EasyChair Preprint

№ 3754

Heuristic Approach Towards Covid -19: Big Data Analytics and Classification with Natural Language Procession.

Sabyasachi Mohanty, Ritika Sharma, Mohit Saxena and Ankur Saxena

July 6, 2020

Heuristic Approach Towards COVID -19: Big Data Analytics and Classification with Natural Language Processing.

SabyasachiMohanty[1], Ritika Sharma[1], Mohit Saxena[2], Ankur Saxena[1*]
[1]Amity University UttarPradesh, Noida, India
[2]Sup'Biotech Paris, 94800, France

sabya.mohanty12@gmail.com, ritika.sharma787878@gmail.com, mohitsaxena1106@gmail.com, asaxena1@amity.edu

Abstract *- Data has tremendously incorporated our lifestyle. With advancements in technology and reduced internet cost, data usage has increased many folds resulting in and generation of huge heaps of unstructured data called as Big Data. This unstructured Big Data is difficult to handle using existing database management technology. We observed that genetic information related to Coronavirus is tremendously increasing everyday. With implementation of Big Data analytics, these databases will be easily manageable leading to advancements in COVID-19 research. In this article, we have used HDFS system for efficient data management. In our work we classified gene classes present in complete sequence so as to quickly detect mutation in no time. To achieve this, we predicted Machine Learning models to classify gene sequences faster in-class with libraries like matplotlib to construct detailed graph of the data. We choose 3 different sequences to classify gene sequence using Natural Language processing technique of Sklearn library and tested our results using logical regression.*

Keywords-*Covid-19, Matplotlib, Sklearn, Natural language processing,Classification, Logistic regression*

1. Introduction

The coronavirus outbreak hit millions of people's lives and came across with numerous deaths. The threat of this virus increases every day with new cases [1]. However, the countries affected by this virus are taking different measures using Artificial Intelligence, big data technologies [2].

The available data is now processed by machine learning software to recognize the pattern and after this formed algorithmic models [3]. After processing these models used to predict the number of currently infected cases [4]. Disease control centres use the big data from several years and forecasting tools to predict the models from different multiple sources to contain the coronavirus. According to the World Health Organization (WHO) Artificial Intelligence and big data plays a very important role in response to covid_19. Many countries use big data analytics to minimise the further risk of coronavirus[5]. The outbreak of Novel Coronavirus has spread rapidly from its origin in Wuhan, Hubei Province of China in late December of 2019. A case of pneumonia with unknown ethology was reported to the WHO country office in China on December 31st, 2019. Early investigations suspect that the cause of COVID outbreak to humans could be the Wuhan South China Seafood and wildlife market where all varieties of seafood and other kinds of meat were sold altogether pointing towards a single torrent crisis from an animal reservoir. Previous investigations of publicly available sequence data have presented phylogenetic estimates of SARS- CoV-2 time of most recent common ancestor (TMRCA) and growth rates using Bayesian phylogenetic methods [6]. Further studies revealed that many clinical, epidemiological and radiological features of COVID-19 are similar to those of severe acute respiratory syndrome coronavirus (SARS) with pangolins or bats as the most likely animal reservoir which outbreak in 2003, middle East respiratory syndrome (MERS) outspread in 2012. Immediately it was found that the virus transmitted between human to human in hospital, family setting and other gatherings.

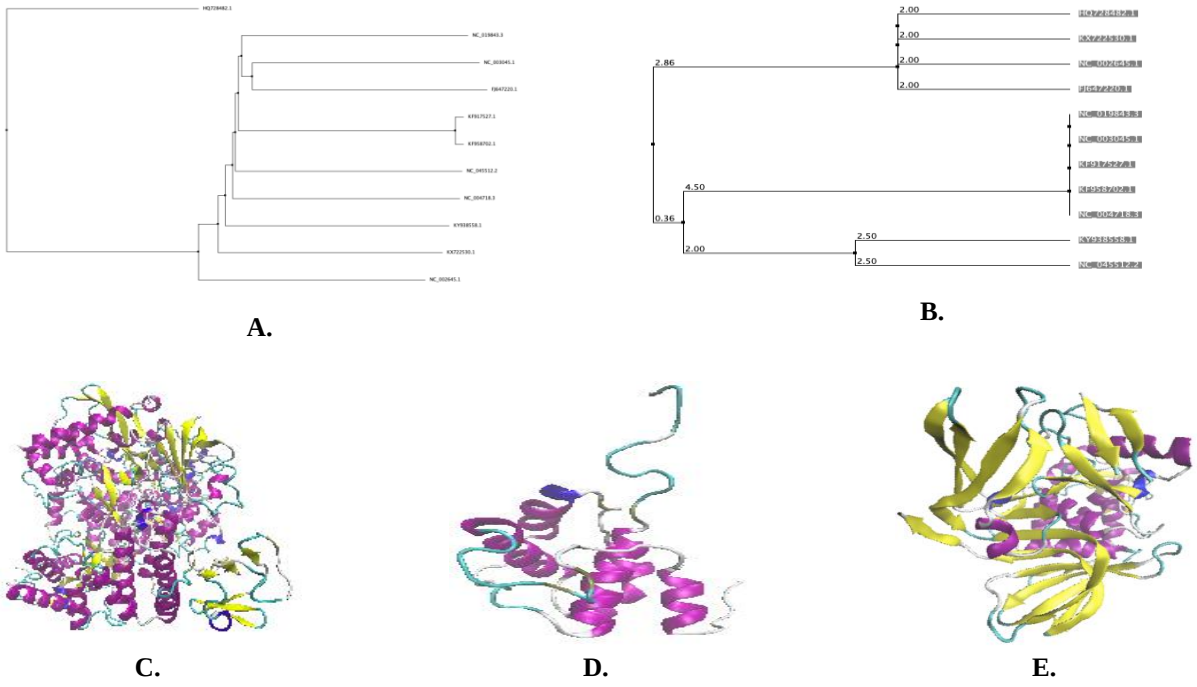

Fig.1.Increased Genetic big data expression of Coronavirus over the past years (a)Phylogenetic treeusing neighbouring joining using BLOSUM62 on region from Original of MAFFT Multiple sequence alignment of coronavirus disease host retrieved from Uniprot (b) Phylogenetic tree using average Distance between coronavirus disease hosts as mentioned in Fig. 2.1 using BLOSUM62 on region from Original of MAFT by JalView v2.5.10 (c,d,e) VMD predicted Cartoon structure of SARS-CoV-2, SARS-CoV and MERS respectively obtained from crystal structures available on Protein Data Bank.

From the above figure gene expression analysis of big data generated based on the research done on the corona virus we can see that the genetic information has been increasing exponentially everyday so we can draw a conclusion that over the next years we need more accurate fast machine learning models which can be applied to big data in order to solve or cure such pandemic cases like COVID-19 case right now and seeing the increasing amount of the data we worked on machine learning model on big data that can classify the genes in sequence faster as compared to earlier models. These three viruses belong to the same genera beta coronavirus.

Table No.1 Recent works on Artificial Intelligence and Machine Learning for Covid_19

S.N	Title	Objective	Results	Conclusion	Refer ence
1	Artificial intelligence and machine learning to fight COVID-19	Use of AI and machine learning for biological processes	The big data is managed and processed with AI resulting in better therapeutics and prevention operation and diagnosis	A cyberinfrastructure algorithm is necessary to deal with this pandemic	[7]
2	Deep learning based drug screening for	Potential drugs to be tested using ML	Sequence have been processed and result in	Modelled structure can be used for the	[8]

	novel coronavirus 2019- nCov	including big data on nCoV protease	proper dock match in 18 patients	deep learning based drug screening.	
3	Mapping the landscape of Artificial Intelligence applications against COVID-19	Use of AI, to study COVID-19 crisis at different scales including molecular, clinical, and applications.	Research highlighted the angiotensin-converting enzyme 2 (ACE2) protein, facilitates the virus' entry into host cells	Applications of AI were used to better understand the proteins involved in SARS-Cov-2 infection .	[9]
4	Leveraging Deep Learning to Simulate Coronavirus Spike proteins has the potential to predict future Zoonotic sequences	Use of deep learning in order to control the Spike Protein formation to predict the future zoonotic sequence	With test set of 100 simulated sequences, all had best BLAST matches to Spike proteins in searches against NCBI non redundant dataset	Simulated sequences from neural network may guide us in future with prospective targets for vaccine discovery.	[10]
5	COVIDier: A Deep-learning Tool For Coronaviruses Genome And Virulence Proteins Classification	Deep learning algorithm that can predict virus family from its genome	Deep learning tool that uses Sci-kit learn packagesto classify between Alpha& Beta coronavirus, MERS, SARS-CoV- 1, SARS-CoV-2, and bronchitis-CoV genomes can give an accuracy above 90% which can be used	A tool like COVIDier has potential to replace blast.	[11]
6	Artificial Intelligence against COVID-19: An Early Review	Role of AI in the fight against COVID-19 I	AI models do not have enough open datasets and models to work on, problems of big data hubris, non-adjustment of algorithms, outlier data	Innovations in AI's impact is limited and any innovation further in this domain may be a result of this pandemic situation.	[12]
7	COVID-19 Epidemic Analysis using Machine Learning and Deep Learning Algorithms	Measure everyday exponential behaviour along with the prediction of future reachability	Number of cases were scaled using minmax scaler to fit the LSTM model; predicted cases were rescaled to original range.	Early prediction of transmission can help to take necessary actions	[13]
8	Review of Big Data, Artificial Intelligence and Nature-Inspired Computing Models for Performance Improvement towards Detection of COVID-	Computing models that can be adopted to enhance the performance of detecting and predicting the COVID-19	Identified a blend of CNN and Whale optimization algorithm for COVID-19 diagnosis and prediction for patient response to treatment	NIC and big data Preprint analytics tools can be adopted in contact tracing of COVID-19 pandemic case to identify "hot spots",	[14]

	19 Pandemic Case and Contact Tracing	pandemic cases		and to alert people.	

2. Methodology

About the Dataset

We have collected this dataset through means of using Google.com then we found the genetic information about the coronavirus. The data was collected from NCBI by doing an advanced search by doing main search as (COVID-19,CORONAVIRUS, HUMAN BEINGS , SARS, MERS,GENOMICS,GENES).We further refined the search results by selecting the information available of past 30 years only so the data can be said that it is latest or up to date.

Upon finding out the proper gene sequences further processing was done that involved converting the fasta sequence format data to text format and then to csv format of the data as it is easy to handle with help of pandas ,aligning of the sequences together assigning the gene classes of gene sequence file in a structured format, and assigning gene classes to the gene sequence .The dataset processing was instrumental to correct unambiguous presentation and seamless execution of ML tools on the data.

Primary sequences on which the work is based

https://www.ncbi.nlm.nih.gov/nuccore/NC_045512
https://www.ncbi.nlm.nih.gov/nuccore/NC_004718.3?report=fasta
https://www.ncbi.nlm.nih.gov/nuccore/NC_019843.3?report=fasta

Data Storage

The role of data storage is very crucial in this analysis experiment as to easily keep the data separated and for fast and effective workflow of the analysis process.(although the work can be done without any data storage format but we preferred the data storage format in order to understand these viruses more and to easily handle the data)

Without data storage format it would have taken a very long time to do the analysis as there wouldn't have been a proper structure.

The Data Downloaded was stored in a MAP REDUCE format.

- First, we divide the input into three splits as shown in the figure. This would help to spread the work among all the map nodes.
- Then, we assign a value
- Now, a list of primary-value pairs will be created where the value is nothing but the individual words and value which is one. So, for the first line (COVID-19 SARS MERS) we have 3 key-value pairs – COVID, 1; SARS, 1; MERS, 1. The mapping process remains the same on all the nodes.
- After the mapping phase, a partition process took place where we did the sorting and shuffling process so that all the tuples with the same value are sent to the corresponding reducer.
- Now each reducer has a unique value and a list of values corresponding to that very key. For example, MERS, SARS etc.
- Then the Reducer counts the values which are present in that list of values. Reducer got a list of values like SARS(bats)(1,1,1). Then, it counts the number of ones in the very list and gives the final output as – Bats, 3.
- Finally result is given

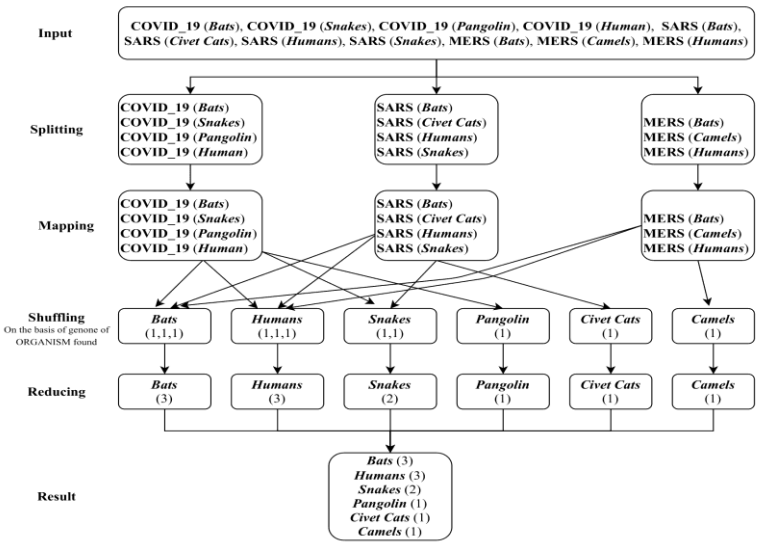

Fig.2.1.Data organising in Map Reduce Format

Environment Setup

We started the work by installing Anaconda, as it allows the library installation process to be pretty easier, it is used with python version 3.7.

We used Jupyter Notebook as our primary IDE as it is said to be one of the most precious standard IDE used in the field of machine learning as it has a very simple user interface and very user friendly as well.

Starting

In our work we used advanced machine learning techniques like Natural Language Processing with most popular libraries like Sklearn in our work .The Data was imported in CSV (comma separated values) format.

The Dataset looks like:

	sequence	clas
0	attaaaggtttatacttcccaggtaacaaaccaaccaactttcga...	
1	ttctgcaggctgcttacggtttcgtccgtgttgcagccgatcatca...	
2	agactccgtggaggaggtcttatcagaggcacgtcaacatcttaaa...	
3	cgaaataccagtggcttaccgcaaggttcttcttcgtaagaacggt...	
4	atgcactttgtccgaacaactggactttattgacactaagaggggt...	

Fig.2.2.Dataset of COVID-19

	sequence	class
0	ATATTAGGTTTTTACCTACCCAGGAAAAGCCAACCAACCTCGATCT...	1
1	TGCAGTCGATCATCAGCATACCTAGGTTTCGTCCGGGTGTGACCGA...	2
2	CCTTAAGCACCAATCACGGCCACAAGGTCGTTGAGCTGGTTGCAGA...	0
3	GTGACGAGCTTGGCACTGATCCCATTGAAGATTATGAACAAAACTG...	1
4	CCGAACAACTTGATTACATCGAGTCGAAGAGAGGTGTCTACTGCTG...	0

Fig2.3.Datasetof SARS

	sequence	class
0	GATTTAAGTGAATAGCTTGGCTATCTCACTTCCCCTCGTTCTCTTG...	1
1	TCGTGTCTCTTGTACGTCTCGGTCACAATACACGGTTTCGTCCGGT...	0
2	GGTTCATGGATGGCGAAAATGCCTATGAAGTGGTGAAGGCCATGTT...	2
3	ATTGCTTGTGAAAATCCATTCATGGTTAACCAATTGGCTTATAGCT...	2
4	ACCTCTTGCCCTGAGTGGATGGACGATTTTGAGGCGGATCCTAAAG...	3

Fig.2.4.Dataset of MERS

.head() function of pandas shows the first 5 rows of the dataset

The complete process can be summarized as :

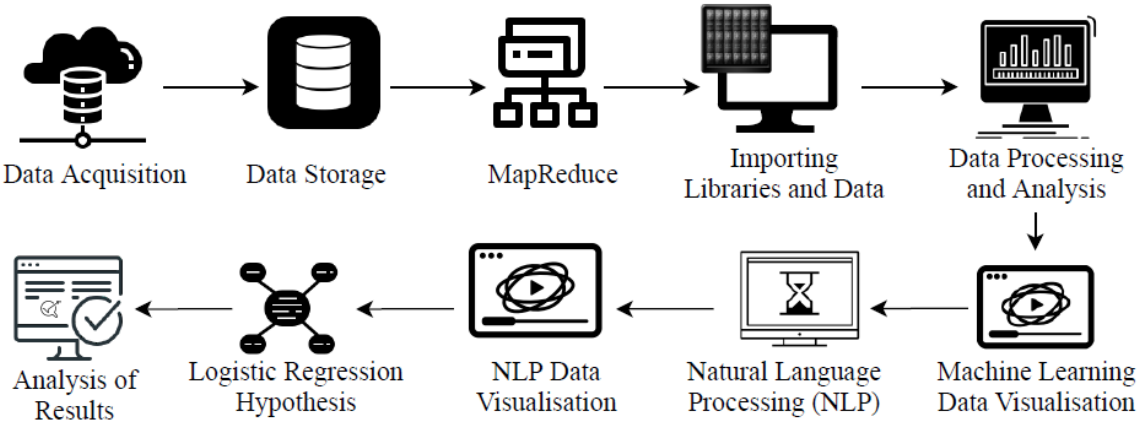

Fig2.5.The workflow of study

After we installed the jupyter notebook, which is an integrated development environment, on our PCs, we utilised the pre installed libraries in the IDE for further structuring like matplotlib for plotting graphs , Sklearn for classification training and natural language processing , pandas, numpy for mathematical operations and then pandas library to import our sequence datasets into jupyter. We performed a natural language processing technique for training classification(DNA sequencing) and then visualized the data by plotting graphs. Then we compared the individual sequences of SARS,MERS and COVID-19 by comparing their Nucleotides against each other .

3. Results

K-mer formation of sequence or Natural language processing of the data

The dataset above is not in vector form or in uniform length, which is a necessity for data to which is to be classified or a regression algorithm.

DNA and protein sequence can be said to be the transcript of life. This transcript codes information for molecules that are found in all life forms. This transcript is written as the (book)genome ,gene class or family (sentence or chapter),kmers ->motifs (words),and nucleotide bases and amino acids (alphabets or characters).

First the biological sequences were broken to k-mer length overlapping “words”. Like “ATGCATGTCA” becomes: ‘ATGCAT’, ‘TGC ATG’, ‘GCATGC’, and ‘CATGCA’.
Therefore the above example is broken into 4 parts .

We used 6 character words which can be changed according to the user or parameter setter. The length of the words and amount of overlaps are needed to be figured out for any application.

In the field of genomics this sort of data manipulation technique is called k-mer counting or counting the number of each k-mer sequence .We used python’s natural language processing as it is very easy. After k-mer formation the dataset was further converted to words.

	class	words
0	1	[atatta, tattag, attagg, ttaggt, taggtt, aggtt...
1	2	[tgcagt, gcagtc, cagtcg, agtcga, gtcgat, tcgat...
2	0	[ccttaa, cttaag, ttaagc, taagca, aagcac, agcac...
3	1	[gtgacg, tgacga, gacgag, acgagc, cgagct, gagct...
4	0	[ccgaac, cgaaca, gaacaa, aacaac, acaact, caact...

Fig.3.1.K-MER FORMATION OF COVID-19

	class	words
0	1	[attaaa, ttaaag, taaagg, aaaggt, aaggtt, aggtt...
1	1	[ttctgc, tctgca, ctgcag, tgcagg, gcaggc, caggc...
2	1	[agactc, gactcc, actccg, ctccgt, tccgtg, ccgtg...
3	0	[cgaaat, gaaata, aaatac, aatacc, atacca, tacca...
4	2	[atgcac, tgcact, gcactt, cacttt, actttg, ctttg...

Fig.3.2.K-MER FORMATION OF SARS

	class	words
0	1	[gattta, atttaa, tttaag, ttaagt, taagtg, aagtg...
1	0	[tcgtgt, cgtgtc, gtgtct, tgtctc, gtctct, tctct...
2	2	[ggttca, gttcat, ttcatg, tcatgg, catgga, atgga...
3	2	[attgct, ttgctt, tgcttg, gcttgt, cttgtg, ttgtg...
4	3	[acctct, cctctt, ctcttg, tcttgc, cttgcc, ttgcc...

Fig.3.3.K-MER FORMATION OF MERS

Individual class comparison of dataset in form of graphs or histograms which is very important for data analysis:

This step is a necessary step which is done in order to compare the amount of characterised gene classes in a particular viral sequence.

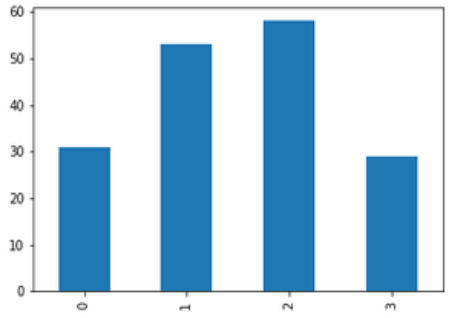

Fig3.4.Individual parameter study of classes in form of graph of COVID-19 dataset

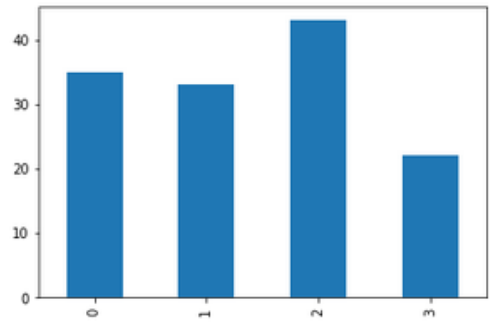

Fig.3.5.Individual parameter study of classes in form of graph of SARS dataset

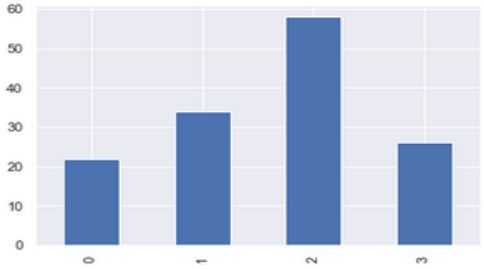

Fig.3.6.Individual parameter study of classes in form of graph of MERS dataset

(Fig3.4) From the above graph we can conclude that COVID-19 dataset has the most 2nd class of genes at maximum and the 3rd class has the least amount.

2>1>0>3

(Fig3.5) From the above graph we can conclude that the SARS dataset has the most 2nd class of genes at maximum and the 3rd class is of the least amount.

2>0>1>3

(Fig3.6) From the above graph we can conclude that the MERS dataset has the most 2nd class of genes at maximum and the 0th class is of the least amount.

2>1>3>0

From all 3 we can conclude that 2nd class of gene is maximum in all three sequences

Individual Nucleotide comparison of dataset in form of graphs or histograms which is very important for data analysis:

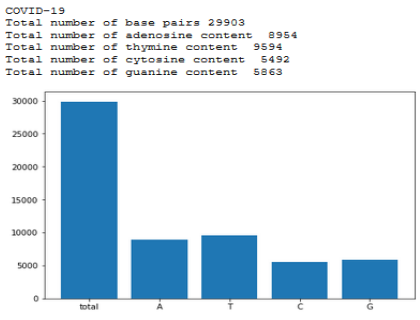

Fig.3.7.Nucleotide parameter study of COVID-19 dataset in form of graph

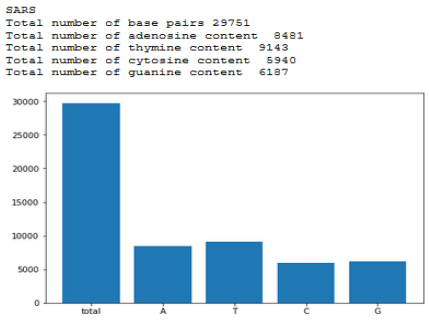

Fig.3.8.Nucleotide parameter study of SARS dataset in form of graph

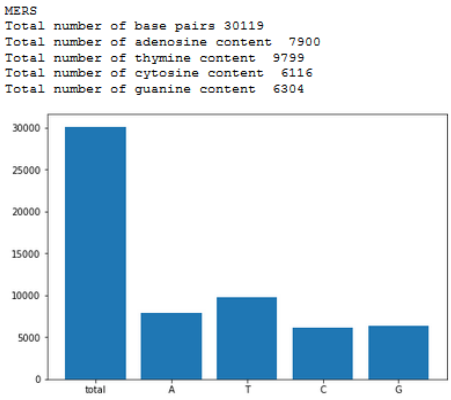

Fig.3.9.Nucleotide parameter study of MERS dataset in form of graph

From the above three analysis we can conclude that all sequences are in format from maximum amount to minimum T>A>G>C .

For better analysis we did a 1 to 1 nucleotide comparison of data:

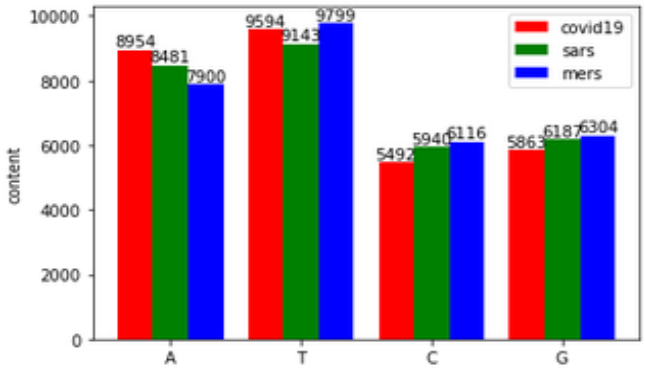

Fig.3.10.The graph shows the comparison of nucleotide comparison of all 3 sequences

Interpretation from the above plot:

The above graph shows the ATGC content of COVID-19, SARS and MERS which represents the comparison between these three sequences. The red colour rectangle represents COVID 19, the green one shows SARS and the blue rectangle shows MERS.

Maximum Content	**Minimum Content**
A-COVID-19	A-MERS
T-MERS	T-SARS
C-MERS	C-COVID-19
G-MERS	G-COVID-19

Logistic regression applied in order to classify the 3 Sequences:

logistic regression was applied after the Natural language processing as if we had done a normal base pair sequence the outcome won't predict anything whereas the result after performing the natural language processing the result predicts the potential direction of gene mutation.

logistic regression is a statistical technique used to model a binary dependent variable. It is another technique by machine learning in the field of statistics. Logistic regression is named as logistic function which is also called sigmoid function. Mathematically it is dependent on two class values ‘0’ and ‘1’and it is an S-shaped curve.

Formula: 1/(1+ e^-value) where e is the base of the natural algorithms.

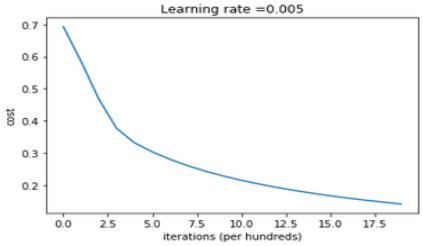

Fig3.11..The graph shows the logistic Regression equation applied on COVID-19 sequence

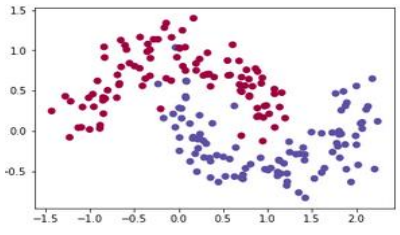

Fig3.12..The graph shows the logistic Regression equation applied on COVID-19 sequence and SARS sequence

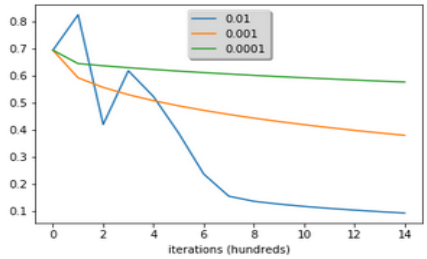

Fig3.13..The graph shows the logistic Regression equation applied on COVID-19 sequence ,MERS sequence and SARS sequence

(Fig3.11) After we did the Kmers formation using Natural Language Processing technique we applied the logistic regression equation to the Covid-19 gene sequence in order to check the changes in the data and notice the patterns in it. Then the result was plotted with help of matplotlib library.

(Fig3.12) Further we applied and noticed the pattern in the sars and covid 19 sequence as they have been worked more upon and have suitable drug candidates for curing the ill patients or for testing first on research animals and then for human trials.

(Fig3.13) Here the orange line depicts COVID-19 gene sequence , blue line depicts SARS gene sequence and green line depicts MERS gene sequence. The above result shows that the most active sequence according is MERS although it got eliminated very quickly and then followed by COVID-19 which is going on now and then SARS
Thus we can conclude that although all 3 viruses are of the same family as they follow the same pattern of sequence and genes but no 2 viruses of the corona family are the same.

4. Discussion

In this paper we highlighted on how humongous amount data or also called as big data can be easily structured/organised in an efficient way and can be worked upon using modern era technique to handle the data and with machine learning training the personal computers to predict and analyse the Genomic sequence data accucratedly without human intervention [7]. We used techniques such as natural language processing in order to structure the sequence data and then worked further upon it for increasing accuracy using different machine learning models . In the previous papers or studies the work has been done on huge scale on bigger level such as classifying then virus but not concentrated to specific gene so taking inspiration from the recent studies on we developed a model to classify the genes present in the complete sequence effectively so that mutations can be noticed easily saving time and then we compared the sequences with each other drawing attention more towards the abnormalities present the Viral genomes which has many interpretation like one of those is that all the mutation are Natural and not Synthetically done.
Epidemiologists have estimated the worldwide spread of Covid-19 and different mutated morphological features using machine learning and deep learning algorithms[13].Researchers have used Nature Inspired Computing models and Artificial Intelligence Models along with big data analytics in order to draw attention towards COVID-19 detection prevention and

diagnosis[14].Scientists have been constantly tracking and predicting the spread using AI dashboards like for previous pandemic zika virus using models and deep neural networks; the same model is being trained with the large dataset generated due to COVID-19[12]. With a strong file storage system such as HDFS file system of storage and machine learning techniques including deep learning a faster pipeline for drug screening process can be established where all potential drug targets are tested against nCoV protease effectively[8].The research institutions or researchers need more

and more data each data for example 100 sequences dataset of corona virus using deep learning algorithms and multimodal neural networks could predict the sequence alignment more accurately than BLAST which is widely being used plus the use of this system works more effectively and is more memory efficient with low memory usage as compared to BLAST and it can help in with suitable target selection for vaccine test or discovery in a novel zonnois case[10].The researchers have drawn attention towards the role of angiotensin-converting enzyme 2 (ACE2) protein the receptor that makes the Covid virus enters easily into the human cells or other carrier organisms[9].The scientist have also been able to prepare multiple deep learning algorithms that can relate the family of viruses from its genome [11].

5. Conclusion and Future Scope

Thus we can conclude that from all the big data generated in the past days needs a very robust system such as map reduce for storage and processing of the data , and the above analysis also shows that that although all 3 viruses are of same family as they follow the same pattern of sequence and genes but no 2 viruses of corona family are same so no same treatment can be done for the viruses whereas similar sort of molecules used in drugs used to treat any virus can be used and with help of machine learning we can analyse its effect on the virus virtual without spend any big amount of money in experimenting in highly expensive lab setups. This model requires better, more robust entries that are accurate and curated. We will work on the other zoonotic sequences too. Since, the diagnosis is not specific so it cannot be analysed with just a few parameters as more information is needed to be analysed due to difference in multiple ailments. The data should be 98
% accurate for it to be acceptable in real time diagnostic tool development. The dataset is required to be trained rigorously to make the analysis more efficient. Also, the future work may involve deep learning and neural networks like BERT and other better algorithms after an improvised dataset is formed.

6. Acknowledgements

We would like to thank Amity Institute of Biotechnology and our family, without their support throughout the process this paper would have not been accomplished. We would also like to thank Amity University for giving us this great opportunity.

7. References

1. Volz E, Baguelin M, Bhatia S, Boonyasiri A, Cori A, Cucunubá Z, et al. Report 5: Phylogenetic analysis of SARS-CoV-2.
2. A Agarwal and A Saxena”Comparing Machine Learning Algorithms toPredict Diabetes inWomen and Visualize Factors Affecting It the Most—A Step Toward Better HealthCare forWomen”International Conference on Innovative Computing and Communications, https://doi.org/10.1007/978-981-15-1286-5_29,2019.
3. A Saxena, N Kaushik, A Chaurasia and N Kaushik”Predicting theOutcome of an Election Results Using Sentiment Analysis of Machine Learning”International Conference on Innovative Computing and Communications,https://doi.org/10.1007/978-981-15-1286-5_43,2019.
4. Agarwal, A. and A. Saxena. Comparing Machine Learning Algorithms to Predict Diabetes in Women and Visualize Factors Affecting It the Most—A Step Toward Better Health Care for Women. in International Conference on Innovative Computing and Communications. 2020. Springer.
5. Ying S, Li F, Geng X, Li Z, Du X, Chen H, et al. Spread and control of COVID-19 in China and their associations with population movement, public health emergency measures, and medical resources. medRxiv. 2020.
6. Zhou, Y., Hou, Y., Shen, J., Huang, Y., Martin, W., & Cheng, F. (2020). Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. Cell discovery, 6(1), 1-18

7. Alimadadi, A., Aryal, S., Manandhar, I., Munroe, P. B., Joe, B., & Cheng, X. (2020). Artificial intelligence and machine learning to fight Covid-19.
8. Zhang, H., Saravanan, K. M., Yang, Y., Hossain, M. T., Li, J., Ren, X., & Wei, Y. (2020). Deep learning based drug screening for novel coronavirus 2019-nCov
9. Bullock, J., Pham, K. H., Lam, C. S. N., & Luengo-Oroz, M. (2020). Mapping the landscape of artificial intelligence applications against COVID-19. arXiv preprint arXiv:2003.11336.
10. Crossman, L. C. (2020). Leverging Deep Learning to Simulate Coronavirus Spike proteins has the potential to predict future Zoonotic sequences. bioRxiv.
11. Habib, P., Alsamman, A. M., Saber-Ayad, M., Hassanein, S. E., & Hamwieh, A. (2020). COVIDier: A Deep-learning Tool For Coronaviruses Genome And Virulence Proteins Classification. bioRxiv.
12. Naudé, W. (2020). Artificial Intelligence against COVID-19: An early review.
13. Punn, N. S., Sonbhadra, S. K., & Agarwal, S. (2020). COVID-19 Epidemic Analysis using Machine Learning and Deep Learning Algorithms. medRxiv.
14. Agbehadji, I. E., Awuzie, B. O., Ngowi, A. B., & Millham, R. C. Review of Big Data, Artificial Intelligence and Nature-Inspired Computing Models for Performance Improvement towards Detection of COVID-19 Pandemic Case and Contact Tracing.
15. Mohagaonkar, S., A. Rawlani, and A. Saxena. Efficient Decision Tree using Machine Learning Tools for Acute Ailments. in 2019 6th International Conference on Computing for Sustainable Global Development (INDIACom). 2019. IEEE