



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

# IJIRSET

International Journal of Innovative Research in  
**SCIENCE | ENGINEERING | TECHNOLOGY**



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 11, Issue 5, May 2022

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

Impact Factor: 8.118

# Genome Sequence Classification using LSTM

**Sushmita Kulkarni, Geeta Thorat, Vishakha Patil, Srushti Pawar, Prof. Kanchan S Warke**

BE Students, Department of Computer Engineering, Bharati Vidyapeeth's college of Engineering for Women,  
Pune, India

Assistant Professor, Department of Computer Engineering, Bharati Vidyapeeth's college of Engineering for Women,  
Pune, India

**ABSTRACT:** The Corona virus which originated in Wuhan, China has since spread throughout the world and is affecting millions of people. When there is a novel virus outbreak, it is crucial to quickly determine if the epidemic is a result of the novel virus or a well-known virus. We propose a deep learning algorithm that uses a Long Short Term Memory (LSTM) neural network, for the classification of the severe Coronavirus and other viruses. The experimental results show that at peak performance, the proposed Long Short Term Memory (LSTM) neural network model achieves a classification accuracy of 90.66%, for the classification of Coronavirus and other viruses. These results are encouraging enough to recognize deep learning algorithms as alternative avenues for detecting Coronavirus and other viruses.

**KEYWORDS:** Genome Classification, LSTM, Feature Extraction, Classification

## I. INTRODUCTION

Recent pandemic of COVID-19 (Coronavirus) caused by severe acute respiratory syndrome Coronavirus 2 (SARS-CoV-2) has been growing lethally with unusual speed. It has infected millions of people and continues a mortifying influence on the global population's health and well-being.

In this situation, genome sequence analysis and advanced artificial intelligence techniques may help researchers and medical experts to understand the genetic variants of COVID-19 or SARS-CoV-2. Genome sequence analysis of COVID-19 is crucial to understand the virus's origin, behavior, and structure, which might help produce/develop vaccines, antiviral drugs, and efficient preventive strategies.

Our Project introduces an artificial intelligence based system to perform genome sequence analysis of COVID-19 and alike viruses, e.g., SARS, middle east respiratory syndrome, and Ebola.

## II. LITERATURE SURVEY

1. Wei Zhao, Zheng Zhong, Xingzhi Xie, Qizhi Yu and Jun Liu "Relation Between Chest CT Findings and Clinical Conditions of Coronavirus Disease (COVID-19) Pneumonia: A Multicenter Study," American Journal of Roentgenology: 1-6. 10.2214/AJR.20.22976

Wei Zhao et.al [1] states that In December 2019, a series of cases of pneumonia of unknown causation emerged in Wuhan, Hubei, China, and quickly raised intense attention around the world [1]. A novel bat-origin coronavirus, 2019 novel corona-virus, was identified by means of deep sequencing analysis [2]. The virus, named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [3], is phylogenetically closest to bat SARS-like coronavirus but in a separate clade, which means that a novel coronavirus is spreading [2]. As of February 17, 2020, 72,436 laboratory-confirmed cases were consecutively reported in 31 provinces (municipalities and regions) in China, including 11,741 severe cases, 1868 fatal cases, and 6242 suspected cases [4]. The study had several limitations. First, only 101 patients with confirmed COVID-19 were included; negative results and infections with other viruses were not included in the analyses. Comprehensive investigation of the imaging features of patients with negative results and other virus infections may help us to differentiate COVID-19 pneumonia from other lung infections and then to screen patients with highly suspected cases. Second, we did not evaluate follow-up CT findings in our study. Exploring the CT changes and comparing them with the clinical parameters may help us monitor and predict outcome and support clinical decision making.

2. Wang W, Xu Y, Gao R, "Detection of SARS-CoV-2 in Different Types of Clinical Specimens. JAMA" Published online March 11, 2020. doi:10.1001/jama.2020.3786

Wang W et.al [2] proposed that In this examination, SARS-CoV-2 was recognized in examples from numerous

destinations of 205 patients with COVID-19, with lower respiratory tract tests regularly testing positive for the infection. Significantly, the live infection was identified in defecation, inferring that SARS-CoV-2 might be sent by the fecal course. A little level of blood tests had positive PCR test results, proposing that contamination now and then might be fundamental. Transmission of the infection by respiratory and extrarespiratory courses may help clarify the fast spread of illness. What's more, trying of examples from different destinations may improve the affectability and diminish bogus negative test outcomes. Two more modest investigations detailed the presence of SARS-CoV-2 in butt-centric or oral swabs and blood from 16 patients in Hubei Province,<sup>3</sup> and viral burden in throat swabs and sputum from 17 affirmed case.

3.Tao Ai, Zhenlu Yang, Hongyan Hou, Chenao Zhan, Chong Chen, Wenzhi Lv, Qian Tao, Ziyong Sun, Liming Xia ,” Correlation of Chest CT and RT-PCR Testing in Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases” Feb 26 2020<https://doi.org/10.1148/radiol.202000642>

Tao Ai,et.al [3] introducing Key Results. The positive rates of reverse-transcription polymerase chain reaction (RT-PCR) assay and chest CT in our cohort were 59% (601 of 1014 patients) and 88% (888 of 1014 patients), respectively, for the diagnosis of patients suspected of having coronavirus disease 2019 (COVID-19) chest CT accuracy in diagnosing COVID-19 during the peak of the Italian epidemic: A retrospective correlation with RT-PCR testing. strategy of combining initial RT-PCR and chest CT was analysed to ... confirming COVID-19 infection, incorporating multiple RT- Materials and Methods This study included 1014 patients in Wuhan, China, who underwent both chest CT and RT-PCR tests between January 6 and February 6, 2020. With use of RT-PCR as the reference standard, the performance of chest CT in the diagnosis of COVID-19 was assessed. In addition, for patients with multiple RT-PCR assays, the dynamic conversion of RT-PCR results (negative to positive, positive to negative) was analyzed as compared with serial chest CT scans for those with a time interval between RT-PCR tests of 4 days or more.

4. Laura Matrajt, Tiffany Leung “Evaluating the Effectiveness of Social Distancing Interventions to Delay or Flatten the Epidemic Curve of Coronavirus Disease” pmc article doi: 10.3201/eid2608.201093. Epub 2020 Apr 28.

Tiffany Leung and Laura Matrajt [2] introducing that a numerical model to examine the adequacy of social separating mediations in a moderate sized city. Intercessions decreased contacts of grown-ups >60 years old, grown-ups 20-59 years old, and youngsters <19 years old for about a month and a half. Our outcomes recommend mediations began before in the plague defer the pandemic bend and intercessions began later straighten the pestilence bend. We noticed that, while social removing mediations were set up, most new cases, hospitalizations, and passings were deflected, even with humble decreases in contact among grown-ups. Nonetheless, when intercessions finished, the pestilence bounced back. Our models propose that social removing can give critical opportunity to expand medical care limit however should happen related to testing and contact following of all speculated cases to alleviate infection transmission.

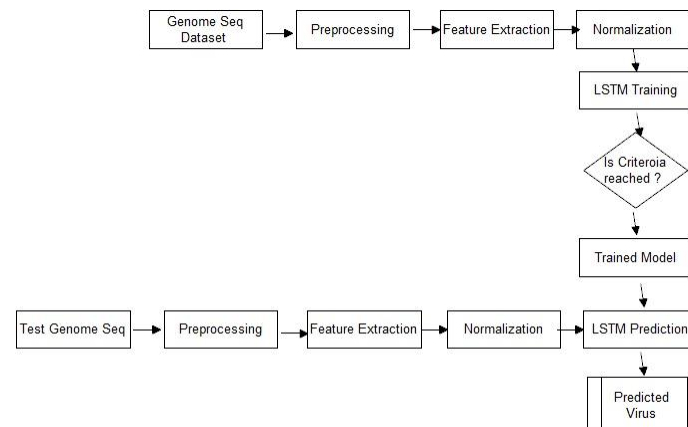
Shashi Yadav “Deep Learning based Safe Social Distancing and Face Mask Detection in Public Areas for COVID- 19 Safety Guidelines Adherence”

International Journal for Research in Applied Science & Engineering Technology (IJRASET) 2020

Shashi Yadav et.al[5] introducing to establish safe climate that adds to public security, we propose a productive PC vision put together methodology centered with respect to the continuous mechanized observing of individuals to recognize both safe social removing and face veils in broad daylight puts by executing the model on raspberry pi4 to screen action and identify infringement through camera. After identification of penetrate, the raspberry pi4 imparts ready sign to control focus at state police base camp and furthermore offer alert to public. In this proposed framework current profound learning calculation have been blended in with mathematical methods for building a vigorous modular which covers three parts of location, following, and approval. Accordingly, the proposed framework favors the general public by saving time and helps in bringing down the spread of Covid. It tends to be actualized adequately in current circumstance when lockdown is facilitated to investigate people in open social events, shopping centers, and so on Robotized investigation decreases labor to assess people in general and furthermore can be utilized in any place.



### III. PROPOSED SYSTEM APPROACH



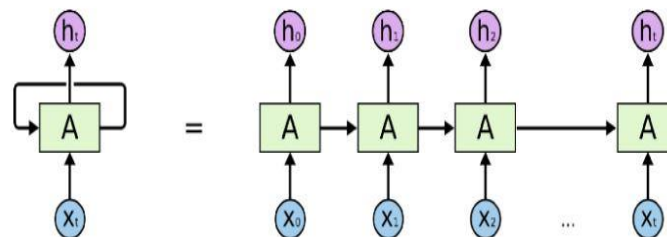
**Fig.2 Block Diagram of Proposed System**

Proposed method takes Genome Sequence Data Set. It process on input text file. After that it compares the sequence with that of normal and deduce the result. Proposed model will help with a growing workload to be able to focus on complex clinical cases. With its deep learning and LSTM algorithms, it automatically highlights abnormalities, segments anatomies. Proposed model gives better accuracy for Dataset. For real time classification large dataset is needed.

### V. METHODOLOGY USED IN PROPOSED SYSTEM

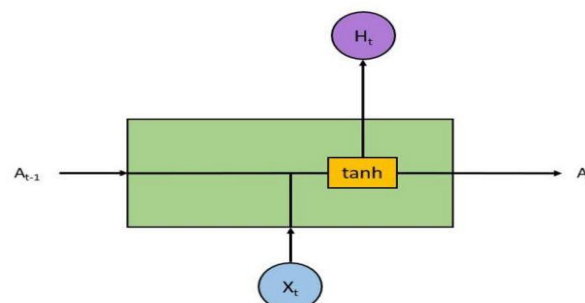
#### 1 LSTM

Long Short Term Memory networks usually just called LSTMs are a advanced version of RNN, capable of learning long-term dependencies. and were refined and popularized by many people in following work.



**Figure Architecture of LSTM**

They work tremendously well on a large variety of sequence modelling problems, and are now widely used. LSTMs are explicitly designed to avoid the long-term dependency problem. Remembering information for long periods of time is their default behaviour. The RNN unit takes the current input (X) as well as the previous input (A) to produce output (H) and current state (A).



**Figure Architecture of One Block**

LSTMs also have a similar structure though the internals have different components as compared to a single tanh (activation) layer in the RNN. There are 4 layers inside an LSTM block which interact together:

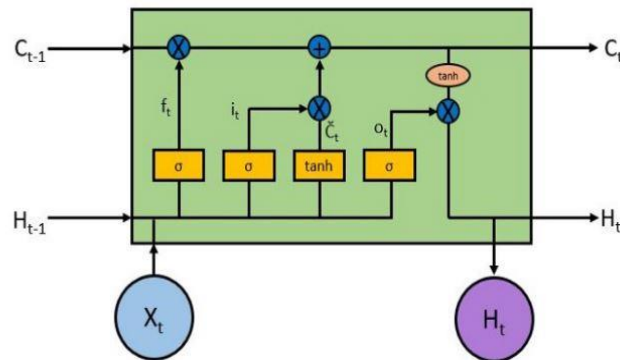


Figure. Layers of LSTM

At first it looks pretty complicated and intimidating but let's try to break it down and understand what the purpose of each layer and block is. The key to the operation of LSTM is the top horizontal line running from left to right enclosed in the highlight below. With some minor linear interactions along this line the cell state  $C$  allows information to flow through the entire LSTM unchanged which enables LSTM to remember context several time steps in the past. Into this line there are several inputs and outputs which allow us to add or remove information to the cell state. The addition or removal of information is controlled by gates. These are the sigmoid layers (Yellow boxes inside the RNN cell). They output numbers between zero and one, describing how much of each component should be let through. A value of zero means let nothing through, while a value of one means let everything through. An LSTM has three of these gates to control the cell state.

## VI. RESULT

As per the survey mentioned, we've developed the system of Genome Sequence Classification System using machine learning. Result of the same are below:



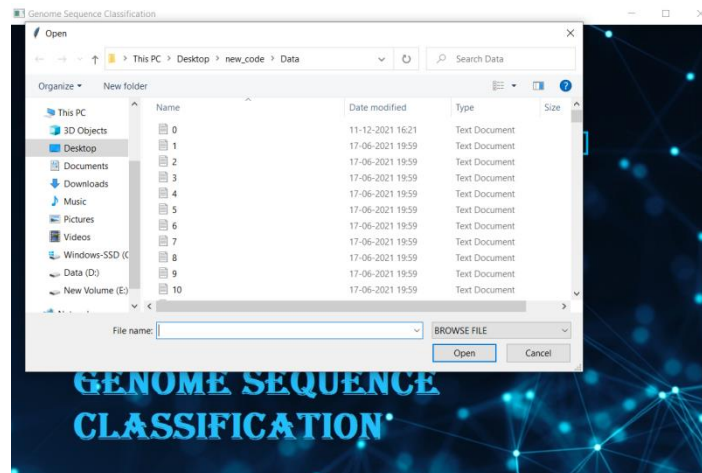


FIGURE: INPUT

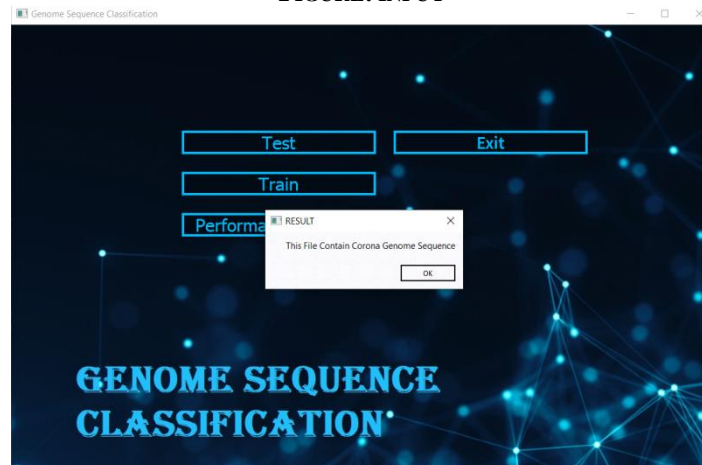


FIGURE: OUTPUT

## VII. CONCLUSION

We have introduced a system for Genome Sequence Classification of Covid-19 and other viruses using text files of genome sequence as input. User will select input file and system will process to classify the result as presence of corona or dengue genome sequence.

Due to available built in commands, design and development time get reduced. With minimal Mathematics behind deep learning, we can design and test various architectures of neural network.

## REFERENCES

1. Wei Zhao, Zheng Zhong, Xingzhi Xie, Qizhi Yu and Jun Liu "Relation Between Chest CT Findings and Clinical Conditions of Coronavirus Disease (COVID-19) Pneumonia: A Multicenter Study," American Journal of Roentgenology: 1-6. 10.2214/AJR.20.22976
2. Wang W, Xu Y, Gao R, "Detection of SARS-CoV-2 in Different Types of Clinical Specimens. JAMA" Published online March 11, 2020. doi:10.1001/jama.2020.3786
3. Tao Ai, Zhenlu Yang, Hongyan Hou, Chenao Zhan, Chong Chen, Wenzhi Lv, Qian Tao, Ziyong Sun, Liming Xia, "Correlation of Chest CT and RT-PCR Testing in Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases" Feb 26 2020 <https://doi.org/10.1148/radiol.2020200642>
4. Laura Matrajt, Tiffany Leung "Evaluating the Effectiveness of Social Distancing Interventions to Delay or Flatten the Epidemic Curve of Coronavirus Disease" pmc article doi: 10.3201/eid2608.201093. Epub 2020 Apr 28.
5. Shashi Yadav "Deep Learning based Safe Social Distancing and Face Mask Detection in Public Areas for COVID-19 Safety Guidelines Adherence"
6. International Journal for Research in Applied Science & Engineering Technology (IJRASET) 2020



**INNO SPACE**  
SJIF Scientific Journal Impact Factor  
**Impact Factor: 8.118**



**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
**INDIA**



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



**www.ijirset.com**

Scan to save the contact details