Indian Journal of Clinical and Medical Research (INDJCMR)
Volume 1, Issue 2 (May- August 2026), PP 62-69.
https://fdrpjournals.org/indjcmr

Modelling COVID-19 Transmission and Viral Mutation Dynamics in India

Sanghita Mitra[1], Rojina Khatun[2], Malavika Bhattacharya[3]*
[1,2,3]Department of Biotechnology, Techno India University, EM 4, Salt Lake, Sector V, Kolkata, West Bengal, India.

To Cite this Article: *Sanghita Mitra[1], Rojina Khatun[2], Malavika Bhattacharya*, "Modelling COVID-19 Transmission and Viral Mutation Dynamics in India", Indian Journal of Clinical and Medical Research, Volume 01, Issue 02, May-August 2026, PP: 62-69.*

Abstract: *Indeed, the SARS-CoV-2 epidemic has been among the most impactful events in recent public health history. It swept through 216 countries and claimed more than 590 million confirmed infections and 6.4 million fatalities before achieving a full scientific understanding of the virus itself. With great genetic agility, the virus adapted through constant evolution, facilitated by replication in billions of human hosts, with surviving strains showing increased transmissibility and efficiency in completing the infection cycle. Sometimes, these mutations manifested as moderate-to-severe symptoms due to the interaction between a high viral load and the host's immune system.A distinctive feature of SARS-CoV-2 is the fact that its evolutionary fitness is constantly changing due to mutations within its genome. Similar to many RNA viruses, SARS-CoV-2 makes numerous replication mistakes, leading to the formation of various genetic strains. Some viral lineages were defined as variants of concern by global health bodies if they showed signs of increased transmissibility, higher mortality, resistance to immunisation, and/or reduced vaccine efficacy. Five variants of concern were officially recognised by the WHO by the end of 2021. They are known as Alpha (UK, late 2020), Beta (South Africa, December 2020), Gamma (Brazil, January 2021), Delta (India, December 2020) and Omicron (South Africa, November 2021). However, the delta variant was specifically relevant for India, since it became the main driver behind the second wave of the coronavirus outbreak in the country.*

Key Word: *Covid-19, India, viral mutation, Delta variant, healthcare, mutation modelling.*

I. INTRODUCTION

Virus mutation happens when a virus makes a copy of its material and makes a mistake. For viruses like SARS-CoV-2, this is more likely to happen. The thing that helps copy the viral material, called RNA-dependent RNA polymerase, is more likely to make mistakes than the systems that other living things use to copy their genetic material. This means that the virus can change quickly.

SARS-CoV-2 viruses are a bit different. They have a tool that helps fix mistakes when they copy their genetic material. This tool is called a 3′–5′ exonuclease. It checks for mistakes. Fixes them so the virus can make a more accurate copy of itself. Some scientists, like Yin and his team, figured out how this tool works and how it affects the virus. They found out that the virus can keep its genome without falling apart because of too many mistakes.

With this tool, the virus is still changing. Important changes can happen in the proteins that make up the virus, like the spike, membrane, envelope and nucleocapsid proteins. These changes can affect how easily the virus gets into cells, how it avoids the system, and how it survives. Scientists see that different versions of the virus have amounts of genetic material, which shows that it is complicated how the virus makes copies of itself, how the host affects it and what the proteins do. It is really important to understand how all this works so we can predict what the virus will do next and come up with a plan to deal with it. The SARS-CoV-2 virus and its mutations are crucial to understand. The SARS-CoV-2 virus is still evolving [1].

II. VIRUS MUTATION MECHANISM

The rate at which a virus gets mutations depends on things. Stranded viruses usually mutate faster than double-stranded viruses. Also, RNA viruses often evolve faster than DNA viruses. Smaller viral genomes often have mutation rates. The accuracy of the polymerase, the template strand structure and nucleotide availability also affect mutation rates. The conditions inside the host cell and repair mechanisms play a role, too. Host enzymes like cytidine and adenine deaminases can make genetic changes in the viral genome. This shows that mutation rates are not the same and can vary depending on characteristics and host-related pressures [2].

In the case of RNA viruses, the error rate during replication is usually between 10^{-3} and 10^{-6} errors per nucleotide. This error rate helps RNA viruses adapt quickly to changing environments. Coronaviruses are different. They have a proofreading mechanism that removes errors made during replication. This mechanism helps keep the genome stable. This is likely why SARS-CoV-2 can have a genome without becoming unstable. RNA viruses usually have genomes. The proofreading ability of Coronaviruses is a factor in their genetic stability. It allows them to maintain a genome. This is important for their survival and evolution[3].

2.1. Importance of Modelling in Predicting Pandemic Trends

When the World Health Organisation said that COVID-19 was a pandemic on March 11 2020, India acted fast by putting a lockdown on the whole country starting on March 25. At that time, there were no vaccines or medicines that could cure COVID-19, so using math to make models was very important for the people who made decisions and the people who worked on health. These models helped a lot in understanding how COVID-19 might spread and what could be done to stop it.

What the models found out was that how many people got better from COVID-19 was very important in deciding what would happen with the outbreak. If a lot of people got better from COVID-19, then the COVID-19 problem would get smaller over time. If a lot of people were still getting COVID-19, then COVID-19 would keep spreading and become a big problem in the country. Some studies in four states in India found that the COVID-19 infection was spreading fast with a reproduction number or R_0 that was always more than 1.0. This meant that COVID-19 was still growing. In the time most of what the models said showed that the number of new COVID-19 cases every day and the total number of COVID-19 cases would keep going up[4].

2.2. Key Epidemiological Modelling Frameworks

One of the ongoing challenges in epidemiology is figuring out not only if a certain exposure causes an outcome, but also how that effect happens. Researchers often want to know how much of the effect happens through a certain intermediate factor and how much happens on its own. This aids in comprehending the mechanism through which a disease disseminates or influences health outcomes. A prevalent method for this is traditional mediation analysis, which incorporates mediating variables as covariates in regression models [5].

There have been several different modelling frameworks used in the context of COVID-19, each one made to answer different kinds of research questions and real-world problems.The model selection frequently hinges on the study's objective, such as forecasting transmission trends, assessing intervention effects, or elucidating the role of particular risk factors.

2.2.1. SIR Model

The SIR model is one of the most important and widely used models in the study of infectious diseases. It splits a closed population into three main groups: susceptible (S), which includes people who are likely to get sick; infected (I), which includes people who are currently carrying and spreading the disease; and removed (R), which includes people who have either recovered and become immune or died.

The best thing about this model is that it is easy to use. It only has three compartments, which makes it easier to quickly guess how an outbreak will spread over time and to figure out important numbers like the basic reproduction number (R_0). Even though it might not show every biological detail of how diseases spread, its simple structure makes it very helpful for understanding the overall pattern and course of an outbreak [6].

2.2.2. SEIR Model

The SEIR model adds a fourth compartment called "exposed" (E) to the SIR model. This compartment includes people who are infected but not yet infectious. This is very important for SARS-CoV-2, which has a well-known pre-symptomatic phase during which transmission can and does happen. The flow goes from exposed to infectious to removed. This order is more like how COVID-19 really spreads than the simpler SIR structure [7].

2.2.3. Agent-Based Models

Agent-Based Models (ABMs) take a very different approach than traditional compartmental models. These models don't group people into broad categories. Instead, they simulate individual people, or "agents," and how they act and interact in a certain setting. These individual interactions and behaviours then lead to larger trends in the population, like the spread of an infection.

ABMs are great for looking at groups of people who are very different from each other and for understanding complicated social behaviours, movement patterns, and differences between people. This makes them especially useful for figuring out how diseases spread in the real world. But one of the biggest problems with using ABMs is calibrating them. Setting agent-level parameters in a way that accurately reflects real-world data can be hard, and this often makes the model's predictions less accurate [8].

2.2.4. Stochastic Models

A degree of randomness is introduced into disease modelling through stochastic models that may be more challenging to address with deterministic approaches. Such uncertainty is especially critical in small or isolated populations, where a single random event can have substantial impacts on the likelihood of an epidemic's success or failure. Stochastic models can explicitly accommodate variability in disease transmission, recovery, birth rates, death rates, or external variables like environment or climate, to mention just a few possible stochastic components. In general, outputs from stochastic models are also much richer; rather than producing fixed estimates for outcomes, stochastic models yield distributions of probabilities associated with possible future scenarios, including the probability of epidemic extinction [9].

Section	Model	Core Mechanism	Covid 19 Relevance	Key Limitation
2.2.1	SIR Model	Partitions a closed population into	Enabled rapid estimation of outbreak trajectories	Sacrifices biological realism; cannot

		susceptible, infectious and removed compartments; tracks flow between states via transmission and recovery rates.	and the basic reproductive number (R_0) during early pandemic phases.	account for the pre-symptomatic infectious period characteristic of SARS-CoV-2.
2.2.2	SEIR Model	Extends SIR by inserting an Exposed(E) compartment for individuals who are infected but not yet contagious; follows the sequence E>I>R.	Captures SARS-CoV-2's documented pre-symptomatic transmission phase, aligning the model structure with the virus's actual epidemiological course.	Still relies on compartmental homogeneity, limiting representation of population-level heterogeneity in contact patterns and behaviour.
2.2.3	Agent-Based Model	Simulates discreet agents acting and interacting within an environment, macro-level epidemiological patterns emerge from individual-level rules.	Well-suited to modelling heterogeneous populations and complex social behaviours relevant to non-pharmaceutical interventions.	Parameter calibration to empirical data is computationally demanding, which can constrain predictive precision.
2.2.4	Stochastic Model	Incorporates random variations in transmission, recovery and environmental drivers; generates probability distributions over outcomes rather than single point estimates.	Particularly valuable for small or isolated populations where chance events can determine whether an outbreak occurs or fades; it quantifies forecasting uncertainty.	Greater computational complexity; results are probabilistic rather than deterministic, which can complicate direct policy communication.

Table 1: Summary of Key Epidemiological Modelling Framework (Section 2.2.1-2.2.4)

III. EVOLUTION OF SARS-COV-2 IN INDIA

It is believed that the human-derived beta coronavirus SARS-CoV-2 is bat-derived and was first transmitted to humans around December 2019. Overall, despite having a lower inherent variability than other RNA viruses, SARS- CoV-2 has gained mutations in its genome over time, producing diverse and genetically different lineages/clades with significantly different behaviour and characteristics [10].

The high population density and number of populations in India have enabled virus transmission, producing high case and death counts. Indian genomic surveillance only provided 1.58% of whole-genome sequences and 0.14% of detected case sequences, limitations being due to resources available, but still enough to produce important scientific information.

A study using population genetics during the first 17 months provided documentation of which lineages have spread across all Indian states and union territories, while monitoring VOCs and VOIs throughout and noted that both variants Kappa and Delta evolved first in India before travelling elsewhere in the world [11].

The mutational profiles have been observed through month to month, study of mutation burden on nearly 17,000 protein sequences across all 26 viral proteins with entropy analysis and were seen to be dynamic early in the pandemic before levelling out during four major structural protein analysis, while the non-structural proteins were observed to increase their variation more consistently; potentially an indication of limited new available mutations over time[12].

3.1. Mutation Patterns and Genetic Sequencing Data from ICMR, INSACOG, and GISAID

Several phylogenetic classification systems have been implemented to keep track of the diversity of SARS-CoV-2 lineages. In India, a total of 330 viral isolates were whole-genome sequenced using next-generation sequencing platforms. Also, a phylogenetic analysis of 3,014 viral sequences from 20 states and union territories from January to September 2020 was done by retrieving data from the GISAID database. Different lineage classification schemes, such as the Nextstrain clades, the Pangolin lineages, and the GISAID clade designations, were compared, and a few computer programs on the Datamonkey server were also used to determine sites under positive or purifying selection within the genomes. By mid-2020, most Indian states were predominantly characterised by GR and GH clades, but states such as West Bengal, Uttarakhand, and Haryana still exhibited circulation of the GISAID clade O even post July. Two lineages, such as B.1.1.8 and B.1.113, were observed that seem to have evolved in India itself [13].

The Kappa variant was dominant in the earlier stages of the second wave of COVID-19 in India, while the Delta variant gained predominance as the second wave peaked. To tackle the rapidly evolving complex genomic landscape of the virus, the Indian SARS-CoV-2 Genomics Consortium (INSACOG) was set up in December 2020 under the Ministry of Health and Family Welfare and was assisted by CSIR, DBT and ICMR. Initially comprising ten major national laboratory hubs, it was subsequently expanded by introducing satellite facilities at various locations. The primary objective of this network was to enhance sample collection and transport, while bolstering genomic surveillance for Variants of Concern (VOCs) and Variants of Interest (VOIs) [14].

3.2. Regional Differences in Variant Emergence

India's first reported case of COVID-19 was a university student who returned to Thrissur from Wuhan on 30 January 2020. Cases started trickling in in scattered locations in states such as Kerala, Karnataka, Maharashtra, Gujarat, Delhi, Punjab and Ladakh in the ensuing weeks; in the vast majority of early cases, the contact was via international travel. One of the key early superspreading events was a large religious meeting that took place in Delhi, and that was attributed by health authorities to more than 4,000 cases across India and at least 27 deaths. The transmission routes from this event spanned across multiple states, such as the Andaman and Nicobar Islands, showing the rapid ability of the disease to spread throughout travel and social networks. By April 2020, infection chains began spreading along major East-West corridors. It is possible to say that Eastern India experienced comparatively less spread during the initial phase, and this might partly be due to low levels of urbanisation in several districts away from Kolkata. On the other hand, Southern India, namely Kerala, Tamil Nadu and Karnataka, recorded a relatively high proportion of cases by early April, and by the end of the month, major upticks were witnessed in states like Maharashtra, Goa, Madhya Pradesh, Gujarat, Delhi, Haryana, Punjab and part of the Jammu and Kashmir region. Mumbai, Delhi, Ahmedabad, Bangalore, Indore and Kolkata were major hotspots, which is more or less expected given their populations, high levels of mobility, and population density [15].

IV. IMPACT OF MUTATIONS ON DISEASE DYNAMICS

The rapid development of new strains of SARS-CoV-2 caused genuine alarm, both in the scientific community and in public health infrastructure, and with good reason. Several of the variants, officially recognised by WHO asVariants of Concern(VOCs), were identified as being capable of compromising the protective efficacy of vaccines designed against the original ancestral strain. This development, along with continued new emergent variant production, has compounded the struggle for the realisation of herd immunity, as these newer variant strains of the virus possess the capacity to partially escape from immunity elicited by vaccination or infection. The spike (S) protein is central to this evolutionary pressure, since it is the mediating factor in enabling viral entry into the host cell and is the primary target for neutralisingantibodies and for the immune system in general. Mutations within the S protein have frequently been shown to confer higher levels of infectivity along with a diminished capacity for neutralisation by antibodies, which affects vaccine efficacy (against the original strain) as well as generating new variants in a cycle that is yet to be clearly elucidated [16].

4.1. Transmission Rate and Virulence Changes Due to Mutations

The SARS-CoV-2 virus is an enveloped, positive-sense,single-stranded RNA virus in the genus Beta coronavirus, and like most RNA viruses, possesses a naturally high propensity to mutate, continuing to acquire genetic variations as the pandemic progresses. This process is occurring at a quantifiable rate annually, with mutations playing a crucial role in virus evolution. Since the beginning of the pandemic,several different variant lineages have been detected from around the globe, with the majority of functionally important mutations occurring in the spike gene. Consequently,many epidemiologically significant variant strains have been found to exist in the environment, such as alpha from the UK, delta from India, beta from South Africa, and the Cluster 5 variant first isolated from Denmark.[17].

4.2. Immune Escape and Vaccine Breakthrough Infections

Vaccines then became the cornerstone of the world's response to COVID-19, and the initial data on clinical trials were highly positive and effective, demonstrating over 90% protection against serious disease and death. However, the advent of the Omicron variant of SARS-CoV-2 created a difficult situation, as it spreads much more readily than the prior strains, and infections with Omicron are now being detected in fully vaccinated populations. Concerns have arisen as to whether the spread can be stopped. The factors involved in breakthrough infection are numerous and often interact simultaneously: they may be broadly divided into immune factors such as the gap between mucosal immunity and systemic immunity and duration of immunity, host factors including age, chronic co morbidities, immunological status and the use of immunosuppressive drugs, vaccine-specific factors such as vaccine platform type, dosage, schedule and method of administration, and virus specific factors such as the strain of virus, viral load, rate of replication and its ability to evade immunity. High-powered, large population-based studies, as well as neutralisation antibody assays, are both difficult and resource-intensive, and there has also been some evidence that slightly lower rates of breakthrough infection may be found in those who have suffered from prior SARS-CoV-2-related problems (e.g., long COVID and MIS-C) at a younger age[18].

4.3. Modelling Outcomes Showing Impact of New Variants on Hospitalisations and Mortality

The initial months of the pandemic saw hospitals forced to treat critically ill patients with a virtually non-existent body of knowledge on effective treatment and limited established protocols; much of the early response was driven by rapidly changing clinical guidelines and rapid clinical decision-making, leading to alarmingly high hospital Case Fatality Rates (CFRs), as close to 1 in 3 hospitalised cases died in early reports. By Summer 2020, many single-centre studies were reporting a decrease in in-hospital mortality, a finding later corroborated by large-scale national analyses. A study analysing data from nearly 600 hospitals in the US and encompassing over 32 million admissions from Jan 2017 to Dec 2020 reported a dramatic surge in COVID-19-related deaths during the early stages of the outbreak, followed by a drastic decrease until the rate fell below 10% in June 2020, staying relatively low until the end of the study. Nonetheless, overall hospital mortality, even among patients without respiratory failure, continued to remain above pre-pandemic levels; patients who died included most elderly, and mechanically ventilated patients, and approximately 1 in 10 hospitalised COVID-19 patients died since the summer of 2020 [19].

Logistic regression, through multivariate analyses, also revealed common predictors of death. Risk factors more likely

to cause death have repeatedly been demonstrated to be male gender, age >60, blood group O+, hypertension, type 2 diabetes, and obesity. The common symptoms noted at the time of presentation were fevers, malaise, breathlessness and fatigue. Bilateral ground glass opacities, categorised as BiRad 5 on CT imaging, were the most common radiological finding of the sickest patients. Predictive scoring with a model incorporating age, fevers, cough, sore throat, fatigue, dyspnoea, unilateral consolidation on CT, and several other lab markers showed 76% accuracy in predicting outcome [20].

V. VIRAL MUTATION MODELLING APPROACHES USED IN INDIA

Accurately predicting the fitness impact of new mutations on a virus is arguably one of the most difficult problems in the field. For SARS-CoV-2, by far the most important interaction during infection is the binding of the viral spike protein with host ACE2 receptors, while being the target of neutralising antibodies for immune protection.

The proposed model takes a structure-based, computational approach that unifies ACE2 binding and antibody sensitivity into a single viral fitness model to predict how particular mutations could affect viral fitness.

The predicted impacts for close to 380,000 unique mutation scenarios were generated using a combination of 10ACE2 and 10 antibody-spike protein complexes. By making a model prediction using a simple relative measure of how the change in binding affinities for the antibody and ACE2 complexes differs, we have minimised modelling error for these systematic predictions of viral fitness.

In an environment with both ACE2 and antibody presence, it is more likely for the mutation to persist if its interaction with ACE2 affinity is enhanced and it confers antibody resistance, making the model a potentially powerful tool for characterising immune properties of existing variants, and for designing future vaccines [21].

5.1. Epidemiological Models

A slightly more sophisticated epidemiological framework is to incorporate testing behaviour into models of disease transmission, and the way an individual's decision on whether to test is informed by the prevalent level of infection in society. This is a better model, as public perception often follows perceived levels of risk.

The practical implication of this is quite significant. Often, epidemic models that ignore testing behaviour will underestimate epidemic rates of transmission and death during the beginning and ending phases of the epidemic, where the number of known cases is relatively small. Thus, the public health implication of this is quite obvious and highlights the need for not only investment in increased testing capacity for the purpose of monitoring and reporting but for epidemic control purposes as well [22].

5.2. Genomic Modelling

In global sequence analyses (data since late 2020), a large number of mutations were found in the ORF7 b and ORF8 regions of SARS-CoV-2. A high prevalence of mutations co-occurring within these regions, however, was not seen in sequence data from the United Kingdom, India, or Brazil. This suggests mutation distribution is not simply a function of which regions the virus is transmitting at the greatest extent.

Concurrently, global analysis suggests increasing rates of genetic change among SARS-CoV-2 proteins are occurring globally. When pooled across countries, mutations appear in many regions of the genome at readily-accessible sites within amino acids. Going forward, most recurring mutations are predicted within the proteins ORF3a, the membrane protein, and ORF8.

The pattern of emergence location of new strains may also not reflect that seen in other beta coronaviruses; there may be lessons from this trend regarding initial origins and adaptability of the virus [23].

5.3. Computational Modelling

AI and computer vision have been extensively employed for tasks in the diagnosis of COVID-19 from CT images, such as lesion segmentation, classification and image analysis. The application of these technologies has garnered much attention due to their potential to assist in the efficient and rapid diagnosis of COVID-19. A systematic review identified 114 studies on the topic by collating literature from many leading scientific sources such as PubMed, Scopus, IEEE Xplore, Web of Science, ScienceDirect, Google Scholar and Kaggle. It was found that the diagnostic tools based on AI have shown great promise in automatically diagnosing COVID-19 from images. Although the accuracy results are encouraging, it was also recommended that there is a need for further validation before their integration into routine clinical practice [24].

5.4. Mathematical Approaches

In the process of developing mathematical models for the COVID-19 pandemic, several issues that were not addressed in traditional epidemic models had to be taken into consideration. One of the main problems encountered was undiagnosed infected people transmitting the virus in the community, unaware of their status and unable to be tracked through testing. Different people in the population have different contact rates, and also, transmission in settings of different contact densities, i.e., household, workplace, public transport, and urban dense areas, must be taken into account.

Several crucial adjustments were made in the model to improve its realism, such as correction for the discrepancy between reported and real cases, introduction of mechanisms to incorporate import of cases of non-diagnosed people from countries to other countries and from region to region, adding mechanisms for contact tracing, and adding isolation of incoming travellers to quarantine places. Through integrating the real conditions into these models, the models are more adequate for the description of pandemic progression and thus more valuable for policy decisions [25].

Section	Approach	Methodology	Key Findings / Application	Limitations / Outlook
5.1	Epidemiological Modelling	Transmission models were refined by incorporating individual-level testing behaviour — specifically, how perceived disease prevalence shapes the decision to seek testing — into standard epidemiological frameworks.	Omitting testing-behaviour dynamics causes models to overestimate true transmission and mortality, particularly at the early and late phases of an epidemic when reported case counts are low.	Findings highlight expanded testing capacity as a direct policy lever for outbreak management, not merely a surveillance tool; however, model accuracy depends on reliable behavioural data.
5.2		Global SARS-CoV-2 sequence datasets were analysed to identify mutation hotspots across the viral proteome, with geographic stratification to compare mutational profiles across regions including India, the UK, and Brazil.	Mutations cluster disproportionately in ORF7b and ORF8 globally; future strains are projected to carry common mutations in ORF3a, membrane, and ORF8 proteins. Geographic variant distribution diverges from patterns observed in related beta coronaviruses.	Mutational burden is accelerating across most accessible amino acid positions. Regional disparities in sequencing depth may bias global estimates; evolutionary origins remain incompletely understood.
5.3	Computational Modelling	A systematic review synthesising 114 studies drawn from multiple academic databases evaluated AI and computer-vision methods applied to CT imaging for COVID-19 diagnosis, covering lesion segmentation, disease classification, and related tasks.	AI-assisted CT diagnostic tools demonstrate meaningful promise for automating COVID-19 detection, with encouraging accuracy metrics reported across the reviewed studies.	Broader clinical deployment requires further validation; generalisability across varied imaging equipment and patient populations remains a principal concern before routine implementation.
5.4		Standard compartmental frameworks were adapted to reflect COVID-19-specific complexities: undetected infectious individuals, heterogeneous contact structures, cross-border importation of undiagnosed cases, contact tracing protocols, and quarantine durations.	Embedding these real-world features into formal model structures substantially improved correspondence between model projections and observed outbreak dynamics, increasing practical utility for policy formulation.	Each added parameter requires reliable empirical data for calibration; model complexity trades off against transparency, which can complicate communication to non-technical stakeholders.

Table 2: Summary of Viral Mutation Modelling Approaches Used in India (Sections 5.1–5.4)

V. CONCLUSION

This increasing trend in the analysis of SARS-CoV-2 mutation modelling in India truly shows how it's crucial to have genomic data merged with epidemic data. These studies mentioned above unequivocally establish that combining transmission

dynamics with phylogenetics and computational forecasting has greatly increased the capability of the country to track evolving strains and to assess the likely threat posed by them, as well as the likely epidemic burden over time.

At the same time, the above articles are equally open to question and state clearly the problems that remain. The uneven distribution of sequencing across various regions of India, the massive disparity in capacities across different states, and disparate institutional sharing of information clearly hamper prompt analyses of emergent strains, and while we might suggest a rise in sequencing centres, an effective, harmonised and nationally coordinated data ecosystem which links epidemiological data with clinical and genomic information nearly in real time, would address these issues efficiently.

The Indian pandemic story is then in some way proof of what we know is possible and what it takes for effective viral mutation surveillance. Despite resource constraints, India has been able to produce data that has established how the synergistic linking of data types can inform critical health decisions. This is crucial not only for India, but for health systems worldwide, which are seeking robust pandemic responses in future.

Acknowledgment
We would like to express our heartfelt gratitude to the Chancellor of Techno India University.

Author Contributions
Sanghita Mitra: Data Collection, Formal Analysis, Writing – Original Draft
Rojina Khatun: Resources, Writing-Editing
Malavika Bhattacharya: Conceptualisation, Supervision

References

1. Yin X, Popa H, Stapon A, Bouda E, Garcia-Diaz M. Fidelity of Ribonucleotide Incorporation by the SARS-CoV-2 Replication Complex. J Mol Biol. 2023 Mar 1;435(5):167973. doi: 10.1016/j.jmb.2023.167973. Epub 2023 Jan 20. PMID: 36690070; PMCID: PMC9854147.
2. Sanjuán R, Domingo-Calap P. Mechanisms of viral mutation. Cell Mol Life Sci. 2016 Dec;73(23):4433-4448. doi: 10.1007/s00018-016-2299-6. Epub 2016 Jul 8. PMID: 27392606; PMCID: PMC5075021.
3. Domingo E, García-Crespo C, Lobo-Vega R, Perales C. Mutation Rates, Mutation Frequencies, and Proofreading-Repair Activities in RNA Virus Genetics. Viruses. 2021 Sep 21;13(9):1882. doi: 10.3390/v13091882. PMID: 34578463; PMCID: PMC8473064.
4. Khajanchi S, Sarkar K. Forecasting the daily and cumulative number of cases for the COVID-19 pandemic in India. Chaos. 2020 Jul;30(7):071101. doi: 10.1063/5.0016240. PMID: 32752627; PMCID: PMC7585452.
5. Richiardi L, Bellocco R, Zugna D. Mediation analysis in epidemiology: methods, interpretation and bias. Int J Epidemiol. 2013 Oct;42(5):1511-9. doi: 10.1093/ije/dyt127. Epub 2013 Sep 9. PMID: 24019424.
6. McLean AR. Infectious Disease Modeling. Infectious Diseases. 2012 Dec 5:99–115. doi: 10.1007/978-1-4614-5719-0_5. PMCID: PMC7121366.
7. Hethcote, H. W., & Tudor, D. W. (1980). Integral equation models for endemic infectious diseases. *Journal of Mathematical Biology*, 9(1), 37–47
8. Monti, C., Pangallo, M., De Francisci Morales, G., &Bonchi, F. (2022). *On Learning Agent-Based Models from Data*. arXiv.
9. Centres for Disease Control and Prevention (CDC). Technical Explainer: Infectious Disease Transmission Models — Stochastic Models. *CDC Modeling& Forecasting*
10. Potdar VA, Cherian SS. Evolution of the SARS-CoV-2 pandemic in India. Med J Armed Forces India. 2022 Jul;78(3):264-270. doi: 10.1016/j.mjafi.2022.05.006. Epub 2022 May 16. PMID: 35599988; PMCID: PMC9108071
11. Limaye S, Kasibhatla SM, Ramtirthkar M, Kinikar M, Kale MM, Kulkarni-Kale U. Circulation and Evolution of SARS-CoV-2 in India: Let the Data Speak. Viruses. 2021 Nov 8;13(11):2238. doi: 10.3390/v13112238. PMID: 34835044; PMCID: PMC8619538.
12. Santoni D, Ghosh N, Saha I. An entropy-based study on mutational trajectory of SARS-CoV-2 in India. Infect Genet Evol. 2022 Jan;97:105154. doi: 10.1016/j.meegid.2021.105154. Epub 2021 Nov 19. PMID: 34808395; PMCID: PMC8603812.
13. Potdar, Varsha1; Vipat, Veena1; Ramdasi, Ashwini$; Jadhav, Santosh2; Pawar-Patil, Jayashri$; Walimbe, Atul2; Patil, Sucheta S.2; Choudhury, Manohar L.1; Shastri, Jayanthi3; Agrawal, Sachee3; Pawar, Shailesh4; Lole, Kavita5; Abraham, Priya$; Cherian, Sarah2,* ICMR-NIV NIC Team. Phylogenetic classification of the whole-genome sequences of SARS-CoV-2 from India & evolutionary trends. Indian Journal of Medical Research 153(1-2):p 166-174, Jan–Feb 2021. | DOI: 10.4103/ijmr.IJMR_3418_20
14. Sarkar, P., Banerjee, S., Saha, S.A. *et al.* Genome surveillance of SARS-CoV-2 variants and their role in pathogenesis focusing on second wave of COVID-19 in India. *Sci Rep* **13**, 4692 (2023).
15. Gupta, D., Biswas, D. &Kabiraj, P. COVID-19 outbreak and Urban dynamics: regional variations in India. *GeoJournal* **87**, 2719–2737 (2022).
16. Faraji N, Zeinali T, Joukar F, Aleali MS, Eslami N, Shenagari M, Mansour-Ghanaei F. Mutational dynamics of SARS-CoV-2: Impact on future COVID-19 vaccine strategies. Heliyon. 2024 Apr 25;10(9):e30208. doi: 10.1016/j.heliyon.2024.e30208. PMID: 38707429; PMCID: PMC11066641.
17. Kannan S, Shaik Syed Ali P, Sheeza A. Evolving biothreat of variant SARS-CoV-2 - molecular properties, virulence and epidemiology. Eur Rev Med Pharmacol Sci. 2021 Jun;25(12):4405-4412. doi: 10.26355/eurrev_202106_26151. PMID: 34227076.
18. Amanatidou E, Gkiouliava A, Pella E, Serafidi M, Tsilingiris D, Vallianou NG, Karampela I, Dalamaga M. Breakthrough infections after COVID-19 vaccination: Insights, perspectives and challenges. Metabol Open. 2022 Mar 17;14:100180. doi: 10.1016/j.metop.2022.100180. PMID: 35313532; PMCID: PMC8928742.
19. Phillips MC, Sarff L, Banerjee J, Coffey C, Holtom P, Meurer S, Wald-Dickler N, Spellberg B. Effect of mortality from COVID-19 on inpatient outcomes. J Med Virol. 2022 Jan;94(1):318-326. doi: 10.1002/jmv.27332. Epub 2021 Sep 18. PMID: 34516010; PMCID: PMC8662145.
20. Yupari-Azabache IL, Briceno RK, Díaz-Ortega JL, Otiniano NM, Paredes-Díaz SE. Mortality Due to Covid-19 in Hospitalized Patients: A Prediction Model Based on Different Risk Factors. *Risk Manag Healthc Policy*. 2025;18:3181-3198
21. Yupari-Azabache IL, Briceno RK, Díaz-Ortega JL, Otiniano NM, Paredes-Díaz SE. Mortality Due to Covid-19 in Hospitalized Patients: A Prediction Model Based on Different Risk Factors. *Risk Manag Healthc Policy*. 2025;18:3181-3198

22. Thakur S, Planeta Kepp K, Mehra R. Predicting virus Fitness: Towards a structure-based computational model. J Struct Biol. 2023 Dec;215(4):108042. doi: 10.1016/j.jsb.2023.108042. Epub 2023 Nov 4. PMID: 37931730.
23. Sarkar J. To test or not to test? A new behavioural epidemiology framework for COVID-19. PLoS One. 2024 Dec 17;19(12):e0309423. doi: 10.1371/journal.pone.0309423. PMID: 39689139; PMCID: PMC11651578.
24. Hassan SS, Basu P, Redwan EM, Lundstrom K, Choudhury PP, Serrano-Aroca Á, Azad GK, Aljabali AAA, Palu G, Abd El-Aziz TM, Barh D, Uhal BD, Adadi P, Takayama K, Bazan NG, Tambuwala MM, Lal A, Chauhan G, Baetas-da-Cruz W, Sherchan SP, Uversky VN. Periodically aperiodic pattern of SARS-CoV-2 mutations underpins the uncertainty of its origin and evolution. Environ Res. 2022 Mar;204(Pt B):112092. doi: 10.1016/j.envres.2021.112092. Epub 2021 Sep 22. PMID: 34562480; PMCID: PMC8457672.
25. Hassan H, Ren Z, Zhao H, Huang S, Li D, Xiang S, Kang Y, Chen S, Huang B. Review and classification of AI-enabled COVID-19 CT imaging models based on computer vision tasks. ComputBiol Med. 2022 Feb;141:105123. doi: 10.1016/j.compbiomed.2021.105123. Epub 2021 Dec 18. PMID: 34953356; PMCID: PMC8684223.