long Short-Term Memory (LSTM)

It is a type of Recurrent Neural Network (RNN) designed to address the issue of vanishing gradients that can occur in traditional RNNs when processing long sequences of data. LSTMs have a memory mechanism that allows them to store information over a longer period and selectively overwrite it as needed, making them well-suited for processing sequences with complex dependencies. An LSTM cell has three gates—input, forget, and output gates—that control the flow of information into and out of the cell, allowing the network to preserve important information and discard irrelevant information. LSTMs have been widely used for a variety of NLP tasks such as language translation, text classification, and speech recognition.

3.5 Neural Language Model (NLM)

It is a type of deep learning model that predicts the probability distribution of the next word in a sequence of words, given the previous words as context. It is trained on a large corpus of text data and learns to generate coherent and semantically meaningful sequences of words. NLMs can be implemented using various deep learning architectures, such as Recurrent Neural Networks (RNNs), Convolutional Neural Networks (CNNs), and Transformer-based models. The goal of an NLM is to model the probability distribution of sequences of words in a language and use that knowledge to generate new sequences of text that are contextually relevant and semantically meaningful. NLMs have found applications in various NLP tasks such as machine translation, text generation, and text classification.

3.6 Bi-LSTM

It stands for Bidirectional Long Short-Term Memory. It is a type of recurrent neural network (RNN) architecture commonly used for tasks such as language modeling and sequential classification. In a Bi-LSTM, the input sequence is processed in both forward and backward direction, allowing the model to capture contextual information from both past and future. The outputs from the forward and backward LSTMs are concatenated and fed into the next layer, providing a rich representation of the input sequence.

3.7 Activation Function

An activation function is a mathematical operation applied to the output of each neuron in a neural network. Its purpose is to introduce non-linearity into the output of the neuron, allowing the neural network to learn and represent complex relationships between inputs and outputs. Some common activation functions include: [1]Sigmoid: maps any input value to the range of 0 and 1, representing a probability. [2]Tanh: maps input values to the range of -1 and 1. [3]ReLU (Rectified Linear Unit): replaces negative values with 0 and retains positive values. [4]Softmax: maps input values to a probability distribution over several classes, used as the activation function in the output layer for multi-class classification problems. The choice of activation function depends on the specific task and type of neural network being used.

3.8 N-gram models

They are a type of statistical language modeling that predict the next word in a sequence based on the preceding N-1 words. N can be any positive integer, but most commonly, N is set to 2 (bigrams) or 3 (trigrams). N-gram models consider the context of the previous words in the sequence to generate the next word. [1]Unigram Model: The simplest type of n-gram model, an unigram model only considers the current word without considering any context. [2]Bigram Model: A bigram model considers the current word and the preceding word to make predictions. [3]Trigram Model: A trigram model considers the current word and the preceding two words to make predictions. N-gram models are commonly used in natural language processing tasks such as speech recognition, machine translation, and text classification. They are also used as building blocks for more advanced language models, such as neural language models.

3.9 Federated Learning

It is a machine learning technique where models are trained on decentralized data sources, such as mobile devices or edge devices, without exchanging the underlying data. The models are trained locally on individual devices, and then the updated model parameters are aggregated in a centralized location to update the global model. This allows for privacy-preserving training on sensitive data while still enabling machine learning models to be trained at scale.

3.10 Coupled Input and Forget Gate (CIFG)

Coupled input and forget gates are a modification to the standard LSTM architecture in which the input and forget gates are combined into a single gate. This helps to reduce the number of parameters and computations in the LSTM, making it more computationally efficient. This architecture has been found to be effective in some applications, but its impact on performance can vary depending on the task and data being used.

IV. RELATED WORKS

There have been many review papers on Long Short Term Memory in general and Recurrent Neural Networks(RNN) in particular. The details about different review papers are tabulated below in Table.1.

S. No	Paper	Year	Next Word Prediction	Authors
1.	Predicting next Word using RNN and LSTM cells: Statistical Language Modeling	2019	Language Model	Aejaz Farooq Ganai,Farida Khursheed
2.	Next Words Prediction Using Recurrent Neural Networks	2021	Recurrent Neural Networks(RNN)	Sourabh Ambulgekar, Sanket Malewadikar, Raju Garande, and Dr. Bharti Joshi
3.	LSTM Neural Networks for Language Modeling	2012	Long Short-Term Memory (LSTM)	Martin Sundermeyer, Ralf Schlüter, and Hermann Ney
4.	A RNN based Approach for next word prediction in Assamese Phonetic Transcription	2018	Long Short-Term Memory (LSTM)	Partha Pratim Barman , Abhijit Boruahaa
5.	Next Word Prediction in Hindi Using Deep Learning Techniques	2019	Long Short-Term Memory (LSTM), Bidirectional Long Short-Term Memory (Bi-LSTM)	Radhika Sharma, Nishtha Gael , Nishita Aggarwal, Prajyot Kaur and Chandra Prakash
6.	NEXT WORD PREDICTION USING LSTM	2022	Long Short Term Memory (LSTM)	Afika Rianti , Suprih Widodo, Atikah Dhani Ayuningtyas, Fadlan Bima Hermawan
7.	An Approach for a Next-Word Prediction for Ukrainian Language	2021	Long Short Term Memory (LSTM) , Markov chains	Khrystyna Shakhovska ,Iryna Dumyn ,Natalia Kryvinska , and Mohan Krishna Kagita
8.	Understanding LSTM –a tutorial into Long Short-Term Memory Recurrent Neural Networks	2019	Long Short-Term Memory Recurrent Neural Networks (LSTM-RNN)	Ralf C. Staudemeyer, Eric Rothstein Morris
9.	Activation Functions: Comparison of Trends in Practice and Research for Deep Learning	2018	Activation Functions (AFs)	Chigozie Enyinna Nwankpa, Winifred Ijomah, Anthony Gachagan, and Stephen Marshall
10.	FEDERATED LEARNING FOR MOBILE KEYBOARD PREDICTION	2019	Finite state transducer (FST) , Coupled Input and Forget Gate (CIFG), Recurrent Neural Networks (RNN), Long Short Term Memory (LSTM)	Andrew Hard, Kanishka Rao, Rajiv Mathews, Swaroop Ramaswamy, Francoise Beaufays Sean Augenstein, Hubert Eichner, Chloe Kiddon, Daniel Ramage

Table 1. Different Review Works

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 1, February 2023
Impact Factor: 7.301

V. COMPARISON OF VARIOUS ASPECTS OF ALGORITHMS

5.1 Datasets Review

S. no	Model	Dataset Used
1	Language Model	Project Gutenberg from ebook author Franz Kafka
2	Recurrent Neural Networks(RNN)	Project Gutenberg from ebook author Franz Kafka
3	Long Short-Term Memory (LSTM)	The Quaero2 project's French corpora and the English Treebank-3 Corpus
4	Long Short-Term Memory (LSTM)	Dataset taken from the Assamese author Padmanath Gohainbaruah's book Bhanumati.
5	Long Short-Term Memory (LSTM) ,Bidirectional Long Short-Term Memory(Bi-LSTM)	This dataset was created by IIT Bombay at the Indian Language Technology Center.
6	Long Short Term Memory (LSTM)	Indonesian destinations
7	Long Short Term Memory (LSTM) , Markov chains	Ukrainian poems

Table 2. Datasets used in the Models

5.2 Performance Accuracy

S. No	Model	Accuracy of performance
1	Language Model	46%
2	Recurrent Neural Networks(RNN)	For 10 input nodes, the training accuracy ranges from 54% to 55%, while the testing accuracy ranges from 56% to 54%.
3	Long Short-Term Memory (LSTM)	Experiments show that typical recurrent neural network architectures can perform 8% better in terms of confusion.
4	Long Short-Term Memory (LSTM)	The model's accuracy rate is 88.20% when it has a top layer, each with 128 neurons, a learning rate of 0.001, and 98000 iterations.
5	Long Short-Term Memory (LSTM) ,Bidirectional Long Short-Term Memory(Bi-LSTM)	The derived LSTM model's accuracy is 70.89%, while the Bi-LSTM model's accuracy is 79.54%. But their respective validation accuracy rates are 59.46% and 81.07%.
6	Long Short Term Memory (LSTM) , Markov chains	an accuracy of 75%

Table 3. Accuracy used in the Model

VI. CHALLENGES

There are several challenges associated with next word prediction:

6.1. Contextual understanding: Next word prediction models must accurately understand the context and meaning of a sentence in order to make accurate predictions. This can be challenging, especially when dealing with idiomatic expressions, sarcasm, and other forms of figurative language.

6.2. Data diversity: Next word prediction models are typically trained on large corpora of text, which may not represent the diverse range of writing styles, languages, and cultures in the world. This can result in models that are not well-suited for certain populations or types of text.

6.3. Personalization: While personalization of next word prediction models can improve accuracy, it also requires large amounts of data from individual users, which may not always be available or feasible to collect.

6.4. Real-time performance: Next word prediction must be fast and responsive, as users expect a seamless and uninterrupted typing experience. This can be a challenge for models that require significant computational resources or latency.

6.5. Privacy and security: When collecting and using data to train and personalize next word prediction models, there are important considerations around privacy and security, as users' personal information and writing may be sensitive. These challenges must be overcome in order for next word prediction to continue to advance and be widely adopted in various applications.

VII. CONCLUSION AND FUTURE

In conclusion, next word prediction is a fundamental task in natural language processing that has a wide range of applications in various fields such as speech recognition, machine translation, text generation, and human-computer interaction. The advancements in deep learning have led to the development of powerful language models that have greatly improved the performance of next word prediction. However, despite these advancements, there are still several challenges that need to be addressed in order to further improve the performance of next word prediction models. These challenges include handling out-of-vocabulary words, idiomatic expressions, ambiguity, different writing styles, long-term dependencies, low-resource languages and missing data among others. Approaches like using LSTMs and activation functions have shown to be effective in overcoming some of these challenges, but the field of natural language processing is constantly evolving, and new techniques and approaches are being developed to further improve the performance of next word prediction models. The advancements in this field will continue to pave the way for more natural and efficient human-computer interaction and improve the performance of a wide range of natural language processing tasks. In summary, next word prediction is a complex task that is constantly evolving, and the field is continuously making progress in addressing the challenges that come with it. The developments in this field will continue to pave the way for more natural and efficient human-computer interaction and improve the performance of a wide range of natural language processing tasks. In our future of next, word prediction is likely to involve continued advancements in natural language processing and machine learning techniques, resulting in more sophisticated and accurate models. This could potentially lead to a wider adoption of next word prediction in various applications, such as typing assistance on smartphones and computers, improved language translation, and personalization in virtual assistants.

REFERENCES

[1]. Aejaz Farooq Ganai,Farida Khursheed ,2019,” Predicting next Word using RNN and LSTM cells:Statistical Language Modeling”
[2]. Sourabh Ambulgekar, Sanket Malewadikar, Raju Garande, and Dr. Bharti Joshi,2021,” Next Words Prediction Using Recurrent Neural Networks”
[3]. Martin Sundermeyer, Ralf Schlüter, and Hermann Ney,2012,” LSTM Neural Networks for Language Modeling”
[4]. Partha Pratim Barman , Abhijit Boruahaa,2018.” A RNN based Approach for next word prediction in Assamese Phonetic Transcription ”
[5]. Radhika Sharma, Nishtha Gael , Nishita Aggarwal, Prajyot Kaur and Chandra Prakash,2019,”Next Word Prediction in Hindi Using Deep Learning Techniques”
[6]. Afika Rianti , Suprih Widodo, Atikah Dhani Ayuningtyas, Fadlan Bima Hermawan,2022,” NEXT WORD

PREDICTION USING LSTM”

[7]. Khrystyna Shakhovska ,Iryna Dumyn ,Natalia Kryvinska ,and Mohan Krishna Kagita,2021,” An Approach for a Next-Word Prediction for Ukrainian Language”

[8]. Ralf C. Staudemeyer, Eric Rothstein Morris,2019,” Understanding LSTM –a tutorial into Long Short-Term Memory Recurrent Neural Networks”

[9]. Chigozie Enyinna Nwankpa, Winifred Ijomah, Anthony Gachagan, and Stephen Marshall,2018,” Activation Functions: Comparison of Trends in Practice and Research for Deep Learning ”

[10]. Andrew Hard, Kanishka Rao, Rajiv Mathews, Swaroop Ramaswamy, Francoise Beaufays Sean Augenstein, Hubert Eichner, Chloe Kiddon, Daniel Ramage,2019,”FEDERATED LEARNING FOR MOBILE KEYBOARD PREDICTION”