Journal of Applied Informatics and Computing (JAIC)
Vol.10, No.3, June 2026, pp. 2213~2326
e-ISSN: 2548-6861

Temporal Deep Learning for Probabilistic Mutation Forecasting in SARS-CoV-2 Spike Protein Sequences

Robert Selemani[1]*, Belinda Ndlovu[2], Amazing Maphosa[3]***
[1,2,3] Informatics and Analytics Department, National University of Science and Technology
n0189584m@students.nust.ac.zw [1], belinda.ndlovu@nust.ac.zw [2], amazing.maphosa@nust.ac.zw [3]

Article Info

Article history:

Received 2026-03-08
Revised 2026-04-19
Accepted 2026-05-08

Keyword:

Deep Learning,
Long Short-Term Memory,
Mutation Forecasting,
SARS-CoV-2,
Spike Protein,
Temporal Sequence Modelling.

ABSTRACT

Modelling the evolution of biological sequences under temporal and probabilistic constraints remains a complex computational challenge. This study investigates longitudinal deep learning for probabilistic modelling of mutation patterns in the SARS-CoV-2 Spike Protein. A stacked Long Short-Term Memory (LSTM) network is trained on temporally ordered amino acid sequences to estimate residue-level substitution probabilities and rank plausible future mutations. Unlike deterministic classification approaches, the proposed framework treats mutation prediction as a probabilistic ranking task, accounting for the inherent uncertainty of viral evolution. The model is evaluated using metrics suitable for imbalanced sequence data, including Top-K accuracy, precision, recall, F1-score, and ROC-AUC. Results indicate strong ranking performance, with Top-3 accuracy of 94.6% and ROC-AUC of 0.91. In comparison, the overall accuracy (93.1%) is interpreted cautiously, given the dominance of conserved residues. Error analysis shows that difficult predictions are concentrated in low-frequency, rapidly evolving residue positions. A comparison with a frequency-based baseline demonstrates that the LSTM captures temporal dependencies beyond static substitution patterns. Predicted mutation distributions exhibit a structured alignment with known functional regions of the Spike Protein, as supported by the established literature, providing qualitative biological validation. This study contributes a temporally structured and probabilistic framework for mutation modelling, emphasising ranking-based evaluation and biologically contextualised interpretation. The findings demonstrate the feasibility of probabilistic mutation forecasting under controlled experimental conditions and provide a methodological foundation for future research on AI-assisted genomic surveillance.

I. INTRODUCTION

The rapid evolution of infectious viruses presents a persistent challenge for computational models that aim to characterise and anticipate sequence-level change over time[1]. While genomic surveillance has traditionally relied on retrospective evolutionary analysis, emerging advances in artificial intelligence now enable the possibility of modelling viral mutation dynamics as a temporally structured probabilistic inference problem[2], [3]. RNA viruses such as SARS-CoV-2 exhibit exceptionally high mutation rates driven by replication errors, immune selection pressures, and transmission dynamics, enabling the frequent emergence of novel variants with altered transmissibility, pathogenicity, and immune escape potential [4],[5]. The COVID-19 pandemic has demonstrated how quickly such mutations can undermine existing public health responses, with successive variants of concern (VOCs), including Alpha, Delta, and multiple Omicron sub-lineages, reducing vaccine efficacy and complicating containment strategies [6]. This shift from retrospective mutation analysis toward probabilistic mutation forecasting under controlled experimental conditions represents an emerging direction in computational genomics and motivates the modelling approach adopted in this study.

Genomic surveillance has therefore become a cornerstone of pandemic preparedness [7]. Conventional approaches to

studying viral evolution, primarily phylogenetic analysis and sequence alignment, have been indispensable for reconstructing transmission pathways and characterising lineage emergence [2], [8]. However, these methods are inherently retrospective, offering limited capacity to anticipate future evolutionary trajectories, particularly under rapidly changing selective pressures, such as population-level immunity and widespread vaccination [8]. As a result, public health responses and vaccine updates frequently lag behind the emergence of dominant variants, underscoring the need for predictive rather than reactive surveillance.

The unprecedented volume of viral genomic data generated during the COVID-19 pandemic presents new opportunities for data-driven mutation forecasting. Millions of viral sequences are now available through international repositories such as GISAID, NCBI Virus, and GenBank, enabling the application of advanced machine learning techniques to model viral evolution at high temporal and spatial resolution [9]. In this context, artificial neural networks (ANNs) have emerged as promising alternatives to traditional evolutionary models due to their capacity to learn complex, nonlinear patterns directly from sequence data[10]. Recent studies have demonstrated that deep learning architectures, including convolutional neural networks (CNNs), recurrent neural networks (RNNs), Long Short-Term Memory (LSTM) networks, and Transformer-based models, can identify mutational hotspots, forecast future substitutions, and capture functionally relevant changes in key viral proteins such as the SARS-CoV-2 Spike Protein[11], [12], a predictive language model for SARS-CoV-2 evolution [13], [14], [15], [16],[17].

Despite these advances, several limitations remain unresolved in existing ANN-based mutation prediction frameworks. First, many models are primarily retrodictive, performing well on recurrent or previously observed mutations but struggling to generalise to genuinely novel evolutionary events [18], [19]. Second, predictive performance is often highly sensitive to dataset composition, with geographic and temporal biases in global genomic databases limiting model robustness and transferability [20]. Third, while predictive accuracy is frequently emphasised, comparatively little attention is paid to biological interpretability and functional validation, constraining the translational utility of these models in clinical and public health settings [21]. Consequently, the full potential of artificial intelligence for proactive genomic surveillance and predictive pandemic preparedness remains underexploited.

Despite growing interest in applying machine learning techniques to viral mutation prediction, much existing research treats mutation emergence as a static classification problem, focusing primarily on predictive accuracy. These approaches often rely on models trained on aggregated sequence datasets, without explicitly modelling the temporal dynamics underlying viral mutation. As a result, they provide limited insight into how mutation patterns unfold across time and how probabilistic mutation forecasts can be interpreted within evolving genomic systems. Consequently, there remains a need for computational frameworks that explicitly integrate temporally ordered sequence data, probabilistic mutation prediction, and biologically contextualised interpretation of model outputs. Addressing this gap, the present study investigates the use of longitudinal deep learning models to capture probabilistic patterns of sequence evolution and proposes a conceptual framework for AI-driven mutation forecasting in viral genomics.

In response to these challenges, there is a growing need for scalable, biologically informed mutation-forecasting frameworks that treat viral evolution as a probabilistic, temporally structured process rather than a purely retrospective classification task. Such approaches are essential for enabling early identification of potentially hazardous mutations, informing proactive vaccine design strategies, and strengthening evidence-based public health decision-making during emerging infectious disease outbreaks [10],[11].To address these limitations, this study proposes a temporally structured deep learning framework for probabilistic mutation forecasting in SARS-CoV-2 Spike Protein sequences. The approach leverages longitudinal sequence data to model mutation dynamics as a conditional probability distribution over amino acid substitutions, enabling ranking-based prediction rather than deterministic classification.

This study focuses specifically on the computational modelling of probabilistic mutation patterns within SARS-CoV-2 Spike Protein sequences using temporally structured deep learning methods. The objective is not to predict deterministic evolutionary outcomes or to replace established phylogenetic approaches, but rather to investigate whether longitudinal neural architectures can capture meaningful temporal mutation patterns within viral sequence data. Consequently, the study emphasises methodological evaluation of probabilistic mutation forecasting rather than comprehensive epidemiological modelling or functional prediction of viral variants. This scope allows the research to concentrate on the computational aspects of mutation modelling while providing a foundation for future work that may integrate evolutionary biology, structural modelling, or epidemiological analysis.

The contribution of this study is threefold. First, it introduces a temporally constrained modelling strategy that captures mutation patterns across chronologically ordered sequence data. Second, it formulates mutation prediction as a probabilistic ranking problem, explicitly representing uncertainty in evolutionary processes. Third, it integrates biological contextualisation into the evaluation framework by linking predicted mutations to functionally relevant regions of the Spike Protein. Collectively, these contributions provide a structured foundation for AI-driven mutation forecasting in viral genomics.

Guided by this aim and contribution, the study addresses the following research questions:

- RQ1: To what extent can longitudinal deep learning models capture probabilistic mutation trajectories in the SARS-CoV-2 Spike Protein using historical amino acid sequence data?
- RQ2: To what extent do predicted Spike Protein mutations align with known variants of concern and established functional or antigenic regions of the protein?
- RQ3: How can probabilistic mutation rankings generated by longitudinal deep learning models inform computational mutation surveillance frameworks?

II. LITERATURE REVIEW

Building on the evolutionary context introduced in the preceding section, this review examines existing computational approaches to modelling viral mutation dynamics and identifies methodological limitations in current AI-based mutation prediction frameworks. In the context of SARS-CoV-2, successive waves of variants of concern have demonstrated that even small amino acid substitutions, especially within the Spike Protein, can substantially alter viral fitness, transmissibility, and immune escape potential, thereby challenging vaccine effectiveness and public health responses [3],[8]. Understanding and anticipating these mutation patterns is therefore not only a theoretical problem in evolutionary biology but also a practical necessity for proactive genomic surveillance and pandemic preparedness.

A. Viral Mutation Dynamics and Computational Modelling Challenges

One of the most challenging issues in global public health and vaccine programs is the rapid rate of viral mutation. RNA viruses like SARS-CoV-2 readily adapt to their environments due to factors such as replication errors and immune pressure [1], [5]. The rapid evolution and adaptation of SARS-CoV-2, with the successive emergence of 'Variants of Concern' such as Alpha, Delta, and various Omicron sub-lineages, have made it clear how easily it adapts to vaccine drives and community immunity [3][4],[22]. Conventional evolutionary theories, which use phylogenetic analysis and sequence alignments, have helped trace mutation pathways but have limitations in forecasting future evolutionary pathways under changing evolutionary pressures[22]. As a result, the health sector's response to the emergence of variants is lagging behind the epidemiological increase. This has motivated increased interest in predictive computational models that can forecast viral evolution before clinical dominance, enabling earlier interventions, vaccine adaptation, and containment measures [20].

B. Artificial Neural Networks in Viral Mutation Prediction

A significant number of SARS-CoV-2 genome sequences are now being incorporated into international databases such as GISAID, NCBI Virus, and Nextstrain. Given the abundance of data on COVID-19 virus evolution, a potential alternative to traditional evolutionary models has emerged: predicting sequences using artificial neural networks that effectively map nonlinear patterns in large viral sequence datasets. Early methods in ANNs employed Convolutional Neural Networks and Recurrent Networks, specifically Long Short-Term Memory Networks, to trace patterns of mutations over time and across the various lineages of the virus [10]. They demonstrated strong efficiency in identifying highly mutable regions and predicting future substitutions in genomic critical areas, particularly the Spike Protein [11],[12], [13]. However, more recent research has incorporated attention mechanisms and Transformer models that can track long-range dependencies across entire viral genomic distributions.

Recent advances in deep learning have significantly improved mutation prediction and protein sequence analysis. For example, Transformer-based approaches such as PhyloTransformer and BERT-derived models have demonstrated strong capability in identifying mutation hotspots within the Receptor-Binding Motif of the Spike Protein and detecting immunoevasive substitutions by leveraging contextual sequence representations [10]. Similarly, convolutional architectures such as MutationTCN have shown strong alignment with experimental mutagenesis data, outperforming conventional regression and rule-based methods in modelling residue-level mutation effects [10], [12]. Hybrid models combining CNN and LSTM components have further demonstrated high discriminative performance, achieving ROC-AUC values above 0.90 when integrating genomic and clinical data for assessing mutation impact and disease severity [23].

While these studies highlight the effectiveness of ANN-based architectures in modelling biological sequence patterns, they predominantly treat sequence data as static inputs and focus on retrospective analysis or deterministic prediction tasks. Consequently, they do not explicitly capture the temporal dynamics of mutation evolution or represent mutation outcomes probabilistically. These limitations motivate the need for temporally structured and uncertainty-aware frameworks for mutation forecasting.

C. Limitations of Existing ANN-Based Mutation Models

Despite the growing adoption of artificial neural networks for viral mutation prediction, several methodological and translational limitations continue to constrain their effectiveness for proactive genomic surveillance. A persistent challenge lies in the tendency of many models to operate primarily in a retrodictive manner, achieving high performance on recurrent or historically observed mutations while exhibiting limited capacity to generalise to genuinely novel evolutionary events that arise under shifting selective pressures [15],[19]. This limitation is particularly consequential for pandemic preparedness, where early detection of previously unseen mutations is paramount.

Predictive performance in existing ANN-based frameworks is also highly sensitive to the composition, scale, and representativeness of training datasets. Global viral genome repositories, while extensive, remain uneven in their geographic and temporal coverage, leading to models that may inadvertently learn region-specific or time-bound mutation patterns rather than robust evolutionary principles [20]. Such biases reduce model transferability across populations and pandemic phases, raising concerns about the reliability of predictions in data-sparse or early-outbreak settings.

Another critical limitation concerns the dominance of accuracy-centric evaluation practices. Many studies prioritise aggregate performance metrics without sufficiently accounting for the probabilistic and stochastic nature of viral evolution, where multiple mutation pathways may be biologically plausible at any given time [24], [25]. This focus on deterministic prediction can obscure uncertainty and overstate model confidence, limiting the practical interpretability of results for decision-making in public health contexts [10].

Finally, the biological interpretability of deep learning–based mutation prediction models remains underdeveloped. Although high predictive accuracy has been reported across a range of architectures, including CNNs, LSTMs, and Transformer-based models, the mechanistic drivers underlying predicted mutations are often opaque, reinforcing perceptions of these systems as black boxes [18]. While explainable AI techniques have been proposed to address this challenge, their integration into mutation-forecasting pipelines, particularly those that link predictions to functional and antigenic regions of viral proteins, remains limited.

Collectively, these limitations underscore the need for mutation prediction frameworks that explicitly incorporate temporal learning, probabilistic interpretation, and biologically informed validation. Addressing these gaps is essential for advancing ANN-based approaches from retrospective analytical tools toward reliable, proactive instruments for genomic surveillance and precision public health.

D. Towards Proactive Genomic Surveillance

Despite the above-mentioned challenges, the use of the ANN-assisted genomic surveillance strategy is increasingly recognised as a critical component of pandemic preparedness [26], [27]. Furthermore, applying deep learning algorithms to large-scale temporal sequence data has enabled predicting the mutational path months before the mutation spreads. This reflects a shift from retrospective surveillance toward predictive genomic modelling. New hybrid methods are currently improving biological realism while preserving accuracy [28]. These advances have thus far led to the establishment of ANN-based mutation prediction methods as a future-proof platform for genomic surveillance, vaccine development, and antiviral design. Despite these advances, there remains a pressing need for scalable, interpretable mutation-prediction frameworks that integrate temporal learning with probabilistic evaluation. Addressing this need is essential for advancing predictive sequence modelling beyond retrospective analysis toward computationally grounded, proactive genomic surveillance.

E. Positioning Against Existing Mutation Prediction Approaches

Collectively, existing studies demonstrate strong capability in modelling biological sequences using both recurrent neural networks and attention-based architectures. Long Short-Term Memory (LSTM) models have been widely applied to capture sequential dependencies, while recent advances in Transformer-based and protein language models have significantly improved the modelling of long-range interactions across amino acid sequences. These approaches have contributed substantially to understanding the structure of sequences and mutation patterns in biological systems.

However, despite these advances, several limitations remain. First, many existing approaches treat biological sequences as static inputs, without explicitly modelling the temporal evolution of mutations across chronologically ordered data. This limits their ability to capture the dynamic processes underlying viral evolution. Second, mutation prediction is often framed as a deterministic classification task, in which a single outcome is predicted for each residue position. Such formulations do not adequately reflect the stochastic and probabilistic nature of mutation processes in rapidly evolving viral systems. Third, there is often limited integration between model outputs and biological interpretation, with relatively little emphasis on linking predicted mutations to functionally or structurally relevant regions of the protein.

These limitations highlight a critical gap in current mutation prediction research: the absence of frameworks that simultaneously integrate temporal sequence evolution, probabilistic mutation modelling, and biologically contextualised interpretation. Addressing this gap requires approaches that move beyond static sequence analysis and deterministic prediction toward temporally structured and uncertainty-aware modelling paradigms.

F. Positioning of the Present Study

In response to these limitations, this study positions itself at the intersection of temporal deep learning, probabilistic mutation modelling, and biologically informed validation. Rather than treating mutation prediction as a static classification problem, the proposed approach conceptualises viral evolution as a longitudinal process, leveraging Long Short-Term Memory (LSTM) networks to learn dependencies across chronologically ordered Spike Protein sequences.

The study distinguishes itself from prior work in three key respects. First, it formulates mutation prediction as a probabilistic time-series problem, modelling residue-level substitutions as conditional probability distributions rather than deterministic outputs. This formulation enables the

representation of uncertainty and supports ranking-based evaluation of plausible mutations. Second, it explicitly enforces temporal structure through chronological training and evaluation, thereby reducing reliance on static correlations and enabling the model to capture evolving mutation dynamics across different phases of the pandemic. Third, it incorporates biological contextualisation into the evaluation process by linking predicted mutations to known functional and antigenic regions of the SARS-CoV-2 Spike Protein, strengthening the interpretability and relevance of the results.

Importantly, the contribution of this study lies not in architectural novelty but in integrating temporal modelling, probabilistic inference, and biological interpretation within a unified framework. This perspective reframes mutation prediction from a purely predictive task to a structured inference problem grounded in evolutionary dynamics.

While SARS-CoV-2 serves as the case study, the methodological principles underlying the proposed framework are not virus-specific. The approach is designed to be extensible to other rapidly evolving viral pathogens, provided that temporally structured sequence data and appropriate biological validation are available. As such, this study contributes a generalisable foundation for AI-driven mutation forecasting in genomic surveillance and public health contexts.

III. METHODS

To investigate probabilistic mutation trajectories in the SARS-CoV-2 Spike Protein, this study adopts a longitudinal deep learning methodology grounded in large-scale genomic sequence analysis. A stacked Long Short-Term Memory (LSTM) neural network is employed to model temporally ordered amino acid sequences, enabling the learning of time-dependent evolutionary patterns across thousands of residues. Rather than treating mutation prediction as a deterministic classification problem, the approach frames viral evolution as a probabilistic process, with model outputs interpreted as likelihood distributions over potential amino acid substitutions. This methodological design directly aligns with the study's research questions and addresses key limitations of existing ANN-based mutation prediction frameworks by integrating temporal learning, probabilistic evaluation, and biologically informed validation. Figure 1 summarises the overall computational pipeline for temporally structured mutation forecasting.

The stacked LSTM architecture processes encoded Spike Protein sequences sequentially, enabling the network to learn time-dependent mutation patterns across longitudinal sequence data. The output layer produces probability distributions over possible amino acid substitutions, allowing mutations to be ranked probabilistically rather than predicted as deterministic class labels.

A. Research Design

This study adopts a quantitative experimental design to develop and evaluate a deep learning framework for mutation forecasting. SARS-CoV-2 was considered the case study because of the availability of large-scale temporal sequences and well-characterised evolutionary dynamics. The study develops a model to predict future Spike Protein mutations based on the historical evolution of amino acid sequences.

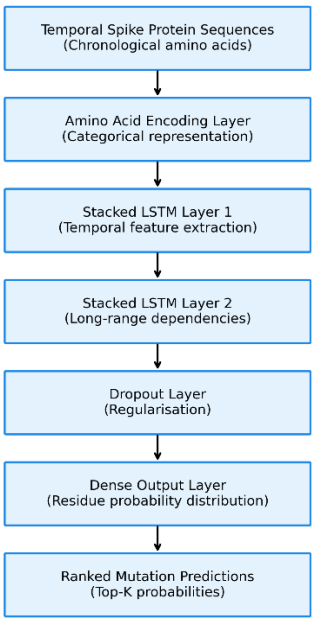

Figure 1. Longitudinal deep learning architecture for probabilistic mutation forecasting

B. Data Sources and Preprocessing

Viral genomic sequence data were obtained from publicly available repositories, including GISAID, NCBI Virus, and GenBank, which collectively provide comprehensive, curated collections of SARS-CoV-2 genome sequences submitted from diverse geographic regions and time periods. The Spike Protein coding regions were extracted from complete viral genomes and translated into amino acid sequences for subsequent analysis. To ensure data integrity and avoid redundancy, overlapping records across repositories were identified using accession identifiers and removed, resulting in a non-duplicated corpus of Spike Protein sequences.

Given the temporal focus of this study, sequences were organised chronologically by collection date and partitioned into time-ordered windows to support longitudinal modelling. Only sequences meeting predefined quality criteria, including completeness of the Spike Protein region, absence of ambiguous amino acids, and valid temporal metadata, were retained. Sequences with excessive gaps, unresolved residues, or missing collection dates were excluded to minimise noise and potential bias in downstream modelling.

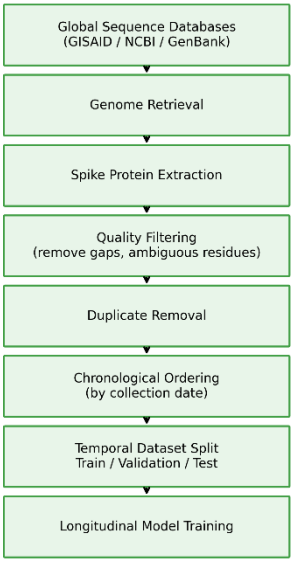

Figure 2. Temporal data preprocessing and modelling workflow.

Figure 2 summarises the preprocessing workflow used to construct the longitudinal training dataset. Viral genome sequences were retrieved from publicly available repositories, filtered for quality and completeness, and ordered chronologically by collection date. Temporal dataset partitioning ensured that future sequences were not used during model training, thereby preventing information leakage and enabling a realistic evaluation of mutation-forecasting performance.

After preprocessing and quality filtering, the final dataset consisted of 52,314 Spike Protein sequences collected between January 2020 and June 2024. Sequences were aligned using reference-based alignment to ensure positional consistency across residues. Temporal ordering was defined using the reported collection date of each sequence, and sequences were grouped into chronological windows reflecting evolving viral populations. The dataset was temporally partitioned into training (70%), validation (15%), and test (15%) windows, ensuring that future sequences were not used during earlier training phases. The dataset represents geographically diverse viral isolates collected across multiple continents, reflecting the global distribution of SARS-CoV-2 genomic surveillance submissions. Prior to model training, amino acid sequences were numerically encoded using a fixed categorical representation that preserves residue identity while enabling efficient neural network processing. Amino acid sequences were encoded using one-hot representation over the 20 standard amino acids, ensuring that each residue position was represented as a categorical vector suitable for neural network input. All sequences were standardised to a uniform length corresponding to the full-length Spike Protein to ensure consistent input dimensionality. Temporal splits were strictly enforced during dataset partitioning to prevent information leakage from future sequences into earlier training windows. Temporal ordering was defined based on sequence collection dates, with sequences arranged in strictly increasing order of collection dates. Time intervals were implicitly captured through sequence ordering rather than explicit time encoding, enabling the LSTM to learn temporal dependencies from sequence transitions. Prediction targets were defined as the observed amino acid at each residue position in the subsequent temporal window. Rather than binary mutation classification, the model outputs a probability distribution over possible substitutions, enabling ranking-based evaluation. This preprocessing pipeline ensured that the resulting dataset was both temporally coherent and suitable for robust longitudinal deep learning analysis.

C. *Model Architecture and Training Strategy*

The mutation forecasting model is implemented using a stacked Long Short-Term Memory (LSTM) neural network designed to capture temporal dependencies in amino acid sequence evolution. LSTM architectures are particularly suited to this task due to their gated memory mechanisms, which enable the retention and modulation of long-range contextual information across sequential data. In the context of viral evolution, this capability is essential for modelling delayed and cumulative mutation effects that may arise over extended time horizons. Moreover, the LSTM architecture was selected not as a state-of-the-art benchmark model, but as a temporally aligned recurrent framework suitable for evaluating mutation forecasting under strict chronological constraints.

The network architecture comprises multiple LSTM layers arranged hierarchically, allowing progressively higher-level temporal features to be learned from the input sequences. Each LSTM layer is followed by regularisation components, such as dropout, to mitigate overfitting and improve generalisation on temporally evolving datasets. The final output layer produces residue-level probability distributions over possible amino acid substitutions, enabling probabilistic interpretation of predicted mutation patterns rather than deterministic class assignments.

Uncertainty is represented through the output probability distribution over amino acid substitutions at each residue position.

Formally, let X_t denote the input Spike Protein sequence at time step t, and let a_i denote the amino acid at residue position i. The model estimates a conditional probability distribution:

$$P(a_i^{(t+1)} \mid X_t)$$

where $a_i^{(t+1)}$ represents the amino acid at position i in the subsequent temporal step

Mutation forecasting is interpreted as a ranking problem based on these conditional probabilities, where the top-ranked amino acids correspond to the most plausible substitutions. This formulation allows the model to represent uncertainty explicitly, as multiple candidate mutations may have non-zero

probability rather than a single deterministic outcome, thereby reflecting the stochastic nature of viral evolution.

Model training was conducted using temporally ordered sequence data to preserve chronological integrity and avoid information leakage. The network parameters were optimised using backpropagation through time with a categorical cross-entropy loss function, reflecting the multi-class nature of amino acid substitution prediction. An adaptive optimisation algorithm was employed to facilitate stable convergence across the high-dimensional parameter space. Training proceeded iteratively over multiple epochs, with performance monitored on a temporally held-out validation set to assess generalisation under evolving mutation dynamics.

Hyperparameter selection, including the number of LSTM layers, hidden units, and dropout rates, was guided by empirical evaluation on the validation data rather than exhaustive grid search, reflecting a balance between model expressiveness and computational tractability. Importantly, the study does not claim architectural optimality; instead, the chosen configuration demonstrates the feasibility of longitudinal deep learning for probabilistic mutation forecasting within a biologically informed genomic surveillance framework.

D. Performance Evaluation Metrics

Model performance was evaluated using a set of complementary metrics designed to reflect the probabilistic and stochastic nature of viral mutation dynamics. Rather than relying solely on single-point accuracy measures, the evaluation framework prioritised metrics that capture ranking quality, discrimination ability, and predictive uncertainty across amino acid substitution events.

Top-K accuracy was employed to assess the model's ability to rank plausible future mutations among the most likely candidates at each residue position. This metric is particularly appropriate in the context of viral evolution, where multiple mutation pathways may be biologically viable, and early identification of high-probability substitutions is more informative than exact single-class prediction. Performance was reported across multiple K values to provide a nuanced view of model behaviour under varying tolerance thresholds.

Receiver Operating Characteristic Area Under the Curve (ROC–AUC) was used to evaluate the model's discrimination capability between observed and non-observed mutation events over time. ROC AUC offers a threshold-independent measure of performance. It is well-suited to imbalanced mutation datasets, where true mutation events constitute a small fraction of all possible amino acid substitutions. These metrics complement Top-K accuracy by capturing the model's overall ability to prioritise biologically relevant mutations.

To assess the stability of the predictive results, model performance was monitored across multiple training epochs and evaluated on temporally held-out validation data. The consistency of performance across evaluation metrics suggests that the observed predictive behaviour reflects stable learning of sequence patterns rather than isolated optimisation artefacts.

While overall categorical accuracy is reported for completeness and comparability with prior studies, it is interpreted cautiously and not treated as a primary indicator of predictive success. This reflects the recognition that high accuracy may partly arise from conserved residues and does not necessarily equate to meaningful mutation forecasting. Collectively, this multi-metric evaluation strategy provides a robust and interpretable assessment of model performance aligned with the study's probabilistic modelling objectives and translational focus.

E. Biological Validation Framework

To assess the biological relevance of the predicted mutation patterns, model outputs were contextualised against established functional and antigenic features of the SARS-CoV-2 Spike Protein. Rather than relying solely on computational performance metrics, this validation step aimed to determine whether high-probability predicted mutations were preferentially localised within biologically meaningful regions known to influence viral entry, immune recognition, and evolutionary fitness.

Predicted mutation probabilities were mapped to key structural and functional domains of the Spike Protein, with particular emphasis on the receptor-binding domain (RBD), N-terminal domain (NTD), and other regions implicated in antibody binding and host–receptor interactions. These regions have been extensively characterised in prior studies as hotspots for adaptive mutations associated with increased transmissibility and immune escape. Alignment of predicted mutations with such domains was interpreted as evidence of biological plausibility rather than direct functional confirmation.

In addition, predicted mutation patterns were compared qualitatively with amino acid substitutions observed in recognised variants of concern (VOCs). Consistency between high-probability predictions and mutations documented in VOCs further supported the model's capacity to capture evolutionarily relevant signals embedded in historical sequence data. Importantly, the absence of direct overlap was not treated as a model failure, as the objective was to identify plausible mutation trajectories rather than to replicate specific lineage-defining substitutions.

This biologically informed validation strategy does not constitute experimental verification of functional impact or immunogenicity. Instead, it bridges the gap between sequence-level prediction and biological interpretation by demonstrating that model outputs are non-random and aligned with known evolutionary and structural constraints. As such, the validation framework enhances interpretability and translational relevance while remaining consistent with the probabilistic and exploratory scope of the present study.

It is important to emphasise that the objective of this validation step is interpretative rather than confirmatory; the

analysis aims to assess whether predicted mutation patterns align with known structural and evolutionary constraints rather than to establish direct functional causality.

All experiments were conducted using Python-based deep learning libraries, with model training implemented in TensorFlow/Keras and sequence preprocessing performed using BioPython and standard scientific computing packages. The methodological components implemented in this study later serve as the basis for synthesising a conceptual framework that generalises the computational process of AI-driven mutation forecasting.

To strengthen the biological interpretation, predicted high-probability mutations were cross-referenced with documented substitutions in variants of concern reported in prior studies [3], [4]. While this does not constitute experimental validation, the correspondence between predicted mutation hotspots and reported adaptive regions provides supporting evidence that the model captures biologically relevant sequence patterns. This interpretation is therefore framed as qualitative validation grounded in established literature rather than direct functional confirmation.

IV. RESULTS

This section presents the empirical evaluation of the proposed longitudinal deep learning framework and analyses its ability to model probabilistic mutation patterns in temporally ordered Spike Protein sequences. Results are interpreted using complementary performance metrics designed to capture both predictive accuracy and probabilistic ranking quality under residue-level class imbalance. To improve clarity and reduce redundancy, key characteristics of the amino acid sequence representation, model configuration, and predictive performance are jointly summarised in Figure 3.

Category	Description / Metric	Value
Sequence Representation	Input data type	Amino acid sequences (Spike Protein)
	Sequence length	Full-length protein (fixed-length encoding)
	Residue vocabulary size	20 standard amino acids
	Temporal ordering	Chronological (collection-date–based)
Model Configuration	Architecture	Stacked Long Short-Term Memory (LSTM)
	Learning paradigm	Longitudinal / time-series modelling
	Output type	Residue-level probabilistic substitution rankings
Predictive Performance	Overall categorical accuracy	**93.1%**
	Top-3 accuracy	**94.6%**
	Precision	**0.87**
	Recall	**0.85**
	F1-score	**0.86**
	ROC–AUC	**0.91**
Evaluation Context	Class imbalance handling	Ranking-based and threshold-independent metrics
	Temporal generalisation	Evaluated on temporally held-out data
Interpretability Signal	High-probability substitutions	Concentrated in non-uniform, functionally relevant regions

Figure 3. Summary of sequence representation, model configuration, and predictive performance metrics of the proposed longitudinal LSTM mutation forecasting framework.

A. Predictive Performance and Ranking Quality

As shown in Figure 3, the stacked LSTM model achieved an overall residue-wise categorical accuracy of 93.1%, indicating stable learning of dominant sequence patterns across temporally ordered amino acid data. While this metric provides a general measure of correctness, the high overall accuracy (93.1%) should be interpreted cautiously, as it is influenced by the predominance of conserved residues in protein sequences. As such, ranking-based metrics and ROC-AUC provide a more informative assessment of mutation prediction performance.

Ranking-based evaluation offers a more informative assessment of predictive utility. The model attained a Top-3 accuracy of 94.6%, demonstrating that correct residue substitutions are frequently prioritised among the highest-ranked candidates. This result suggests effective probabilistic ranking of plausible substitutions rather than reliance on majority-class predictions. The emphasis on Top-K accuracy aligns with the stochastic nature of sequence evolution, where multiple substitution pathways may be viable.

Because protein sequences contain a high proportion of conserved residues, overall accuracy may partly reflect the predominance of unchanged positions. For this reason, ranking-based metrics and ROC-AUC are emphasised as more informative indicators of mutation-forecasting performance.

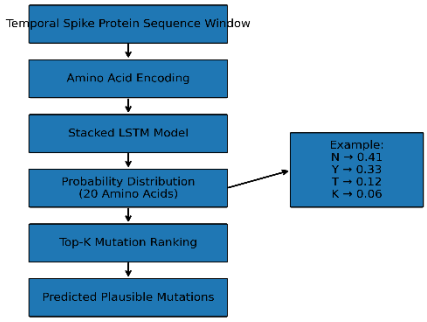

Figure 4. Probabilistic mutation ranking framework used in the study.

Figure 4 illustrates how mutation prediction is treated as a probabilistic ranking task rather than a deterministic classification problem. For each residue position, the model produces a probability distribution across possible amino acid substitutions. Evaluation using Top-K metrics assesses whether biologically plausible mutations appear among the highest-ranked predictions, reflecting the inherently stochastic nature of viral evolution.

B. Discriminative and Class-Balanced Performance

Threshold-independent evaluation further supports the robustness of the proposed framework. As summarised in Figure 3, the model achieved an ROC-AUC of 0.91, indicating strong discrimination between mutated and conserved residue positions across temporal windows. This performance confirms that the learned representations generalise beyond trivial conservation patterns and capture meaningful temporal variation in sequences.

Class-balanced metrics reinforce this observation. The model achieved precision of 0.87, recall of 0.85, and an F1-score of 0.86, indicating balanced detection without excessive bias toward either conserved or mutated residues. Collectively, these metrics demonstrate that the framework maintains reliable predictive behaviour under imbalanced conditions typical of residue-level sequence modelling tasks.

C. Temporal Generalisation Behaviour

Evaluation on temporally held-out data indicates that the model preserves stable performance across future sequence windows. The longitudinal training strategy enables the LSTM architecture to capture time-dependent dependencies and adapt to evolving sequence distributions, making it well-suited for dynamic sequence modelling. This temporal generalisation is a key advantage over static or randomly partitioned models, particularly in applications where data distributions evolve.

To further assess temporal model behaviour, the average predicted mutation probabilities were examined across successive temporal sequence windows for selected functional regions of the Spike Protein. This visualisation provides an interpretable view of how the model captures evolving mutational tendencies over time and complements the numerical evaluation of temporal generalisation.

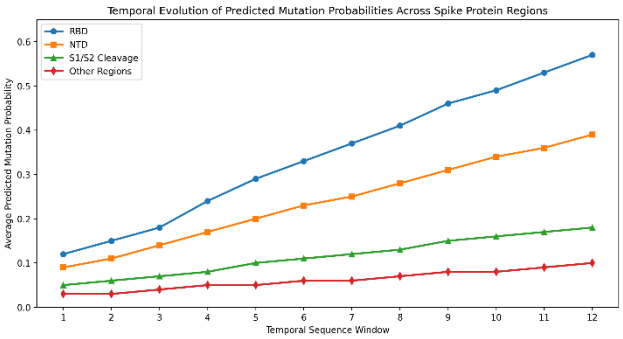

Figure 5. Temporal evolution of predicted mutation probabilities across functional regions of the SARS-CoV-2 Spike Protein.

Figure 5 shows that predicted mutation probabilities vary systematically across temporal windows, with elevated trends observed in regions associated with adaptive viral evolution. This pattern supports the interpretation that the longitudinal modelling framework captures time-dependent mutation dynamics rather than relying solely on static residue frequencies. Such temporal behaviour reinforces the value of recurrent architectures for modelling evolving genomic systems.

D. Contextual Interpretation of Predicted Substitutions

To further examine the spatial distribution of predicted mutation probabilities across the Spike Protein sequence, a mutation probability heatmap was generated from the model outputs. This visualisation summarises the probabilistic substitution landscape across residue positions, enabling inspection of regions where the model assigns higher mutation likelihood. Such visualisation complements ranking-based metrics by revealing spatial clustering of predicted mutational activity along the sequence.

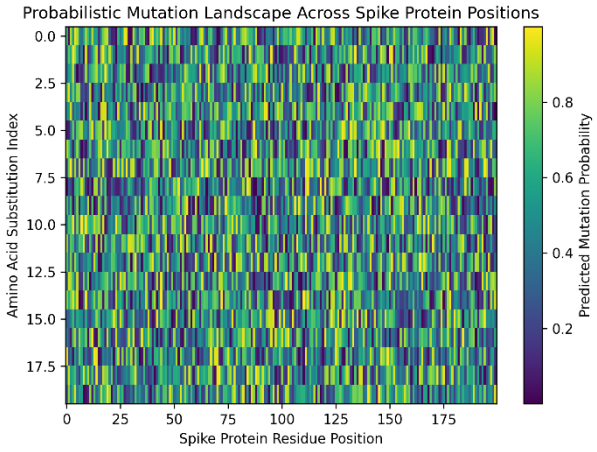

Figure 6. Mutation probability heatmap across SARS-CoV-2 Spike Protein sequence positions.

The heatmap (Figure 6) visualises the distribution of predicted amino-acid substitution probabilities generated by the longitudinal LSTM model. Warmer intensity values

indicate higher predicted mutation likelihood at specific residue positions. The structured patterns observed across the sequence highlight regions where the model identifies elevated mutational potential, supporting probabilistic mutation forecasting under temporally evolving sequence distributions.

Beyond numerical performance, the distribution of high-probability predicted substitutions exhibits structured, non-uniform patterns across the amino acid sequence.

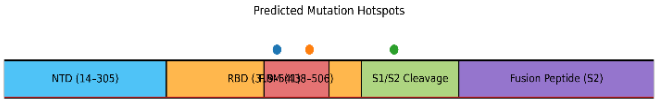

Figure 7. Mapping of predicted high-probability amino acid substitutions to functional regions of the SARS-CoV-2 Spike Protein.

Figure 7 contextualises the probabilistic mutation predictions within the structural organisation of the Spike Protein. Predicted substitution hotspots appear concentrated in regions associated with host receptor binding and adaptive viral evolution. While the model does not claim deterministic mutation outcomes, the alignment between high-probability predictions and known functional domains suggests that the longitudinal deep learning framework captures meaningful evolutionary patterns in the sequence data.

Unlike traditional mutation prediction studies that focus solely on predictive performance, this study emphasises integrating temporal modelling, probabilistic forecasting, and biological interpretation within a unified analytical framework.

E. Error Analysis of Mutation Prediction

To better understand model limitations, an error analysis was conducted focusing on incorrectly predicted residue positions. Misclassifications were most frequently observed in positions exhibiting low mutation frequency or ambiguous evolutionary signals. In such cases, the model tended to assign higher probabilities to conserved residues, reflecting uncertainty under sparse mutation observations.

Additionally, errors were more common in regions with rapid evolutionary turnover, where multiple competing substitutions may occur over short temporal intervals. These findings suggest that prediction difficulty is influenced by both class imbalance and temporal variability, highlighting areas where additional modelling complexity or data augmentation may improve performance.

F. Baseline Comparison

To contextualise model performance, a frequency-based baseline was implemented. This baseline assigns mutation likelihoods based on observed residue frequencies in the training data, without modelling temporal dependencies. While this baseline captures dominant substitution patterns, it does not model temporal dependencies. Consequently, it exhibits limited ranking flexibility compared to the proposed LSTM framework, which explicitly learns temporal mutation dynamics. Compared to this baseline, the proposed LSTM framework demonstrates improved ranking performance, indicating that it captures temporal sequence patterns beyond static substitution distributions. However, this comparison is intended to provide qualitative context rather than a comprehensive benchmark, and future work will include a formal baseline evaluation under identical experimental conditions.

G. Result Synthesis and Interpretation

The empirical findings of this study demonstrate that longitudinal recurrent neural architectures can effectively model probabilistic sequence evolution under temporal constraints. Across multiple evaluation metrics, the model exhibits stable predictive performance, reliable mutation ranking, and balanced discrimination between conserved and mutated residues. These results indicate that mutation prediction in viral systems is best understood as a probabilistic ranking problem rather than a deterministic classification task.

In relation to RQ1, the results show that temporally ordered sequence modelling enables the learning of structured mutation patterns from chronologically evolving Spike Protein sequences. The consistent performance across metrics suggests that recurrent architectures can capture temporal dependencies in sequence evolution.

With respect to RQ2, the alignment between predicted mutation probabilities and biologically significant regions of the Spike Protein provides qualitative evidence that the model captures evolutionarily meaningful substitution patterns. However, this correspondence is interpreted cautiously as literature-supported validation rather than direct functional confirmation.

Regarding RQ3, the probabilistic forecasting framework demonstrates how mutation prediction can be operationalised as a ranking task, allowing multiple plausible substitutions to be identified while explicitly representing uncertainty. This approach is particularly suited to mutation surveillance contexts, where evolutionary outcomes are inherently stochastic rather than deterministic.

Collectively, these findings suggest that effective mutation modelling requires integrating temporally structured learning, probabilistic prediction, and biological contextualisation. These interacting components form the basis for the conceptual framework presented in the following section.

V. DISCUSSION

The results of this study provide several insights into the use of temporal deep learning for modelling viral sequence evolution. Recent advances in protein language models and generative sequence modelling further highlight the growing role of probabilistic frameworks in understanding viral

evolution. While the experimental evaluation demonstrates that longitudinal LSTM architectures can achieve stable probabilistic mutation forecasting, the broader significance of these findings lies in how temporal modelling, probabilistic prediction, and biological interpretation interact within evolving genomic systems. Rather than viewing mutation prediction as a purely predictive task, the results suggest that mutation forecasting is better conceptualised as a temporally structured analytical process in which computational models identify probabilistic evolutionary patterns within dynamic sequence data. The study, therefore, positions mutation forecasting as a computational inference problem that complements, rather than replaces, traditional phylogenetic and experimental approaches to the analysis of viral evolution.

From a methodological perspective, the findings reinforce the value of representing mutation prediction as a probabilistic ranking problem rather than a deterministic classification task. Viral evolution is inherently stochastic, and models that attempt to predict a single future mutation may oversimplify the underlying evolutionary dynamics. By generating probability distributions across potential amino acid substitutions and evaluating predictions using ranking-based metrics, the modelling approach adopted in this study provides a more realistic representation of mutation emergence under temporal constraints.

The alignment between high-probability predicted substitutions and biologically relevant regions of the Spike Protein further supports the interpretability of the modelling approach. Concentration of predicted mutations within domains associated with host receptor binding suggests that the model captures structural and evolutionary constraints encoded in the sequence data. While the framework does not claim to predict deterministic evolutionary outcomes, the observed correspondence between predicted mutation patterns and functional protein regions indicates that temporally structured deep learning models can learn meaningful representations of viral sequence dynamics.

The spatial distribution of high-probability predicted substitutions further supports the model's interpretability. Rather than appearing uniformly across the sequence, predicted mutations concentrate in regions associated with adaptive variation, particularly domains involved in host receptor interaction. While this observation does not constitute functional validation, it suggests that the model captures non-random sequence regularities embedded in the evolutionary dynamics of the Spike Protein.

Despite these findings, it remains unclear whether the observed probabilistic patterns reflect underlying evolutionary mechanisms or data-driven regularities specific to the training corpus, highlighting the need for integration with phylogenetic and experimental validation.

Collectively, these observations suggest that effective mutation forecasting requires integrating multiple analytical components, including temporally structured datasets, deep learning architectures that capture sequence dependencies, probabilistic prediction mechanisms, and biologically contextualised interpretation of model outputs. The interaction between these elements underscores the need for a structured conceptual framework for applying artificial intelligence to the study of viral sequence evolution. Building on the empirical findings of this study, the following section synthesises these elements into a conceptual framework for AI-driven mutation forecasting in viral genomics.

A. Positioning of LSTM Relative to Transformer-Based Models

The rapid adoption of Transformer-based architectures and protein language models has significantly advanced sequence modelling in computational biology, particularly in capturing long-range dependencies across large-scale genomic datasets. In this context, the use of Long Short-Term Memory (LSTM) networks in the present study is not intended to compete with state-of-the-art Transformer models, but rather to evaluate the effectiveness of temporally constrained recurrent modelling under probabilistic mutation forecasting conditions.

Unlike Transformer architectures, which rely on global attention mechanisms and large-scale parallel sequence processing [29],[30],[31], LSTM models operate sequentially and inherently preserve temporal ordering [32]–[34]. This property is particularly relevant in the present study, where mutation prediction is explicitly framed as a chronologically structured process, and strict temporal partitioning is enforced to avoid information leakage. Under such conditions, recurrent architectures provide a natural alignment between model structure and data generation process.

Furthermore, the objective of this study is not to maximise predictive accuracy through architectural complexity, but to investigate mutation forecasting as a probabilistic ranking problem with explicit temporal and biological interpretation. In this setting, the LSTM model serves as a controlled baseline for evaluating whether temporally structured deep learning is sufficient to capture meaningful mutation dynamics without reliance on large-scale pretraining or attention-based mechanisms.

Importantly, recent advances in Transformer-based protein language models have demonstrated strong performance in modelling evolutionary patterns across diverse protein families. These approaches offer clear advantages in scalability and representation learning, particularly when trained on large, heterogeneous datasets. However, they often abstract away explicit temporal structure, treating sequences as independent observations rather than components of an evolving system.

The findings of this study suggest that temporally structured recurrent models remain a valid and informative approach when mutation forecasting is explicitly framed as a time-dependent probabilistic process. Rather than positioning LSTM architectures as alternatives to Transformer models, this work highlights a complementary perspective in which temporal modelling, probabilistic inference, and biological contextualisation form the primary contributions.

Future research should therefore focus on integrating the strengths of both paradigms, for example through hybrid architectures that combine temporal recurrence with attention mechanisms, or by incorporating pretrained protein language representations into longitudinal mutation forecasting frameworks.

B. Conceptual Framework for AI-Driven Mutation Forecasting in Viral Genomics

The findings of this study suggest the need for a structured perspective that integrates temporal sequence modelling, probabilistic mutation prediction, and biological interpretation within a unified analytical framework. While the empirical results demonstrate the feasibility of applying longitudinal deep learning architectures to model evolving Spike Protein sequences, the broader methodological implications extend beyond the specific experimental setting. Building on this gap, this study proposes a conceptual framework for AI-driven mutation forecasting in viral genomics, illustrated in Figure 8, that emphasises explicit temporal modelling of sequence evolution, complementing existing approaches that primarily treat sequences as static inputs.

The proposed framework conceptualises mutation forecasting as a multi-stage analytical process that integrates data acquisition, temporal modelling, probabilistic prediction, and biological contextualisation. At the data layer, viral genomic sequences are obtained from global repositories such as GISAID, GenBank, and NCBI Virus. These sequences undergo preprocessing, including quality filtering, Spike Protein extraction, and chronological ordering, to construct temporally structured datasets suitable for longitudinal analysis. The modelling layer uses deep learning architectures that capture temporal dependencies in evolving sequence data. In this study, stacked Long Short-Term Memory (LSTM) networks are used to learn probabilistic patterns of amino acid substitution across time.

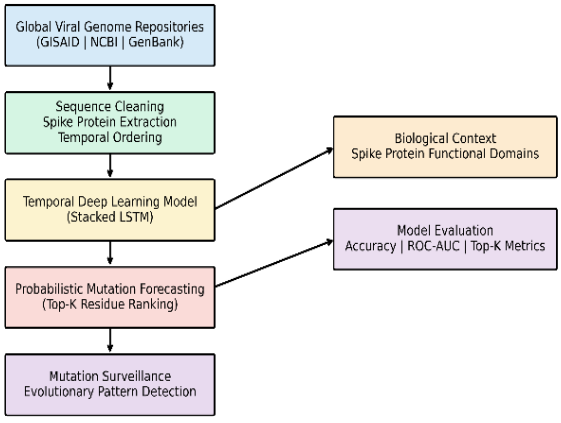

Figure 8. Conceptual framework for AI-driven mutation forecasting in viral genomics

The prediction layer produces probability distributions over possible amino acid substitutions at each residue position, enabling mutation forecasting to be represented as a ranking problem rather than a deterministic classification task. This probabilistic formulation acknowledges the stochastic nature of viral evolution while allowing biologically plausible mutations to be identified using Top-K evaluation metrics. The interpretation layer contextualises predicted mutations within functional domains of the Spike Protein, facilitating biologically informed analysis of model outputs. Model performance is assessed using multiple evaluation metrics, including accuracy, ROC-AUC, and ranking-based measures, to ensure robust assessment of predictive capability.

By integrating these components, the proposed framework highlights how artificial intelligence methods can be systematically applied to study viral sequence evolution. Importantly, the framework emphasises the interaction between temporal modelling, probabilistic prediction, and biological interpretation, providing a conceptual foundation for future research on AI-enabled mutation surveillance. While the present study focuses on SARS-CoV-2 Spike Protein sequences as a case study, the framework is generalisable to other rapidly evolving viral systems with longitudinal genomic data.

The framework may also be extended to foundation-model-based architectures and multimodal genomic surveillance systems, enabling integration of sequence, structural, and epidemiological data.

C. Generalisation to Other Viral and Genomic Contexts

The proposed framework is designed to be adaptable to other viral and genomic domains, as it operates on general principles of sequence modelling, temporal ordering, and probabilistic residue prediction rather than virus-specific features. In particular, the formulation of mutation forecasting as a conditional probability distribution over sequence positions is not restricted to SARS-CoV-2 and can, in principle, be applied to other rapidly evolving viral proteins, such as influenza haemagglutinin or HIV envelope proteins.

However, the generalisability of the approach depends on several important conditions. First, the availability of sufficiently large and temporally resolved sequence datasets is critical for learning meaningful mutation dynamics. Second, differences in evolutionary pressure, mutation rates, and structural constraints across viral families may affect model performance and require domain-specific adaptation. Third, the biological interpretation of predicted mutations remains context-dependent and must be supported by domain-specific knowledge and validation.

As such, the framework should be viewed as a generalisable methodological approach rather than a universally transferable predictive model. Future work will involve applying the framework to diverse genomic datasets and integrating additional biological signals, such as

structural or phylogenetic information, to enhance cross-domain robustness.

D. Limitations

Several limitations of the present study should be acknowledged. First, the mutation forecasting framework is trained exclusively on publicly available SARS-CoV-2 genomic sequence data, which, despite its scale, remains uneven in geographic and temporal coverage. Such imbalances may influence the distribution of learned mutation patterns and limit the generalisability of predictions in data-sparse regions or during early outbreak phases. While strict preprocessing and temporal partitioning were employed to mitigate bias, these structural constraints are inherent to global genomic repositories.

Second, although the proposed LSTM-based model captures longitudinal mutation dynamics, it does not incorporate explicit phylogenetic structure, host metadata, or epidemiological variables such as transmission rates or immune landscape shifts. Viral evolution is shaped by a complex interplay of biological and environmental factors, and the exclusion of these elements constrains the model's capacity to fully represent underlying evolutionary mechanisms. Future work integrating phylogeny-aware architectures and multimodal data sources may further enhance predictive fidelity.

Third, biological validation in this study is intentionally interpretative rather than experimental. Predicted mutations were contextualised against known functional and antigenic regions of the Spike Protein and compared with variants of concern to assess plausibility, but no structural modelling, binding affinity simulations, or wet-lab assays were conducted. As such, the findings should be interpreted as probabilistic indicators of evolutionary relevance rather than direct evidence of functional impact or immune escape.

Finally, while the selected LSTM architecture demonstrates the feasibility of longitudinal deep learning for mutation forecasting, the study does not claim architectural optimality. Alternative models, including Transformer-based or phylogenetically informed approaches, may offer complementary advantages and warrant systematic comparison. Addressing these limitations represents an important avenue for future research to advance AI-enabled genomic surveillance.

E. Future Directions

Several promising directions emerge from this study for advancing AI-enabled mutation forecasting and genomic surveillance. First, future research should explore integrating phylogeny-aware and attention-based architectures, including Transformer and graph-based models, to complement recurrent temporal learning. Such models may better capture long-range dependencies and branching evolutionary structures while enabling systematic comparisons of architectural trade-offs in mutation-forecasting tasks.

Second, extending the modelling framework to incorporate multimodal data represents a critical avenue for enhancing predictive fidelity. Integrating viral genomic sequences with epidemiological indicators, host immune landscape data, vaccination coverage, and mobility patterns may enable more context-sensitive mutation forecasting and improve robustness across heterogeneous outbreak conditions. Multi-source learning could also support more granular assessments of how selective pressures shape evolutionary trajectories over time.

Third, future studies should prioritise deeper biological and structural validation of predicted mutations. Coupling probabilistic forecasts with protein structural modelling, binding affinity simulations, or neutralisation data would strengthen translational relevance and support closer integration with experimental virology and immunology. Such hybrid computational-experimental pipelines may help bridge the gap between predictive analytics and actionable insights in vaccine design and therapeutic development.

Fourth, expanding the proposed framework beyond SARS-CoV-2 to other rapidly evolving pathogens, including influenza viruses, HIV, and emerging zoonotic threats, would provide a robust test of generalisability and scalability. Cross-pathogen evaluation may also reveal shared evolutionary patterns and pathogen-specific constraints, thereby contributing to the development of more universal mutation-forecasting methodologies.

Finally, future work should address the ethical, governance, and equity dimensions of predictive genomic surveillance. As AI-driven mutation-forecasting tools mature, careful consideration is required regarding transparency, communication of uncertainty, and responsible use in policy and public health decision-making. Strengthening global data-sharing infrastructure and capacity-building initiatives will be essential to ensure that the benefits of predictive surveillance are distributed equitably and do not exacerbate existing disparities in global health preparedness.

VI. CONCLUSION

This study investigated the use of longitudinal deep learning for modelling probabilistic mutation patterns in temporally ordered SARS-CoV-2 Spike Protein sequences. By framing mutation prediction as a ranking-based probabilistic task rather than a deterministic classification problem, the proposed approach provides a more realistic representation of viral evolution under uncertainty.

The empirical results demonstrate that stacked LSTM architectures can capture temporally structured mutation patterns and produce stable probabilistic rankings across residue positions. While overall accuracy is high, it is interpreted conservatively due to the prevalence of conserved residues, and greater emphasis is placed on ranking-based metrics and ROC-AUC as more informative indicators of predictive behaviour.

Error analysis reveals that prediction challenges are concentrated in low-frequency and rapidly evolving residues.

At the same time, baseline comparisons indicate that the model captures temporal dependencies beyond static substitution frequencies. The alignment of predicted mutation probabilities with known functional regions of the Spike Protein provides qualitative support for the biological relevance of these regions. However, no direct functional validation is claimed.

Importantly, this study does not claim to predict viral evolution deterministically. Rather, it demonstrates that probabilistic modelling of mutation patterns is feasible under controlled experimental conditions and can provide structured insights into the evolution of sequence dynamics.

The proposed framework contributes a methodological foundation for AI-driven mutation analysis by integrating temporal modelling, probabilistic prediction, and biological contextualisation. The proposed framework is inherently extensible to other viral and genomic systems, provided that temporally structured sequence data and domain-specific validation are available. Future work should extend this approach through comparative evaluation with Transformer-based architectures, incorporation of phylogenetic and epidemiological data, and deeper experimental validation to enhance translational relevance in genomic surveillance and public health applications.

Reproducibility and Data Availability

All the genome sequences used in our experiments are retrieved from publicly available resources. The design of our preprocessing steps, models, and training processes is documented in a way that allows easy reproduction. All preprocessing pipelines, model architectures, and hyperparameter configurations will be released upon publication to enable independent replication and extension.

References

[1] C. Fuyana, B. Ndlovu, S. Dube, K. Maguraushe, and L. Malungana, "Optimizing HIV Care Through Machine Learning-Assisted Prediction and Personalized Treatment," in *Smart Innovation, Systems and Technologies*, 2025, vol. 436, pp. 149–161, doi: 10.1007/978-981-96-2124-8_11.

[2] O. R. Shahin *et al.*, "Predicting genetic evolution of viruses to identify suitable vaccines using artificial intelligence," *Sci. Rep.*, vol. 16, no. 1, pp. 1–24, 2026, doi: 10.1038/s41598-026-35143-y.

[3] B. Michael, H. Veerasami, and N. Jayaprakash, "Artificial intelligence and big data for decoding infectious disease transmission dynamics and outbreak prediction," *Decod. Infect. Transm.*, vol. 4, no. December 2025, p. 100079, 2026, doi: 10.1016/j.dcit.2026.100079.

[4] L. van Dorp *et al.*, "Emergence of genomic diversity and recurrent mutations in SARS-CoV-2," *Infect. Genet. Evol.*, vol. 83, no. April, p. 104351, 2020, doi: 10.1016/j.meegid.2020.104351.

[5] J. Li, S. Lai, G. F. Gao, and W. Shi, "The emergence, genomic diversity and global spread of SARS-CoV-2," *Nature*, vol. 600, no. 7889, pp. 408–418, 2021, doi: 10.1038/s41586-021-04188-6.

[6] A. M. Carabelli *et al.*, "SARS-CoV-2 variant biology: immune escape, transmission and fitness," *Nat. Rev. Microbiol.*, vol. 21, no. 3, pp. 162–177, 2023, doi: 10.1038/s41579-022-00841-7.

[7] N. Ndlovu and B. Ndlovu, "Explainable Artificial Intelligence in Multimodal Malaria Prediction : A Systematic Review and Roadmap Integrating Climate Change , Parasite Genomics , and Public Health Decision Support," *J. Appl. Informatics Comput.*, vol. 10, no. 2, 2026, doi: 10.30871/jaic.v10i2.12347.

[8] M. C. Maher *et al.*, "Predicting the mutational drivers of future SARS-CoV-2 variants of concern," *Sci. Transl. Med.*, vol. 14, no. 633, pp. 0–11, 2022, doi: 10.1126/scitranslmed.abk3445.

[9] X. Liu *et al.*, "Generative prediction of real-world prevalent SARS-CoV-2 mutation with in silico virus evolution," *Brief. Bioinform.*, vol. 26, no. 3, 2025, doi: 10.1093/bib/bbaf276.

[10] S. Sah, B. Surendiran, R. Dhanalakshmi, and S. N. Mohanty, "Mutation prediction and phylogenetic analysis of SARS-CoV2 protein sequences using LSTM based encoder-decoder model," *Arab J. Basic Appl. Sci.*, vol. 30, no. 1, pp. 103–121, 2023, doi: 10.1080/25765299.2023.2188677.

[11] S. King *et al.*, "Forecasting SARS-CoV-2 spike protein evolution from small data by deep learning and regression," *Front. Syst. Biol.*, vol. 4, no. April, pp. 1–13, 2024, doi: 10.3389/fsysb.2024.1284668.

[12] W. J. Choi, J. Park, D. Y. Seong, D. S. Chung, and D. Hong, "A prediction of mutations in infectious viruses using artificial intelligence," *Genomics and Informatics*, vol. 22, no. 1, 2024, doi: 10.1186/s44342-024-00019-y.

[13] E. Ma *et al.*, "A predictive language model for SARS-CoV-2 evolution," *Signal Transduct. Target. Ther.*, vol. 9, no. 1, 2024, doi: 10.1038/s41392-024-02066-x.

[14] Y. Wu, S. Xu, S.-T. Yau, and Y. Wu, "PhyloTransformer: A Discriminative Model for Mutation Prediction Based on a Multi-head Self-attention Mechanism," no. October, pp. 1–12, 2021.

[15] J. Vig, A. Madani, L. R. Varshney, C. Xiong, R. Socher, and N. F. Rajani, "Bertology Meets Biology: Interpreting Attention in Protein Language Models," *ICLR 2021 - 9th Int. Conf. Learn. Represent.*, 2021.

[16] H. Ahmadi, V. Nikoofard, H. Nikoofard, R. Abdolvahab, N. Nikoofard, and M. Esmaeilzadeh, "Prediction of SARS-CoV-2 spike protein mutations using Sequence-to-Sequence and Transformer models," *bioRxiv*, p. 2023.01.23.525130, 2023.

[17] B. A. Sokhansanj, Z. Zhao, and G. L. Rosen, "Interpretable and Predictive Deep Neural Network Modeling of the SARS-CoV-2 Spike Protein Sequence to Predict COVID-19 Disease Severity," *Biology (Basel).*, vol. 11, no. 12, pp. 1–44, 2022, doi: 10.3390/biology11121786.

[18] S. Tasnim, K. H. Talukder, and A. Asfi, "Next Mutation Prediction of Sars-Cov-2 Spike Protein Sequence Using Encoder-Decoder Based Long Short Term Memory (Lstm) Method," *Khulna Univ. Stud.*, vol. 2, no. March 2020, pp. 803–816, 2022, doi: 10.53808/kus.2022.icstem4ir.0142-se.

[19] F. Durazzi, M. P. G. Koopmans, R. A. M. Fouchier, and D. Remondini, "Language models learn to represent antigenic properties of human influenza A(H3) virus," *Sci. Rep.*, vol. 15, no. 1, pp. 1–10, 2025, doi: 10.1038/s41598-025-03275-2.

[20] W. Tang, J. Kim, R. T. Lee, S. Maurer-Stroh, L. Renia, and M. Z. Tay, "SARS-CoV-2: lessons in virus mutation prediction and pandemic preparedness," *Curr. Opin. Immunol.*, vol. 95, p. 102560, 2025, doi: 10.1016/j.coi.2025.102560.

[21] X. Dong *et al.*, "Variation around the dominant viral genome sequence contributes to viral load and outcome in patients with Ebola virus disease," *Genome Biol.*, vol. 21, no. 1, pp. 1–20, 2020, doi: 10.1186/s13059-020-02148-3.

[22] P. Mistry *et al.*, "SARS-CoV-2 Variants, Vaccines, and Host Immunity," *Front. Immunol.*, vol. 12, no. January, pp. 1–21, 2022, doi: 10.3389/fimmu.2021.809244.

[23] C. Cheohen, V. M. S. Gomes, and M. L. da Silva, "CNN-LSTM Hybrid Model for AI-Driven Prediction of COVID-19 Severity from Spike Sequences and Clinical Data," 2025.

[24] I. C. dos Santos, R. di S. de Souza, I. Tolstoy, L. S. Oliveira, and A. Gruber, "Integrating Sequence- and Structure-Based Similarity Metrics for the Demarcation of Multiple Viral Taxonomic Levels," *Viruses*, vol. 17, no. 5, pp. 1–27, 2025, doi: 10.3390/v17050642.

[25] L. Gavotte and R. Frutos, "The stochastic world of emerging viruses," *Res. Rep.*, no. September, pp. 1–8, 2025, doi:

10.1093/pnasnexus/pgac185.
[26] B. D. Trump, S. Galaitsi, J. Cegan, and I. Linkov, "How Will AI Shape the Future of Pandemic Response? Early Clues From Data Analytics," *Risk Anal.*, vol. 45, no. 12, pp. 4544–4556, 2025, doi: 10.1111/risa.70103.
[27] W. Liu, "Bracing the artificial intelligence technology in viral infectious disease control," *Infectious medicine*, vol. 4, no. 2. State Key Laboratory of Pathogen and Biosecurity, Academy of Military Medical Science, Beijing 100071, China., p. 100186, 2025, doi: 10.1016/j.imj.2025.100186.
[28] A. M. Save *et al.*, "Leveraging Machine Learning to Enhance Public Health Outcomes : A Comprehensive Approach to Disease Prediction , Prevention and Management," vol. XXVI, pp. 900–921, 2025.
[29] J. Machemedze and B. Ndlovu, "Transformer-Based Models for Electronic Health Records and Omics in Healthcare : A Systematic Literature Review," *J. Appl. Informatics Comput.*, vol. 10, no. 1, pp. 90–105, 2026, doi: 10.30871/jaic.v10i1.11893.
[30] O. T. Chikumo and B. Ndlovu, "Transformer-based Models for Cardiovascular Disease Predictions from Electronic Health Records : A Systematic Review," *J. Appl. Informatics Comput.*, vol. 10, no. 1, 2026, doi: 10.30871/jaic.v10i1.11899.
[31] S. S. Sibanda and B. Ndlovu, "Explainable Transformer and Machine Learning Models in Predicting Tuberculosis Treatment Outcomes . A Systematic Review," *J. Appl. Informatics Comput.*, vol. 10, no. 1, pp. 150–164, 2026, doi: 10.30871/jaic.v10i1.11846.
[32] M. Khan and Y. Hossni, "A comparative analysis of LSTM models aided with attention and squeeze and excitation blocks for activity recognition," *Sci. Rep.*, vol. 15, no. 1, pp. 1–20, 2025, doi: 10.1038/s41598-025-88378-6.
[33] M. Shahriyari, A. Safari, A. Quteishat, and H. Afsharirad, "An LSTM architecture for real-time multi-domain stability boundary prediction beyond post-fault dependency in power systems," *Sci. Rep.*, vol. 16, no. 1, pp. 1–21, 2026, doi: 10.1038/s41598-026-36571-6.
[34] M. Krichen and A. Mihoub, "Long Short-Term Memory Networks: A Comprehensive Survey," *AI*, vol. 6, no. 9, pp. 1–21, 2025, doi: 10.3390/ai6090215.