Abstract and Introduction summary

➢ Abstract:

- We will talk about “analyzing the genome sequences of the human coronavirus: Predicting virulence and mutation”.
- This virus began to appear about two years ago, threatening the whole world.
- Covid-19 pandemic, caused by the SARS-CoV-2 genome sequence of coronavirus, has affected millions of people all over the world and taken millions of lives.
- We use here an analysis pipeline comprising a classification exercise to identify the virulence of the genome sequences and extraction of important features from its genetic material that are used subsequently to predict mutation at those interesting sites using deep learning techniques.
- The accuracy of the SARS-CoV-2 genome sequencing and predicting mutations at some sites.
- We further use our mutation prediction pipeline to identify and analyze possible parents of a mutated genome sequence.
- We know that all human corona viruses have been traced to animal origins**.**

➢ Introduction:

Covid-19 was declared a global health pandemic on March 11, 2020. It is the biggest public health concern of this century. It has already surpassed the previous two outbreaks due to the coronavirus, namely, Severe Acute Respiratory Syndrome Coronavirus (SARS Cov-2) and Middle East Respiratory Syndrome Coronavirus (MERS-Cov). It is a single stranded RNA virus which is mainly 26,000 to 32,000 bases long on average. The novel coronavirus is spherical in shape and has spike protein protruding from its surface. These spikes assimilate into human cells, then undergo a structural change that allows the viral membrane to fuse with the cell membrane. The host cell is then attacked by the viral gene through intrusion and it copies itself within the host cell, producing multiple new viruses. Overall, we present an analysis pipeline that can be further utilized as well as extended and revised to analyse its virulence, Example (with respect to the number of deaths its predecessors have caused in their respective countries) and to analyse the mutation at specific important sites of the viral genome. To understand the virulence of the genome sequences and the nature of viral mutation, here, we present an analysis pipeline of the genome sequence leveraging the power of machine intelligence.

➢ Related Work:

From the very early stages of bioinformatics and the use of computers to perform biological sequence analyses, The mutation and evolution rate of RNA viruses is dramatically high, up to a million times higher than that of their hosts, and this high rate is correlated with virulence modulation and evolvability, traits considered beneficial for viral adaptation.
The genomes of the RNA viruses can accrue genetic differences while being spatially disseminated during an individual outbreak.
Deletions within the viral genome is a natural phenomenon, almost inevitably related to the attenuation of virus, however it can sometimes be linked with more severe infection.
In case of SARS-CoV-2, several reports pointed out the deletions in throughout the viral genome, often producing deletion variants of non-structural and accessory proteins that may have direct implication upon viral infectivity.
However, still there is a lack of studies bridging all the deletions taken part in the entire genome of SARS-CoV-2 globally, which may contribute to understand the pathogenic dynamics of the virus over time.
This study reveals a number of unreported mutations, which cover both mismatches and deletions in translated and untranslated regions of the SARS-CoV-2 genomes. Moreover, the geo-climate distribution of the mutations deciphered higher unique mutations as well as disease severity in the European temperate countries. Further investigations should focus on structural validations and subsequent phenotypic consequences of the deletions and/or mismatches in transmission dynamics of the current epidemics and the immediate implications of these genomic markers to develop potential prophylaxis and mitigation for tackling the crisis of pandemic COVID-19. Moreover, the identification of the conformational changes in mutated protein structures and untranslated cis-acting elements is of significance for studying the virulence, pathogenicity and transmissibility of SARS-CoV2. This mutational diversity should be investigated by further studies, including their metabolic functional pathway, intra-viral and virus-host interactions analyses.

➢ References:

[1] Coronavirus disease (COVID-19) outbreak situation;. **https://www.who.int/emergencies/diseases/novel-coronavirus-2019**.
[2] WHO coronavirus disease (COVID-19) dashboard; **https://covid19.who.int/**.
[3] Berlit P. SARS-CoV-2-(Severe acute respiratory syndrome coronavirus 2)-Pandemie und Neurologie; 2020.
[4] Single-Stranded RNA Genome of SARS-CoV2;. **https://www.ncbi.nlm.nih.gov/books/NBK554776/figure/article-52171.image.f5/**.
[5] Wang R, Zang T, Wang Y. Human mitochondrial genome compression using machine learning techniques. Human genomics. 2019;13(1):1–8.
[6] Coronavirus in South Asia, June 2020: Cases in India, Bangladesh, and Pakistan Spike;. **www.cfr.org/blog/coronavirus-south-asia-june-2020-cases-india-bangladesh-and-pakistan-spike**.
[7] Genome-wide analysis of SARS-CoV-2 virus strains circulating worldwide implicates heterogeneity; **https://www.nature.com/articles/s41598-020-70812-6**.