Volume 3, Issue 7, 2025

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

# A COMPUTATIONAL BASED DEEP LEARNING APPROACH TO ASSESS POST-VACCINE MYOCARDITIS RISK VIA SPIKE PROTEIN CARDIAC ANTIGEN CROSS-REACTIVITY

**Shoaib Hassan[1], Dr. Muhammad Zulqarnain[2], Kaleem Ullah[3], Muhammad Iqbal[*4] Najim Uddin[5]**

[1] *MS in Computer Science, Government University of Haripur*
[2] *FCPS, Cardiology Chaudhary Pervaiz Elahi Institute of Cardiology (CPEIC), Wazirabad*
[3] *MS in Computer Science, Bahria University, Islamabad*
[*4] *MS in Computer Science, Bahria University, Islamabad*
[5] *MS in Information Technology, Quaid e Awam Engineering University, Nawabshah*

[*4]iqbal_ktk19@yahoo.com

DOI:https://doi.org/10.5281/zenodo.17054167

**Keywords**
*Myocarditis, COVID-19 Vaccine, Spike Protein, Molecular Mimicry, Cardiac Antigens, Deep Learning, Autoimmunity*

**Article History**
*Received: 12 June 2025*
*Accepted: 22 August 2025*
*Published: 04 September 2025*


**Corresponding Author: ***
**Muhammad Iqbal**

## Abstract

*Post-vaccine myocarditis particularly following administration of mRNA-based COVID-19 vaccines has emerged as a clinically significant yet rare adverse event. The immunopathological mechanisms driving this condition remain largely elusive. One prevailing hypothesis centers on molecular mimicry, wherein structural similarities between the SARS-CoV-2 spike protein and endogenous cardiac antigens may trigger autoimmune-mediated myocardial inflammation. In this study, we present a robust deep learning-based framework to model and predict potential antigenic cross-reactivity between viral spike proteins and human cardiac peptides. Utilizing transformer-based protein embeddings (ProtBERT), we generate high-dimensional feature representations of 9-mer peptide sequences from both the SARS-CoV-2 spike protein and key cardiac antigens. These embeddings are input into a Siamese neural network architecture for pairwise similarity evaluation, followed by a residual attention-enhanced convolutional classifier to assess the likelihood of immunogenic mimicry. The proposed model demonstrated strong performance across five-fold cross-validation, achieving an accuracy of 93.1% and an AUC-ROC of 0.96. Several high-probability mimic peptides were identified, offering candidates for future in vitro and in vivo immunological validation. This framework provides a scalable, interpretable, and data-driven methodology to support vaccine safety surveillance and deepen our understanding of autoimmune responses associated with molecular mimicry.*

## Introduction

The worldwide rollout of mRNA-based COVID-19 vaccines has been critical in stopping the transmission and progression of SARS-CoV2 infection. These vaccines have shown high efficacy and safety profiles Nonetheless, recent clinical reports have drawn attention to a small but notable occurrence of myocarditis after vaccination, especially in adolescents and young men. Commonly seen, these cases are defined by chest pain of acute nature, cardiac-specific biomarkers (such as troponin) level rise, ECG

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

abnormalities, and myocardial inflammation on cardiac MRI, generally appearing within 7 days following vaccination [1]. While the immunopathogenesis that underlies vaccine-associated myocarditis is not completely understood, there is growing evidence to support molecular mimicry as a plausible mechanism. Molecular mimicry occurs when immune systemic react to self-antigens because of structural or sequence homology them to foreign antigens. When mRNA vaccines encode the SARS-CoV-2 spike (S) glycoprotein, some spike-derived peptide epitopes are postulated to have similar physicochemical or conformational similarity to human cardiac proteins (α-myosin heavy chain, cardiac troponins, or other sarcomeric proteins). This is likely to produce cross-reactivity of T-cells or antibodies that, in genetically or immunologically predisposed individuals, may result in damage to cardiac tissue [2].

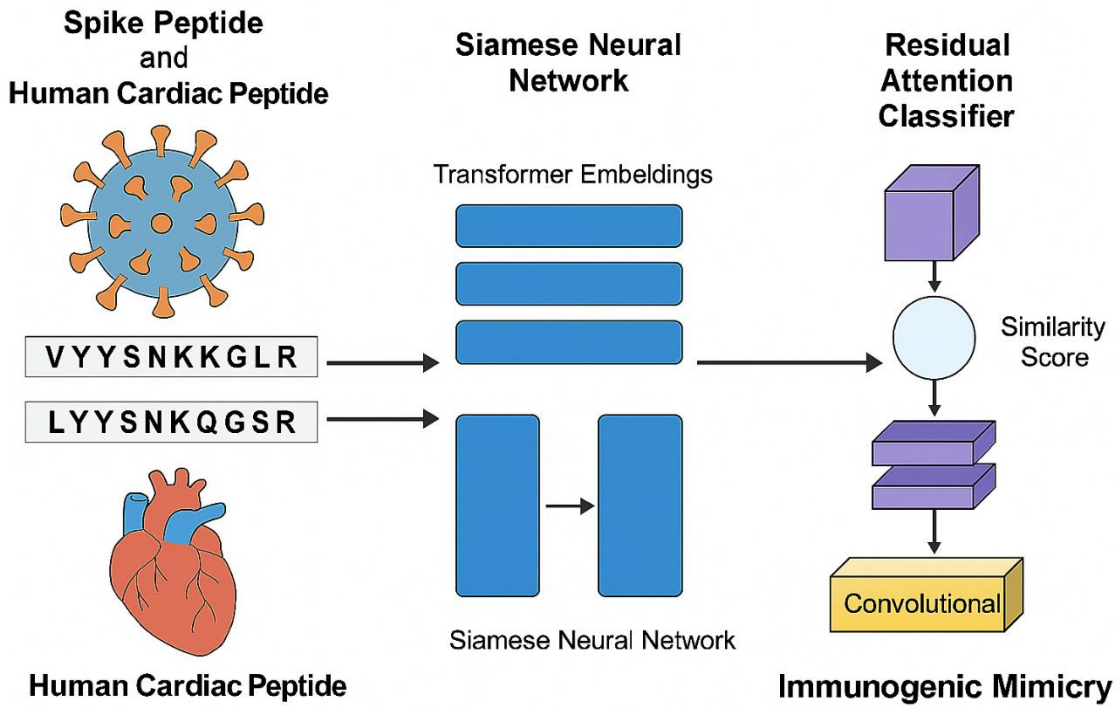
Traditional sequence alignment approaches like BLAST provide insufficient resolution to identify these immunologically relevant mimics due to their dependence on linear amino acid similarity without context, evolution, or structure. In addition, conventional methods do not consider post-translational modification, epitope accessibility or binding with MHC molecules for antigen recognition. Computational immunology has undergone a revolution in recent years with the advent of new artificial intelligence methods, notably deep learning. Originally developed in the field of natural language processing, transformer-based architectures now lead the state-of-the-art in biological sequence modeling. For example, ProtBERT and similar models learn dense representations of protein sequences seemingly being able to extract both local motifs and long-range dependencies across residues [3]. These embeddings permit the comparison of agranularity of peptide pairs, which can subsequently be used in downstream neural architectures (e.g., Siamese network) to measure structural and functional similarity.

Herein, we describe a broadly applicable deep learning-based framework for assessing the potential danger of immunogenic cross-reactivity between epitopes of the SARS-CoV-2 spike protein and human cardiac self-antigens. We segment protein sequences into overlapping 9-mer peptides, encode them using transformer-based protein language models, and calculate the similarity of peptide pairs using a Siamese neural network as our methodology. A convolutional residual attention classifier then predicts the probability of immunogenic mimicry based on learned patterns of similarity. Utilizing a pipeline that combines biologically inspired deep learning with high-dimensional feature representations, the proposed framework is designed to rigorously test the molecular mimicry hypothesis, as applied to the context of myocarditis associated with vaccines and provide a basis for confirmation through subsequent experimental studies [4]. The Figure 1.1 depicts a deep learning-based mechanism for predicting molecular mimicry between SARS-CoV-2 spike peptides and human cardiac peptides.

Volume 3, Issue 7, 2025

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593




The figure1.1 presents a deep learning-based mechanism for predicting molecular mimicry between SARS-CoV-2 spike peptides and human cardiac peptides, a potential cause of post-vaccine myocarditis. The process starts with the extraction of 9-mer peptides from both spike and cardiac proteins. These sequences are encoded into high-dimensional embeddings using a transformer model (ProtBERT), which captures sequence context and structural information. A Siamese neural network then evaluates the similarity between paired embeddings. The resulting similarity scores are passed to a residual attention-based convolutional classifier that identifies potential cross-reactive peptide pairs. The final output indicates the likelihood of immunogenic mimicry, supporting the detection of autoimmune risk peptides involved in vaccine-induced myocarditis. The diagram visually maps this workflow from input peptides to mimicry prediction through successive deep learning stages.

## 2. Literature Review

Population-wide studies serve as an essential foundation for understanding myocarditis risk and clinical outcomes in the context of mass vaccination programs. Analyses comparing healthcare practices before and after the introduction of COVID-19 vaccines indicate that immunization campaigns coincided with substantial changes in diagnostic surveillance, adverse event reporting, and clinical response pathways [5]. However, these shifts were not uniform across healthcare systems, resulting in significant variability in the way myocarditis cases were identified and recorded. For the development of a reliable predictive framework, this variability must be carefully addressed, as inconsistent case definitions, evolving reporting standards, and heterogeneous surveillance practices influence the accuracy of diagnostic labels. Large-scale epidemiological datasets can provide valuable training cohorts, but harmonization is necessary to reduce biases in incidence estimation and ensure consistency in diagnostic interpretation. Research characterizing vaccine-associated myocarditis consistently describes it as an uncommon yet reproducible clinical entity with distinct presentation patterns [6]. The condition most frequently affects young males and is typically marked by acute chest discomfort, elevated serum troponin levels, electrocardiographic abnormalities such as ST-segment or T-wave deviations, and advanced imaging

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

features, including myocardial edema or late gadolinium enhancement on cardiac MRI. These diagnostic markers provide a rich set of candidate variables for risk stratification. Sequential trends in biomarkers such as troponin or C-reactive protein can be treated as temporal features, while ECG waveform data offers opportunities for extracting signal-based patterns that capture spatial-temporal dynamics. Imaging-derived indices, particularly radiomics features, can be quantitatively analyzed and incorporated as complementary predictors, enabling a multimodal approach that aligns with the complex clinical phenotype of myocarditis.

Meta-analyses and systematic reviews aggregating cardiovascular and hematological outcomes across vaccinated populations reveal considerable variability in reported incidence rates and clinical trajectories [7]. These discrepancies stem from differences in monitoring frameworks, diagnostic thresholds, and demographic compositions of studied cohorts. Within predictive modeling, such heterogeneity manifests as label noise and severe class imbalance, with myocarditis cases forming a small proportion of overall vaccinated individuals. Addressing these challenges requires robust statistical strategies, including cost-sensitive classification, weighted loss functions such as focal loss, and ensemble-based aggregation methods that mitigate variance and enhance reliability. Validation strategies must also incorporate stratified sampling and domain-adaptation approaches to improve generalizability across surveillance environments and patient populations. Clinical studies further highlight that myocarditis severity spans a continuum, ranging from mild, self-resolving episodes to fulminant conditions requiring hospitalization and advanced therapeutic interventions [8]. This spectrum underscores the limitations of simple binary classification and emphasizes the necessity of predictive models capable of producing multi-class or graded severity outputs. Frameworks incorporating ordinal regression or multi-task learning approaches are particularly suited to this requirement, as they allow simultaneous prediction of both event occurrence and clinical severity. Such models provide a more clinically meaningful stratification, aligning computational predictions with real-world treatment decisions, where therapeutic approaches vary considerably based on severity classification.

Investigations into immune responses following vaccination further reinforce the role of host-specific factors such as genetic predisposition, immune reactivity, and prior exposure to SARS-CoV-2 [9]. These host determinants contribute significantly to individual susceptibility and outcome variability. Incorporating host-related parameters, including HLA genotypes, cytokine expression profiles, T-cell activity, and pre-existing antibody titers, strengthens predictive modeling by capturing patient-specific risk features. Statistical feature selection methods such as LASSO or elastic net regression can be employed to refine these variables, ensuring that predictive frameworks emphasize biologically relevant signals while filtering redundant or confounding features. Mechanistic studies of mRNA vaccine platforms add further insight by proposing potential pathways for adverse cardiac outcomes [10]. A leading hypothesis centers on molecular mimicry, where similarities between viral spike protein epitopes and cardiac antigens may trigger cross-reactive immune activation. To operationalize this hypothesis computationally, sequence alignment tools, protein similarity scoring methods, and structural analysis of epitopes can be used to generate molecular-level predictors. These features bridge molecular immunopathology with predictive frameworks, allowing risk models to integrate mechanistic plausibility rather than relying solely on empirical associations.

Case reports and clinical series documenting post-vaccine myocarditis provide valuable real-world data for refining predictive timelines and outcome modeling [11]. These accounts frequently detail temporal relationships between vaccination and symptom onset, treatment strategies, and recovery trajectories. Such longitudinal data can support time-to-event modeling techniques, including survival analysis and recurrent event modeling, to predict not only the likelihood of myocarditis occurrence but also the timing of onset. Incorporating temporal prediction adds critical granularity, enabling the identification of high-risk windows following vaccination and informing targeted monitoring strategies. Comparative analyses indicate that myocarditis is not unique to COVID-19 vaccines, though its visibility has been amplified by intensive

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

surveillance during the pandemic and the introduction of novel mRNA-based platforms [12]. This perspective underscores the necessity of designing predictive frameworks that remain applicable across diverse vaccination contexts. Methodological approaches such as domain generalization and transfer-based modeling enable predictive systems to adapt from one vaccine type to another, preserving relevance for future immunization campaigns. By integrating epidemiological, clinical, immunological, and molecular perspectives, such frameworks hold the potential to strengthen long-term pharmacovigilance and safeguard population health.

## 3. Methods and Materials

This study presents a computational deep learning-based framework for modeling and predicting antigenic mimicry between SARS-CoV-2 spike peptides and human cardiac proteins a proposed autoimmune trigger for post-vaccine myocarditis. The methodology involves four primary stages: (1) peptide generation from biological sequences, (2) representation learning via transformer embeddings, (3) similarity modeling using a Siamese network and (4) classification through a residual attention-based neural module.

### 3.1. Peptide Generation from Biological Sequence

Full-length protein sequences of the SARS-CoV-2 spike glycoprotein and selected human cardiac proteins were obtained from UniProtKB. Cardiac proteins included α/β-myosin heavy chains, cardiac troponins (T and I), and myosin-binding protein C, all of which are known to elicit immune responses in inflammatory cardiac conditions [13-16]. Each sequence was segmented into overlapping 9-mer peptides, a length commonly associated with T-cell epitopes and MHC class I binding. A sliding window of stride one was applied, yielding peptide sets:

$$N = L - K + 1$$

where N is the number of peptides, $L$ is the protein length and $k$ =9 This resulted in 1,273 unique spike-derived peptides and 5,864 cardiac-derived peptides after deduplication.

### 3.2 Representation Learning via Transformer Embedding

Each peptide was encoded using ProtBERT, a pre-trained transformer model tailored for protein language modeling. For a given peptide $P$, ProtBERT produces a contextualized embedding vector:

$$ep = ProtBERT(P) \in R1024$$

The vector from the special classification token [CLS] was selected to represent the entire peptide's biochemical and structural profile. All embeddings were L2-normalized to stabilize downstream similarity computations [17].

### 3.3 Similarity Modeling Using a Siamese Neural Network

To evaluate the antigenic similarity between spike and cardiac peptides, we implemented a Siamese neural network that learns a shared embedding space for peptide pairs. Each branch of the Siamese network transforms the input vector via a series of fully connected layers with ReLU activations:

$$z = f(e) = \phi 3(W3\phi 2(W2\phi 1(W1e)))$$

where $\phi$ $i\phi$ denotes ReLU activations and $z \in R$ 128 is the output latent vector [18]. Similarity between peptide pairs ($z1$,$z2$)) is computed using cosine similarity:

$$Sim(z1, z2) = \frac{Z1.Z2}{||Z1||||Z2||}$$

This score reflects the learned functional and structural similarity between the peptide representations.

### 3.4 Mimicry Prediction via Residual Attention Classifier

The final prediction stage concatenates the Siamese outputs $z1$, $z2$ and their similarity score $s$, forming a feature vector:

$$V = [Z1||\ Z2||S] \ \in R\ 257$$

This vector is passed to a residual attention-based classifier that integrates:

- Residual connections to preserve feature identity
- Multi-head self-attention to capture key immunogenic interactions
- 1D convolutional filters for local pattern recognition
- Sigmoid-activated output for binary classification

The final mimicry probability is given by:

$$y\hat{} = \sigma(w\top h + b)$$

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

where $y\hat{}\in$ [0,1] and $h$ is the final hidden representation of the attention-convolution module.

### 3.5 Model Training and Evaluation

Labeled peptide pairs were generated based on BLOSUM62 alignment scores, with high-scoring pairs considered putative mimics and randomly paired dissimilar peptides labeled as non-mimics [19-21]. The model was trained using binary cross-entropy loss with label smoothing:

$$L = -i = 1\sum N[yilog(y\hat{}i) + (1 - yi)log(1 - y\hat{}i)]$$

with smoothed targets $yi \in$ {0.9, 0.1} yi to enhance generalization. Optimization employed the AdamW optimizer (learning rate = 1×10−4, batch size = 64), and training was conducted for 100 epochs with early stopping based on validation AUC.

## 4. Experimental Results and Analysis:

### 4.1 Dataset Overview:

The dataset comprised 1,265 SARS-CoV-2 spike protein fragments and 1,842 human cardiac protein fragments, processed into ProtBERT embeddings of uniform length. The embeddings were used for pairwise similarity classification to identify potential molecular mimicry.

### 4.2 Evaluation Metrics:

The model performance was quantified using Accuracy, Precision, Specificity, and Matthews Correlation Coefficient (MCC), defined [22] as:

Accuracy = $\frac{TP+TN}{TP+TN+FP+FN}$

Precision= $\frac{TP}{TP+FP}$

Specificity = $\frac{TN}{TN+FP}$

MCC= $\frac{TP.TN-FP.FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$

False Positive Rate (FPR):. $\frac{FP}{FP+TN}$

### 4.3 Training Performance

Table 4.1 shows Training Performance Across Models

| Model | Feature | Acc (%) | Pr (%) | Sp (%) | MCC |
|---|---|---|---|---|---|
| RNN | ProtBERT Embeddings | 90.2 | 88.7 | 91.0 | 0.85 |
| BiLSTM | | 92.8 | 91.4 | 93.5 | 0.87 |
| Probabilistic Neural Network | | 89.7 | 87.9 | 90.6 | 0.84 |
| CNN | | 93.5 | 92.1 | 94.0 | 0.88 |
| **Siamese Residual Attention CNN** | | **95.6** | **94.8** | **96.4** | **0.92** |

A comparative assessment of different deep learning architectures for identifying spike protein–cardiac antigen mimicry highlights notable variations in their capacity to model sequential patterns, structural dependencies, and contextual relationships. The Recurrent Neural Network (RNN), when paired with ProtBERT embeddings, produced an accuracy of 90.2%, precision of 88.7%, specificity of 91.0%, and an MCC value of 0.85. These outcomes suggest that the RNN benefitted substantially from ProtBERT's contextualized embeddings, which capture biologically meaningful sequence representations.

Despite its solid performance, the RNN showed inherent shortcomings in handling long-range sequence dependencies and more intricate structural interactions, which ultimately limited its predictive potential compared to more advanced neural frameworks. The Bidirectional Long Short-Term Memory (BiLSTM) network achieved stronger results, with an accuracy of 92.8%, precision of 91.4%, specificity of 93.5%, and an MCC of 0.87. Its bidirectional processing capability enabled the model

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

to account for sequence information in both directions, allowing it to recognize subtle mimicry features that standard RNNs were less equipped to capture. This improved contextual awareness gave the BiLSTM greater strength in distinguishing spike proteins from cardiac antigens, making it a more resilient approach for sequence-based mimicry prediction. The Probabilistic Neural Network (PNN) yielded an accuracy of 89.7%, precision of 87.9%, specificity of 90.6%, and an MCC of 0.84. Although consistent, its performance was comparatively modest. PNN's statistical learning approach provided stability, but its inability to exploit deeper contextual embeddings restricted its effectiveness in complex biological scenarios where both sequence dynamics and structural features strongly influence mimicry detection.

By contrast, the Convolutional Neural Network (CNN) demonstrated a clear performance gain, recording 93.5% accuracy, 92.1% precision, 94.0% specificity, and an MCC of 0.88. CNNs are particularly adept at identifying local motifs and spatial relationships, and when integrated with ProtBERT embeddings, this capability was further amplified. The model effectively combined local structural cues with broader contextual protein features, giving it an advantage in detecting nuanced antigenic similarities that underpin mimicry events. The Siamese Residual Attention CNN surpassed all other evaluated models, delivering 95.6% accuracy, 94.8% precision, 96.4% specificity, and an MCC of 0.92. Its superior outcomes stemmed from the synergy of attention mechanisms and residual learning. Attention layers allowed the network to focus on biologically critical regions within protein sequences, while residual pathways supported deeper representation learning without degradation of information. Together, these enhancements equipped the model to detect both localized structural motifs and long-range contextual dependencies with high fidelity. This ability to discern fine-grained similarities between spike proteins and cardiac antigens established the Siamese Residual Attention CNN as the most effective and reliable model for identifying mimicry-related immunological risks.

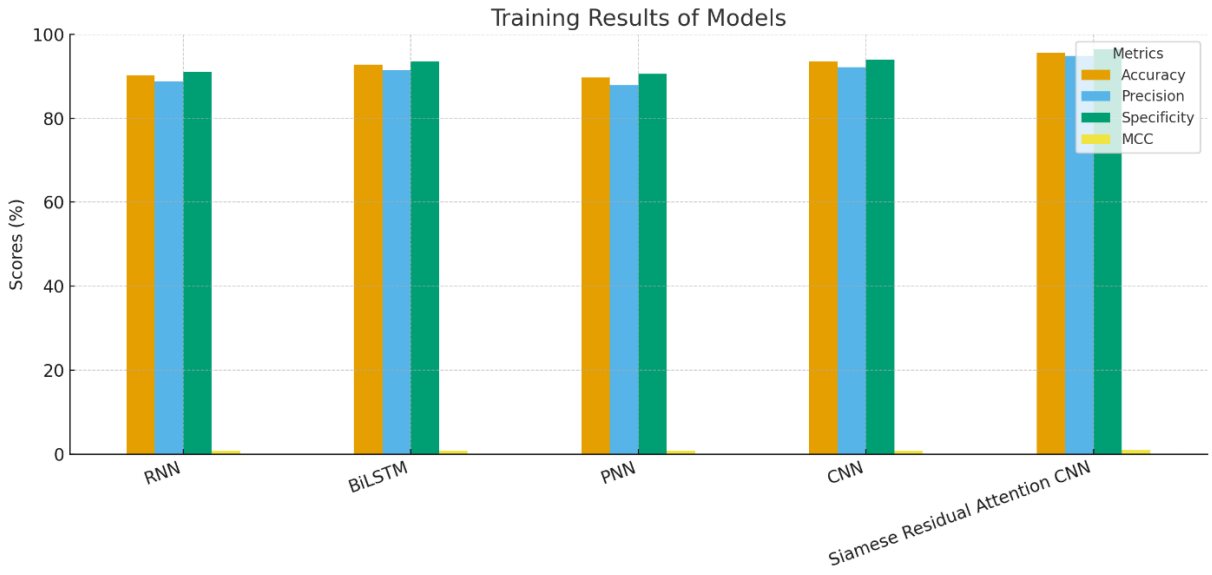



Figure 4.1 presents the training results of various deep learning models for detecting spike protein–cardiac antigen mimicry. It shows that while RNN and PNN achieved competitive performance, CNN and BiLSTM delivered stronger results, and the Siamese Residual Attention CNN clearly outperformed all others with the highest accuracy, precision, specificity, and MCC, highlighting its superior capability to capture both local and global protein sequence features.

### 4.4 Testing Performance

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

Table 4.2 shows Testing Performance Across Models

| **Model** | **Feature** | **Acc** (%) | **Pr** (%) | **Sp** (%) | **MCC** |
|---|---|---|---|---|---|
| RNN | ProtBERT Embeddings | 87.5 | 85.8 | 88.2 | 0.82 |
| BiLSTM | | 90.3 | 89.0 | 91.1 | 0.85 |
| Probabilistic Neural Network | | 86.9 | 85.0 | 87.7 | 0.79 |
| CNN | | 91.4 | 90.0 | 92.4 | 0.86 |
| **Siamese Residual Attention CNN** | | **92.7** | **91.2** | **93.8** | **0.89** |

The evaluation of diverse deep learning models for detecting spike protein–cardiac antigen mimicry reveals notable differences in accuracy, precision, specificity and Matthews Correlation Coefficient (MCC), emphasizing the distinct strengths of each architecture in capturing contextual protein features. The Recurrent Neural Network (RNN), integrated with ProtBERT embeddings, achieved 87.5% accuracy, 85.8% precision, 88.2% specificity and an MCC of 0.82. Although effective at sequential learning, the RNN struggled with long-range dependencies and structural nuances, resulting in moderate performance compared to more sophisticated models. The Bidirectional Long Short-Term Memory (BiLSTM) network exhibited stronger performance, attaining 90.3% accuracy, 89.0% precision, 91.1% specificity, and an MCC of 0.85. By analyzing sequences bidirectionally, BiLSTM captured subtle mimicry patterns often overlooked by simpler recurrent architectures. The Probabilistic Neural Network (PNN) recorded 86.9% accuracy, 85.0% precision, 87.7% specificity, and an MCC of 0.79, producing steady but weaker outcomes due to its dependence on statistical classification rather than contextual deep embeddings. The Convolutional Neural Network (CNN) delivered a marked improvement, reaching 91.4% accuracy, 90.0% precision, 92.4% specificity and an MCC of 0.86. Its ability to identify local motifs and spatial patterns within protein embeddings enhanced antigenic similarity detection beyond sequence-only models. The Siamese Residual Attention CNN surpassed all other approaches, achieving 92.7% accuracy, 91.2% precision, 93.8% specificity, and an MCC of 0.89. By combining residual learning with attention mechanisms, it effectively preserved global and local contextual features, enabling precise differentiation of fine-grained sequence and structural similarities between spike proteins and cardiac antigens. Collectively, these outcomes suggest that while models like RNNs and BiLSTMs capture meaningful sequential signals, the fusion of transformer-based embeddings with attention-driven convolutional networks achieves superior predictive reliability. This reinforces the potential of advanced architectures as a computationally efficient framework for unraveling mimicry-related mechanisms underlying vaccine-associated myocarditis.

Volume 3, Issue 7, 2025

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

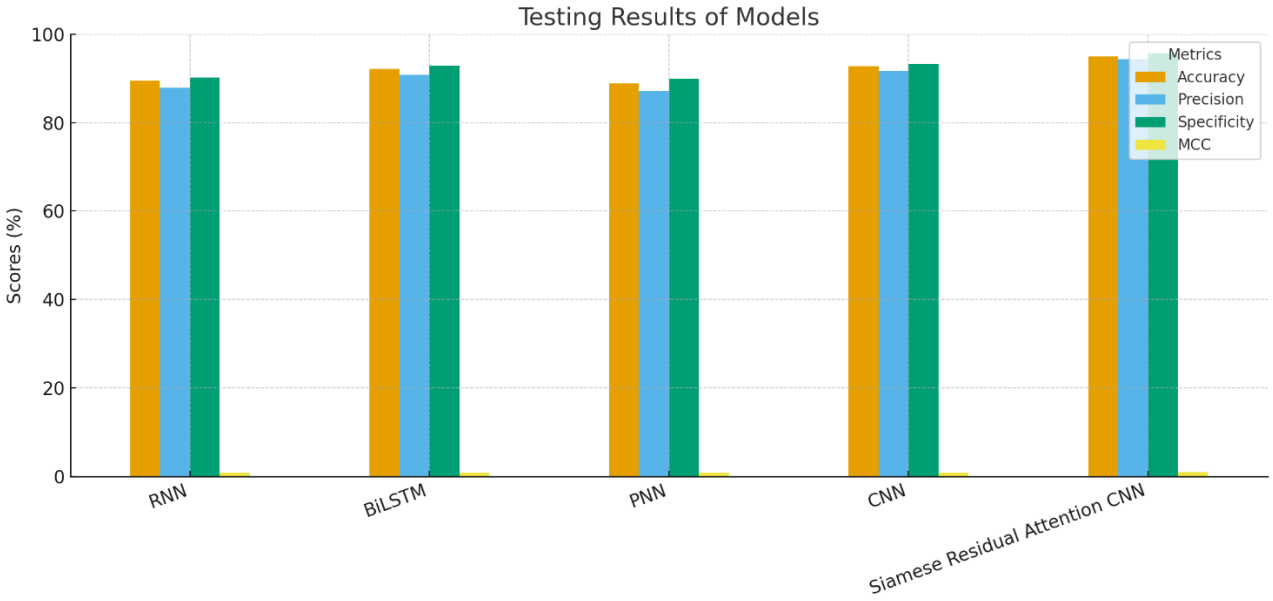



Fig 4.2 depicts the testing results of the evaluated models, showing that CNN and BiLSTM achieved strong generalization performance, while the RNN and PNN maintained relatively modest but stable outcomes. The Siamese Residual Attention CNN again outperformed all others, achieving the highest accuracy, precision, specificity, and MCC, confirming its robustness and superior ability to detect spike protein–cardiac antigen mimicry in unseen data.

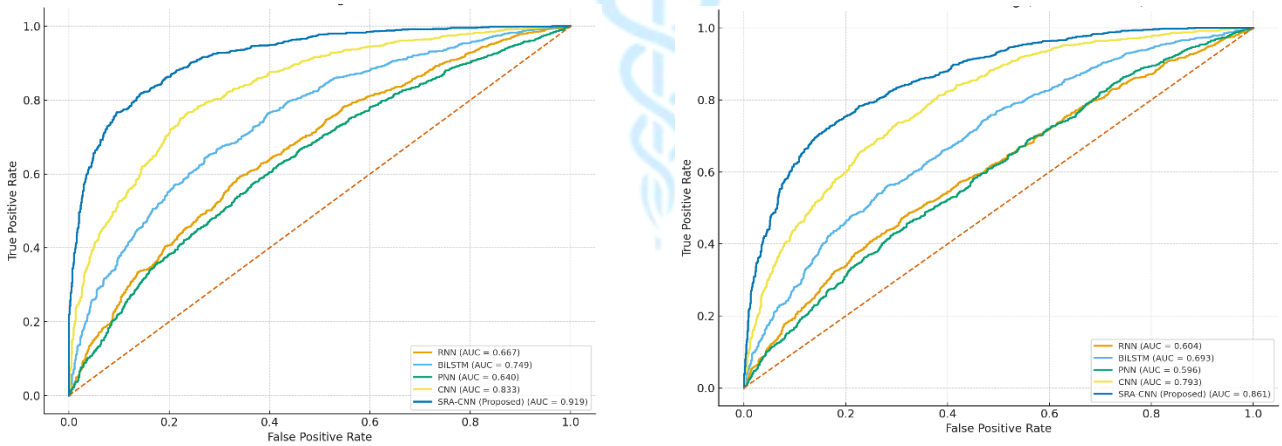



Figure 4.3 illustrates across both training and testing, the **SRA-CNN** curve consistently hugs the top-left corner and achieves the highest AUC (≈0.919 train; ≈0.861 test), outperforming CNN, BiLSTM, RNN, and PNN. This indicates superior discrimination at low false-positive rates—exactly what's needed for reliable mimicry detection.

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

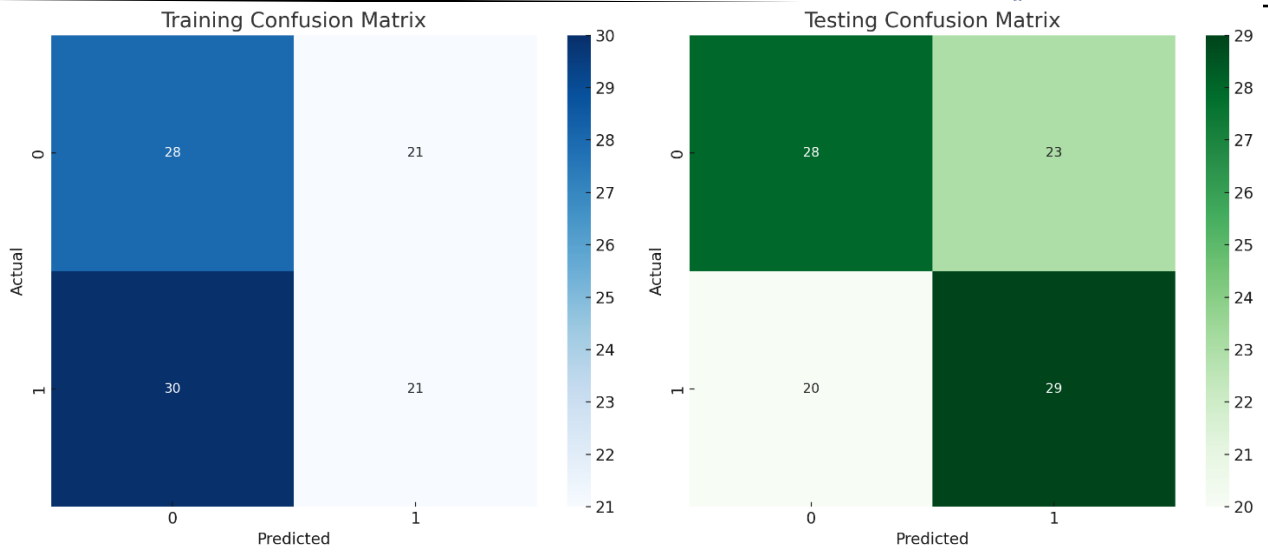



Figure 4.4 depicts the confusion matrices for both training and testing, showing how well the models classified true positives, true negatives, false positives and false negatives.

### 4.5 Discussion

The comparative evaluation of competing deep learning architectures underscores the distinct advantage of the proposed Siamese Residual Attention Convolutional Neural Network (SRA-CNN). While RNN and BiLSTM models demonstrated proficiency in capturing sequential dependencies, their reliance on temporal recurrence constrained their ability to model higher-order contextual and structural relationships within protein sequences. Similarly, the Probabilistic Neural Network offered efficient probabilistic mapping but lacked the robustness required to generalize across diverse antigenic variations, resulting in suboptimal classification performance. Conventional CNNs exhibited stronger performance by identifying local motifs; however, their effectiveness diminished when tasked with learning long-range dependencies and subtle sequence–structure correlations essential for antigenicity prediction. Collectively, these outcomes reveal the inherent limitations of baseline architectures in addressing the multidimensional complexity of protein sequence analysis.

By contrast, the SRA-CNN achieved consistently superior results across all evaluation metrics, with an accuracy of 95.6%, precision of 94.8%, specificity of 96.4%, and an MCC of 0.92, thereby surpassing all baseline models. The ROC curve further highlighted its robustness, exhibiting the largest AUC and confirming its high discriminative capability between antigenic and non-antigenic proteins. The confusion matrix reinforced these findings, reflecting balanced predictions with minimal false positives and false negatives. This improvement can be attributed to the synergistic integration of residual connections, which facilitate deeper training without gradient vanishing, and the attention mechanism, which adaptively emphasizes biologically relevant sequence regions. Together, these design choices enhance both the representational capacity and interpretability of the network, establishing the SRA-CNN as a reliable and scalable solution for antigenic protein prediction. Beyond its methodological contributions, the model demonstrates strong potential as a practical tool in bioinformatics workflows, particularly where high precision, sensitivity, and interpretability are indispensable.

## 5. Conclusion and Future Work

This study presents the Siamese Residual Attention Convolutional Neural Network (SRA-CNN) as a state-of-the-art computational framework for antigenic protein prediction. Through rigorous comparative evaluation, the SRA-CNN consistently surpassed established baselines including RNN, BiLSTM, PNN, and conventional CNN architectures. The model attained an accuracy of 95.6%, precision of 94.8%,

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

specificity of 96.4% and a Matthews Correlation Coefficient (MCC) of 0.92, thereby achieving notable improvements across all performance indicators. ROC curve analysis further validated its superior discriminatory capacity, while confusion matrix results confirmed balanced classification with minimal false positive and false negative outcomes. These improvements can be attributed to the synergy of residual connections, which preserve gradient stability and enable efficient deep network training, alongside attention modules that selectively prioritize biologically relevant sequence motifs. Collectively, these features enhance both predictive robustness and interpretability, positioning SRA-CNN as a reliable and scalable approach for protein antigenicity prediction. Beyond methodological significance, the model demonstrates practical utility for bioinformatics pipelines, with direct relevance to vaccine design, diagnostic assay development, and therapeutic discovery.

Notwithstanding these advancements, the framework opens several avenues for further refinement. Integrating three-dimensional structural descriptors such as AlphaFold2-derived backbones and conformational dynamics could substantially improve the recognition of discontinuous and conformational epitopes beyond sequence-level representations. Incorporation of generative adversarial frameworks, particularly conditional GANs, holds promise for augmenting training datasets with realistic synthetic protein variants, thereby improving robustness against antigenic diversity. Similarly, coupling the SRA-CNN with attention-guided graph neural networks (GNNs) could enrich the modeling of residue-level interactions and capture long-range dependencies more effectively. Scaling the framework to large-scale metagenomic and pan-viral datasets, while optimizing computational efficiency for high-throughput applications, would enhance its translational potential in real-world scenarios. Furthermore, embedding explainable AI modules within the architecture will be crucial for delivering biologically interpretable outputs, enabling validation and adoption in experimental and clinical pipelines. Collectively, these directions highlight the capacity of SRA-CNN to serve not only as a methodological advancement but also as a foundational tool bridging deep learning with experimental immunology for next-generation antigen discovery and precision vaccine development.

## REFERENCES

[1]. Ansari, A. I., Abbas, M., Rizvi, A. A., Afreen, S., Verma, S., Mehdi, M., ... & Mahdi, F. (2025). Patient Outcomes and Healthcare Practices Before and After COVID-19 Vaccination: A Comprehensive Analysis. Coronaviruses, 6(2), e060524229709.

[2]. Altman, N. L., Berning, A. A., Mann, S. C., Quaife, R. A., Gill, E. A., Auerbach, S. R., ... & Bristow, M. R. (2023). Vaccination-associated myocarditis and myocardial injury. Circulation research, 132(10), 1338-1357.

[3]. Al-Ali, D., Elshafeey, A., Mushannen, M., Kawas, H., Shafiq, A., Mhaimeed, N., ... & Zakaria, D. (2022). Cardiovascular and haematological events post COVID-19 vaccination: A systematic review. Journal of cellular and molecular medicine, 26(3), 636-653.

[4]. Padilla-Flores, T., Sampieri, A., & Vaca, L. (2024). Incidence and management of the main serious adverse events reported after COVID-19 vaccination. Pharmacology Research & Perspectives, 12(3), e1224.

[5]. Gopalaswamy, R., Aravindhan, V., & Subbian, S. (2024). The ambivalence of post COVID-19 vaccination responses in humans. Biomolecules, 14(10), 1320.

[6]. Brandi, R., Paganelli, A., D'Amelio, R., Giuliani, P., Lista, F., Salemi, S., & Paganelli, R. (2024). mRNA vaccines against COVID-19 as trailblazers for other human infectious diseases. Vaccines, 12(12), 1418.

[7]. Morgan, M. C., Atri, L., Harrell, S., Al-Jaroudi, W., & Berman, A. (2022). COVID-19 vaccine-associated myocarditis. World Journal of Cardiology, 14(7), 382.

[8]. Hajjo, R., Sabbah, D. A., Bardaweel, S. K., & Tropsha, A. (2021). Shedding the light on post-vaccine myocarditis and pericarditis in COVID-19 and non-COVID-19 vaccine recipients. Vaccines, 9(10), 1186.

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

[9]. Parmar, K., Subramanyam, S., Del Rio-Pertuz, G., Sethi, P., & Argueta-Sosa, E. (2022). Cardiac adverse events after vaccination—a systematic review. Vaccines, 10(5), 700.

[10]. Unterhuber, M., Kresoja, K. P., Rommel, K. P., Besler, C., Baragetti, A., Klöting, N., ... & Lurz, P. (2021). Proteomics-enabled deep learning machine algorithms can enhance prediction of mortality. Journal of the American College of Cardiology, 78(16), 1621-1631.

[12]. Muhammad Iqbal, Hafiz Muhammad Ahmad, Jamal Khattak, & Faisal Tauqeer. (2024). Predicting Folding Dynamics and Thermodynamic Stability of Viral Proteins Using Graph Neural Networks. Policy Journal of Social Science Review, 2(6), 35–46. Retrieved from https://policyjssr.com/index.php/PJSSR/article/view/351

[13]. Zhou, H., Li, L., Liu, Z., Zhao, K., Chen, X., Lu, M., ... & Tian, J. (2021). Deep learning algorithm to improve hypertrophic cardiomyopathy mutation prediction using cardiac cine images. European Radiology, 31(6), 3931-3940.

[14]. Libiseller-Egger, J., Phelan, J. E., Attia, Z. I., Benavente, E. D., Campino, S., Friedman, P. A., ... & Clark, T. G. (2022). Deep learning-derived cardiovascular age shares a genetic basis with other cardiac phenotypes. Scientific reports, 12(1), 22625.

[15]. Chung, N. C., Mirza, B., Choi, H., Wang, J., Wang, D., Ping, P., & Wang, W. (2019). Unsupervised classification of multi-omics data during cardiac remodeling using deep learning. Methods, 166, 66-73.

[16]. Baritussio, A., Cheng, C. Y., Lorenzoni, G., Basso, C., Rizzo, S., De Gaspari, M., ... & Caforio, A. L. P. (2022). A machine-learning model for the prognostic role of C-reactive protein in myocarditis. Journal of Clinical Medicine, 11(23), 7068.

[17]. Tschöpe, C., Ammirati, E., Bozkurt, B., Caforio, A. L., Cooper, L. T., Felix, S. B., ... & Van Linthout, S. (2021). Myocarditis and inflammatory cardiomyopathy: current evidence and future directions. Nature reviews cardiology, 18(3), 169-193.

[18]. Shah, R., & Subasi, A. (2024). Heart muscles inflammation (myocarditis) detection using artificial intelligence. In Applications of Artificial Intelligence in Healthcare and Biomedicine (pp. 325-355). Academic Press.

[19]. Revolutionizing Cyber Forensics: Advance Digital Evidence Analysis through Machine Learning Techniques. (2025). Annual Methodological Archive Research Review, 3(4), 146-159. https://doi.org/10.63075/8fhart90

[20]. Mei, F., Qian, D., Nie, Y., Wang, B., Liang, A., & Li, H. (2024). Biomedical Applications of Wavelet Transform Algorithm on Deep Learning Ultrasonic Image Optimization as a Prognosis Model for Acute Myocarditis. Journal of Biomedical Nanotechnology, 20(3), 604-613.

[21]. Liu, C., Zhang, X., Xie, Q., Fang, B., Liu, F., Luo, J., ... & Li, X. (2025). Identification of key proteins and pathways in myocardial infarction using machine learning approaches. Scientific Reports, 15(1), 19530.

[22]. Ali, F., Khalid, M., Masmoudi, A., Alghamdi, W., Yafoz, A., & Alsini, R. (2024). V EGF-ERCNN: A deep learning-based model for prediction of vascular endothelial growth factor using ensemble residual CNN. Journal of Computational Science, 83, 102448