International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 11 | Nov - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

Genome Analysis of Nipah Virus Using Recurrent Neural Networks (RNN)

K. S. Sree Vidhya
Department of CSE
Arunachala College of Engineering for Women
sreevidhya504@gmail.com

C. Pushpalatha
Associate Professor, Department of CSE
Arunachala College of Engineering for Women
pushrajacse@gmail.com

Abstract—The Nipah virus, a highly pathogenic zoonotic virus, poses significant threats to human health, with high mortality rates and limited treatment options. Understanding its genomic structure is crucial for developing effective diagnostic and therapeutic strategies. This project leverages deep learning techniques, specifically Recurrent Neural Networks (RNNs), to analyze the Nipah virus genome. The primary objective is to identify key genetic features and variations that influence its pathogenicity. RNA sequence data of the virus is processed using Word2Vec for feature representation, transforming nucleotide sequences into vector embeddings. The trained RNN model is then employed to predict potential mutations in the viral genome and assess their implications for viral behavior. The proposed system aims to enhance our understanding of the virus's genomic makeup, enabling more accurate predictions of pathogenic variations and facilitating the development of targeted interventions. Through this approach, we seek to contribute to the advancement of genomic analysis in viral pathogenesis and disease control.

Keywords— Recurrent Neural Networks(RNNs), Nipha virus, Word2Vec, RNA

I. INTRODUCTION

Nipah virus (NiV) is a zoonotic pathogen that has caused severe outbreaks in South and Southeast Asia, representing a significant threat to public health due to its high fatality rate and potential for human-to-human transmission. The disease is typically diagnosed through serological testing or PCR-based methods, but no specific therapeutic treatments or vaccines are currently available. NiV is an RNA virus, and while Ribavirin has been explored as a potential treatment, no definitive solutions have emerged. NiV typically targets immune-compromised individuals, with dendritic cells being one of its primary targets. These cells are an integral part of the immune system and are found among various blood cell types [5].

Traditional approaches to predicting and managing NiV outbreaks have primarily relied on epidemiological models such as the SIR (Susceptible-Infected-Recovered) and SEIR (Susceptible-Exposed-Infected-Recovered) models, as well as phylogenetic analysis. While these models have been useful, they often fail to capture the complexities and uncertainties inherent in disease spread. Conventional models assume a homogeneous mixing of populations and fixed parameters, which do not fully reflect the dynamic and variable nature of real-world disease transmission [1][2]. Additionally, these models can be computationally intensive or require large amounts of labeled data, which is difficult to obtain in the context of emerging infectious diseases [3][4].

The importance of accurate and reliable outbreak prediction cannot be overstated, as it is crucial for timely public health interventions and resource allocation. Existing methods, such as Bayesian inference and machine learning techniques, have shown promise but also come with notable limitations, including difficulties in handling uncertainties effectively [3]. Recent research has explored the use of machine learning techniques to improve prediction and diagnostic accuracy for intractable diseases like NiV. Machine learning models have demonstrated enhanced predictive capabilities, offering promising results for early detection and outbreak management. This approach is being increasingly emphasized by organizations like the World Health Organization (WHO) and

the National Centre for Disease Control (NCDC), as it holds potential for optimizing prediction models and controlling the disease [6].

Gene expression data processing in molecular biology involves experimental and computational methods to analyze gene activity. Techniques such as qRT-PCR, which quantifies gene expression levels, are commonly used to validate high-throughput results. Microarrays, which analyze thousands of genes simultaneously, have been the standard but are gradually being surpassed by RNA sequencing (RNA-Seq) due to its higher resolution and ability to detect novel transcripts. A newer technique, Drop-seq, combines RNA-Seq with microfluidics to study individual-cell gene expression. After experimental data collection, bioinformatics tools such as DESeq2 and edgeR are often employed for normalization and differential expression analysis, enabling a deeper understanding of gene activity and viral behavior.

High-level Biosafety Level-4 (BSL-4) laboratories are required to safely test and handle the Nipah virus due to its potential for severe outbreaks. These labs are equipped with the necessary security and containment measures to study such dangerous pathogens effectively. Domestic animals, particularly bats, play a key role in the transmission of NiV, facilitating the spread of the virus to humans. In India, over 100 species of bats have been identified as hosts for NiV, with 31 species being affected by the virus [6]. Despite the availability of specialized labs and advanced methodologies, there remains a need for more accurate and optimized models for early detection and prediction to control future outbreaks.

The application of deep learning techniques to gene expression data for classification tasks has gained significant attention due to the complexity and high dimensionality of the data, which require advanced computational models. Deep learning algorithms, particularly convolutional neural networks (CNNs) and recurrent neural networks (RNNs), offer a robust framework to uncover intricate patterns in gene expression, making them highly effective for classification tasks. The multi-layered architectures of these models enable them to capture non-linear relationships in gene expression data, which is essential for accurate classification and understanding of gene activity [7].

A key advantage of deep learning models is their ability to autonomously learn feature representations from raw data, eliminating the need for manual feature extraction. This not only reduces the introduction of potential biases but also enhances the model's ability to generalize to new data. When properly trained, deep learning models can minimize classification errors and optimize performance, providing reliable predictive capabilities. Furthermore, models such as autoencoders and variational autoencoders have shown efficacy in performing dimensionality reduction on gene expression data, helping to mitigate issues related to the curse of dimensionality and improving model efficiency [8].

Recurrent neural networks (RNNs), in particular, offer specific advantages over CNNs when processing gene expression data, especially for sequence prediction and understanding temporal dynamics [9]. Unlike CNNs, which typically treat input features as independent, RNNs are designed to handle sequential data, making them ideal for analyzing gene expression time series where temporal dependencies are crucial. RNNs can maintain an internal memory of past inputs, allowing them to capture dynamic temporal behaviors in gene expression profiles. This ability is particularly useful in scenarios where gene expression changes over time, and the model needs to understand the progression or regulation of genes across different time points. Additionally, RNNs can handle sequences of varying lengths without requiring a fixed input size, providing greater flexibility in analyzing data from different experimental setups.

Moreover, RNNs can process bi-directional input, which is advantageous when investigating gene expression profiles with bi-directional influences. The sequential nature of RNNs also allows for more intuitive insights into gene expression pathways and regulatory cascades, potentially offering biologically relevant interpretations of gene activity compared to the hierarchical feature learning in CNNs. However, RNNs are not without challenges. They are prone to issues such as vanishing and exploding gradient problems, which can hinder model training and performance. These challenges must be carefully addressed during the development and application of RNNs in bioinformatics tasks.

The key contribution of this work lies in the innovative use of deep learning techniques, specifically Recurrent Neural Networks (RNNs), to analyze the genomic structure of the Nipah virus (NiV).

- This work leverages Recurrent Neural Networks (RNNs), a deep learning technique, to analyze the genomic structure of the Nipah virus (NiV), offering a novel approach for studying viral genomes.
- RNA sequence data of the virus is processed using Word2Vec to convert nucleotide sequences into vector embeddings, enabling the RNN model to capture complex patterns and genetic features in the virus's genome.
- By identifying key genetic features and variations, this approach can facilitate the development of more targeted and effective diagnostic and therapeutic interventions for NiV, potentially improving disease management.

II. LITERATURE SURVEY

Nipah virus (NiV) is a severe and unpredictable respiratory infection that can lead to coma, encephalitis, or even death. Despite the ongoing research, no approved vaccines or effective treatments are currently available. A systematic review by Aditi and M. Shariff [10] explores the biological aspects of NiV, including its immunopathogenesis and diagnostic approaches for clinical settings. The authors conducted a literature survey from sources like Cochrane Library, Google Scholar, and PubMed, gathering insights into antibody detection techniques and case definitions based on NCDC guidelines. Rodolphe Pelissier et al. [11] discuss the immunopathogenesis of NiV infection, highlighting the protein expression systems involved and their relationship to human immune responses during the disease.

The natural reservoir for Nipah virus (NiV) involves hosts that spread the virus primarily through the respiratory route or throat swabs. Jorge D. Mello-Roman et al. [13] conducted a case study on dengue in Paraguay, demonstrating the effectiveness of combining Artificial Neural Networks (ANN) and Support Vector Machines (SVM) for early dengue diagnosis. SVM, in particular, achieved over 90% accuracy, alongside high specificity and sensitivity, outperforming other models. Similarly, Gaurav Sharma, Seema Bawa, and colleagues [5] applied hybrid machine learning models, including Random Forest and K-Means, for predicting T-cell Lymphotropic Virus, evaluating the performance through K-fold cross-validation and AUROC methods. Additionally, machine learning models such as XGBoost, Random Forest, Decision Tree, and Logistic Regression were used to predict HBsAg seroclearance with a focus on specificity [14].

In response to NiV, Akanksha Rajput, Archit Kumar, and Manoj Kumar [15] developed an "anti-Nipah" web server that integrates data from PubMed and patents into a QSAR model using machine learning techniques. This tool provides valuable resources related to NiV and its inhibitors, though it lacks a detailed methodology for its computational work.

Md. Zakiul Hassan[16] analyzes Nipah virus (NiV) gene expression using NGS RNA-Seq data to identify differentially expressed genes (DEGs). NiV, an enveloped ssRNA paramyxovirus, has a high case fatality rate (>70%). Using statistical tools like limma and bioinformatics platforms such as Cytoscape, Ensembl, and STRING, the study identifies 2707 DEGs (p-value <0.05) from a total of 54359 NiV genes. Key up-regulated genes include EPST1, MX1, IFIT3, RSAD2, and OAS1, while down-regulated genes include SLFN13 and SPAC977.17. The gene interaction analysis reveals no significant association between NiV and viruses like Ebola or Tularemia, highlighting the unique genetic profile of NiV. These findings provide potential biomarkers and candidates for future vaccine or drug development to combat Nipah virus infections.

Sergii Babichev[17], various recurrent neural network (RNN) architectures were tested for gene expression data classification. The performance of models was evaluated using classification accuracy, F1-score, and loss function values. The results showed that a single-layer GRU network with 75 neurons outperformed other models, achieving a 97.2% classification accuracy, slightly better than CNN and LSTM models (97.1%). The GRU model correctly identified 954 out of 981 objects, making it the most effective model for classifying gene expression data in this study.

Rui Yin[18], the challenge of predicting influenza virus mutations is addressed through a time-series mutation prediction model called Tempel. Influenza viruses evolve rapidly, complicating antiviral treatments, so predicting mutations for upcoming flu seasons is critical. Tempel leverages recurrent neural networks (RNNs) with attention mechanisms to model the temporality and dimensionality of influenza A virus glycoprotein hemagglutinin sequences. The attention mechanism improves prediction by focusing on key residue parts in the sequence. Experimental results from three influenza datasets show that Tempel significantly outperforms existing approaches, offering valuable insights into viral mutation dynamics and evolution.

Liping Ma[19], a rapid RT-LAMP assay was developed for detecting Nipah virus (NiV), targeting the nucleocapsid protein (N) gene. The assay, which works at 65°C, was found to be 10 times more sensitive than conventional RT-PCR, with a detection limit of 100 pg of NiV pseudovirus RNA. It showed high specificity with no cross-reactivity to related viruses, providing results in 45 minutes using simple equipment. Clinical testing confirmed the assay's stability and effectiveness for detecting all known NiV strains, offering a promising tool for rapid field detection.

Syed Kannan[20], a prognostic model for early Nipah virus (NiV) diagnosis was developed using Machine Learning (ML), combining clinical factors like symptoms and blood test results. The model used a Restricted Boltzmann Machine (RBM) for feature selection and a stacking ensemble meta classifier (SEMC) for prediction. Trained on data from the 2018-2019 NiV outbreak in Kozhikode, Kerala, the SEMC model achieved 88.3% accuracy and high precision (92.5%) and recall (89.2%). Key indicators like fever, headache, and cough were identified as critical for diagnosis. The model could aid early detection of NiV, though more data is needed for maximum accuracy.

Rodolphe Pelissier[21], Nipah virus (NiV), a deadly zoonotic virus, is examined for its impact on the immune system. While causing severe illness in humans, NiV remains

asymptomatic in its natural hosts, fruit bats, which facilitate outbreaks. The virus disrupts the innate immune response, particularly interferon signaling, allowing it to spread. Human-to-human transmission is common, especially in Bangladesh and India. The review summarizes recent research on NiV's immune modulation and stresses the need for new prophylactic and therapeutic strategies to control this emerging threat.

III. Proposed Methodology

Here, we propose a deep learning-based approach, NiV-RNN, to analyze the genome of the Nipah virus and predict potential mutations that influence its pathogenicity. The NiV-RNN model takes RNA sequence data of the virus as input and converts it into a feature matrix using the Word2Vec method. This involves dividing the RNA sequence into overlapping k-mers (short subsequences of fixed length) and converting each k-mer into a vector representation using the Word2Vec algorithm. The resulting feature matrix is then input into a Recurrent Neural Network (RNN), which is designed to capture the temporal dependencies and complex patterns in the viral genome.

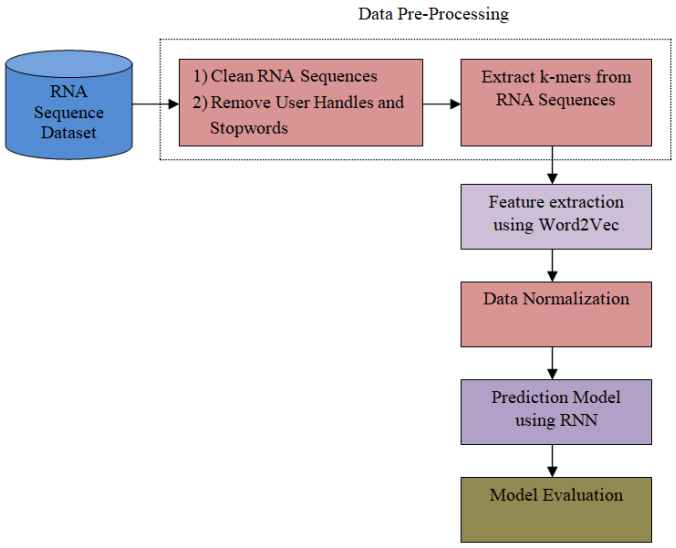

Figure 1. Architecture of the proposed NiV-RNN model

The RNN is specifically trained to identify key genetic features and variations that could affect the virus's pathogenicity and predict how potential mutations may impact viral behavior. The output from the RNN is passed through a fully connected layer and a sigmoid layer, which classifies the sequence based on whether it contains mutations that may affect the virus's ability to cause disease. The proposed system aims to enhance the understanding of the Nipah virus's genomic structure, enabling more accurate predictions of pathogenic variations and contributing to the development of targeted diagnostic and therapeutic strategies.

Through this approach, NiV-RNN seeks to advance genomic analysis in viral pathogenesis, facilitating the identification of mutation hotspots and improving disease control strategies for this high-risk pathogen. The architecture of the proposed NiV-RNN model is illustrated in Figure1.

1) Data Pre-processing

In this step, the RNA sequences undergo a series of text preprocessing techniques to prepare the data for further analysis and model training. The primary goal is to eliminate unnecessary or irrelevant information that could introduce noise into the model and negatively impact its performance.

The first cleaning task involves removing user handles. These handles, which are often present in the raw data, have no meaningful relevance for RNA sequence classification, especially when dealing with biological sequences. Removing these user handles ensures that the data is focused solely on the sequence itself, which is critical for accurate model training and classification.

The next preprocessing step is removing stopwords, which are common words that appear frequently in text but do not contribute significant meaning to the analysis. In traditional natural language processing (NLP), stopwords include words like "the," "a," "and," "is," etc., which are typically excluded from analysis because they do not help in distinguishing between different classes. Although RNA sequences are not text in the conventional sense, the preprocessing of sequences into k-mers (subsequences) can result in the presence of "stop" k-mers — short subsequences that frequently appear but do not provide meaningful information for classification. By removing these, we reduce the noise in the data, making it cleaner and more focused on the features that are important for the classification task.

After cleaning the RNA sequences by removing user handles and stopwords, the next critical step is to extract k-mers from the cleaned sequences. A k-mer is a subsequence of length k (in this case, $k = 6$) extracted from the original RNA sequence. This process involves breaking the sequences into smaller, more manageable units, which are likely to help the model capture local patterns within the data that are essential for distinguishing between classes, such as "Infected" versus "Normal."

The extraction of k-mers is done using a sliding window approach, where a fixed-length window of size 6 moves along the RNA sequence to generate overlapping subsequences. This approach ensures that every possible subsequence of length 6 is captured from the original sequence. For example, for a sequence AGCTAGCTAG, the sliding window will generate the k-mers AGCTAG, GCTAGC, CTAGCT, and so on. By generating k-mers in this way, the RNA sequence is represented in terms of shorter, more manageable components that retain critical local sequence patterns.

Overall, these preprocessing steps—removing irrelevant handles and stopwords, followed by k-mer extraction—help transform the raw RNA sequences into a cleaner, more feature-rich format. This enables machine learning models to better capture relevant patterns in the data, improving classification performance and making the data more suitable for downstream analysis.

2) Feature Extraction

The genetic data, especially from viral genomes like the Nipah virus, can be viewed as a language, where the nucleotide sequences convey crucial biological information. Just as Natural Language Processing (NLP) techniques process human languages, we can apply similar methods to analyze the genomic structure of viruses. In this project, we leverage Word2Vec, a popular NLP method, to convert the nucleotide sequences of the Nipah virus genome into numerical vector representations. Word2Vec captures the contextual relationships between genetic elements (such as k-mers) and transforms them into meaningful embeddings that reflect their biological significance.

In particular, Word2Vec offers two model approaches: Continuous Bag of Words (CBOW) and Skip-Gram. The CBOW [22] model predicts the target nucleotide (or k-mer) based on its surrounding context, learning to understand relationships within sequences by focusing on adjacent nucleotides. Conversely, the Skip-Gram model predicts the surrounding context from a given target nucleotide (or k-mer). The Skip-Gram model is particularly effective for modeling rare k-mers or nucleotide sequences that appear less frequently, as it is capable of generating high-quality vector representations even for infrequent k-mers.

In our approach, the genome sequence is divided into 100 nucleotide (nt) segments, and k-mers (such as 3-mers**,** 4-mers, etc.) are extracted from each segment. These k-mers are then processed using the Skip-Gram model, which learns vector representations for each k-mer. This transformation allows us to capture the relationships between different parts of the genome, providing a comprehensive understanding of the viral sequence. These vector embeddings are then used as a preliminary feature matrix for further analysis.

The core goal of this project is to apply deep learning, particularly Recurrent Neural Networks (RNNs)**,** to analyze the Nipah virus genome. The RNN model is trained to predict genetic mutations and assess their potential impact on the virus's pathogenic behavior. By understanding the variations in the viral genome, we aim to improve the accuracy of predictions regarding the virus's evolution and its potential threat to human health.

In the Skip-Gram model, the objective is to maximize the probability of observing the context k-mers given the target k-mer. The model achieves this by adjusting the word vectors such that the dot product between the vector representations of the target k-mer and its context is large when they frequently co-occur and small otherwise. This helps the model learn the relationships between k-mers, facilitating the classification of RNA sequences and helping identify potential promoters or non-promoters in the virus's genetic makeup.

Mathematically, the objective of the skip-gram model can be expressed as follows:

$$Maximize \sum \quad (i = 1)^n \sum \quad (j \in context(i))^m log p(x_i|x_i) \qquad (1)$$

Where n is the number of k-mers in the training dataset, m is the context window size (the number of surrounding k-mers), x_i is the target k-mer, x_{i+j} is the context k-mer for $p(x_{i+j} +, |x_i)$ is the probability of observing the context k-mer x_{i+j} given the target k-mer x_i.

By maximizing this objective, the model learns to create high-quality vector representations for k-mers, which can then be used to predict important genetic features and their relationships, contributing to a deeper understanding of the virus's genomic structure.

3) Classification

A Recurrent Neural Network (RNN) is a class of artificial neural networks designed for sequence prediction and analysis, where the output depends not only on the current input but also on the previous inputs in the sequence. This is especially useful for tasks involving time-series data, natural language processing (NLP), and sequence classification, such as RNA sequence classification.

In this work, we are utilizing RNN architecture to classify RNA sequences as either "Infected" or "Normal." The architecture of the RNN consists of several layers, each serving a specific function to process and learn from the data. The input layer receives the preprocessed RNA sequences, which are represented as scaled k-mers (subsequences of length 6). These sequences are transformed into a format suitable for neural network processing, where each feature is processed sequentially by the layers that follow. The Flatten layer reshapes the input data from a multi-dimensional format into a one-dimensional vector, ensuring the data can be fed into fully connected layers for further analysis. Next, the model includes dense layers, which are fully connected layers that learn to represent the complex relationships between the input features. Each dense layer uses the ReLU (Rectified Linear Unit) activation function, which introduces non-linearity to the model, enabling it to capture intricate patterns within the data.

To avoid overfitting, dropout layers are introduced, randomly disabling a fraction of the neurons during training, which encourages the model to generalize better and prevents it from relying too heavily on any one feature. Finally, the output layer uses a softmax activation function, producing a probability distribution across the two possible classes: "Infected" or "Normal." The class with the highest probability is selected as the model's prediction. Overall, the RNN architecture is designed to effectively learn from the sequential nature of RNA sequences, making it well-suited for this classification task, while the inclusion of dropout helps prevent overfitting and ensures better performance on unseen data.

IV. Results and Discussion

The performance of the prediction model is evaluated using four commonly used metrics: accuracy (ACC), sensitivity (Sn), specificity (Sp), and Matthews Correlation Coefficient (MCC). These metrics are defined as follows:

$$ACC = \frac{TN+TP}{TP+FN+TN+FP} \quad (2)$$

$$Sn = \frac{TP}{TP+FN} \quad (3)$$

$$Sp = \frac{TN}{FP+TN} \quad (4)$$

$$MCC = \frac{TPXTN - FPXFN}{\sqrt{(TN+FP)(TN+FN)(TP+FP)(TP+FN)}} \quad (5)$$

In addition to these metrics, Receiver Operating Characteristic (ROC) curves are plotted to evaluate model performance, where the Area Under the Curve (AUC) is calculated. AUC values range from 0.5 to 1, with higher values indicating better model performance. Precision-Recall (PR) curves are also used to assess the balance between precision (positive predictive value) and recall (sensitivity). Finally, the confusion matrix is used to visually represent the model's classification performance. Cross-validation is employed to ensure robust evaluation by splitting the data into training, validation, and testing sets.

RNN model outperformed XGBoost across all metrics. In terms of Accuracy, the RNN achieved 91.86%, higher than XGBoost's 89.5%, indicating a more accurate overall classification. When examining Sensitivity, which measures the model's ability to correctly identify positive instances, the RNN scored 92.74%, compared to 86.8% for XGBoost, showing that the RNN is significantly better at detecting positive cases, such as "Infected" RNA sequences. Similarly, the RNN also performed better in Specificity, with 91% compared to XGBoost's 87.9%, meaning it was more effective at correctly identifying negative instances. Furthermore, the Matthews Correlation Coefficient (MCC), which balances all four outcomes (true positives, true negatives, false positives, and false negatives), was higher for the RNN at 0.83, compared to 0.74 for XGBoost. A higher MCC reflects better overall performance, particularly in imbalanced datasets. In conclusion, the RNN model consistently outperforms XGBoost in all aspects, making it a more reliable and effective model for RNA sequence classification, especially when accurate detection of both "Infected" and "Normal" sequences is crucial.

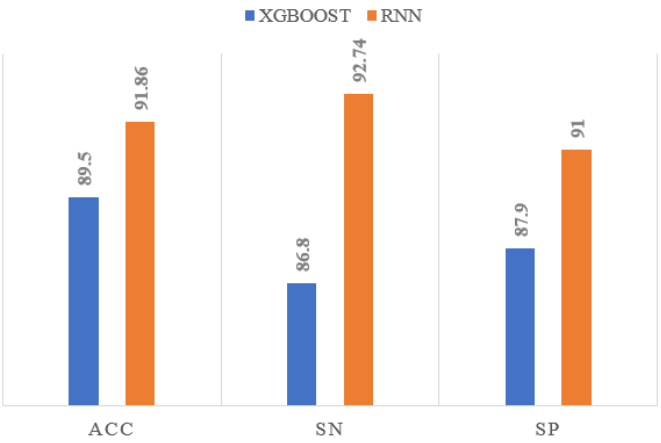

Figure 2. The RNN model comparison with existing method

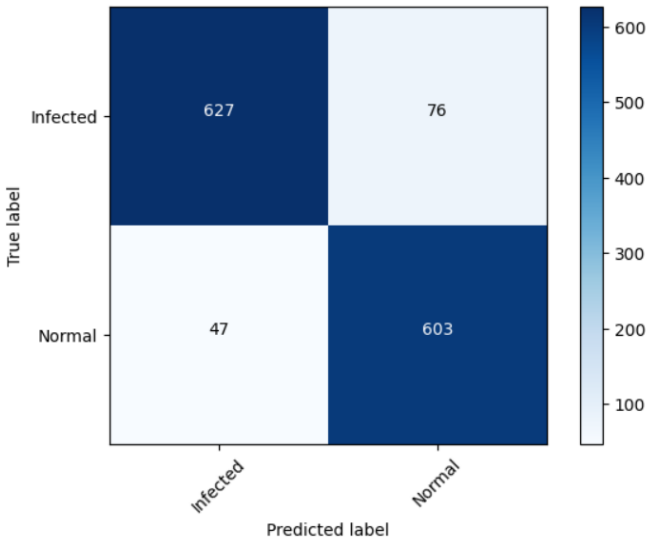

Figure 3. The proposed model's confusion matrix

The Area Under the Curve (AUC) of 0.909 for the Receiver Operating Characteristic (ROC) curve indicates that the model exhibits excellent performance in distinguishing between "Infected" and "Normal". AUC values range from 0 to 1, where a value closer to 1 indicates better discriminatory ability. An AUC of 0.909 suggests that there is a 90.9% chance that the model will correctly rank a randomly selected promoter higher than a randomly selected non-promoter. This indicates that the model is highly effective in identifying promoters in DNA sequences, with a low likelihood of misclassifying them as non-promoters. The ROC curve itself provides a visual representation of the model's ability to distinguish between the two classes by plotting the True Positive Rate (TPR) against the

False Positive Rate (FPR) at various classification thresholds. For a model with an AUC of 0.909, the ROC curve would lie close to the top-left corner, indicating high sensitivity and low false positives. This performance suggests that the model is capable of making accurate predictions with few misclassifications. The threshold setting in the model plays a crucial role in determining the balance between sensitivity and specificity, which could be adjusted depending on the application needs, such as minimizing false negatives or false positives. Overall, the high AUC reflects the strong predictive capability of the model in classifying promoters and non-promoters, which is crucial for applications in genomics and viral pathogenesis prediction.

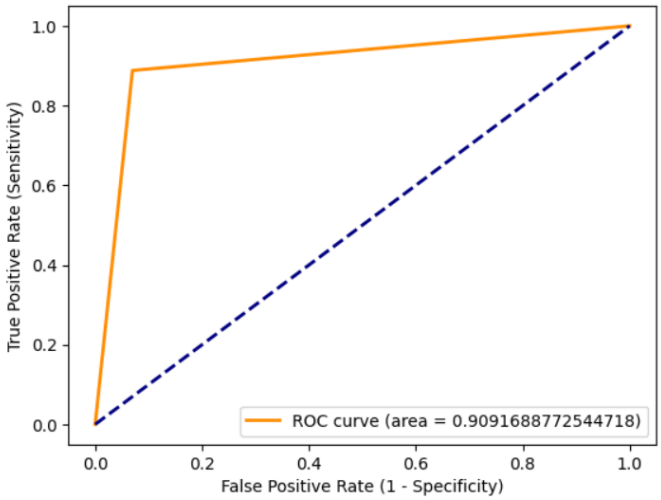

Figure 4. The ROC curves of the RNN model

V. CONCLUSION

In conclusion, this study demonstrates the effectiveness of machine learning models, particularly Recurrent Neural Networks (RNNs), in classifying RNA sequences as either "Infected" or "Normal." The results show that the RNN outperforms other models, such as XGBoost, across all evaluation metrics including accuracy, sensitivity, specificity, and Matthews Correlation Coefficient (MCC). The RNN's superior performance, especially in terms of sensitivity and specificity, highlights its ability to accurately detect both positive and negative cases in RNA sequence classification. These findings suggest that RNNs, with their ability to capture complex patterns in sequential data, are highly suitable for tasks involving biological sequence analysis. Furthermore, the study underscores the importance of preprocessing techniques, such as k-mer extraction and text cleaning, in enhancing the quality of the input data and ensuring reliable model predictions. Future work will explore additional datasets, balanced data handling techniques, and the potential integration of more advanced deep learning architectures to further improve classification accuracy and robustness in real-world applications of RNA sequence classification..

REFERENCES

[1] D. Paul *et al.* Outbreak of an emerging zoonotic Nipah virus: an emerging concern. J. Biosafety Biosecurity(2023)
[2] A.S. Ambat *et al.* Nipah virus: a review on epidemiological characteristics and outbreaks to inform public health decision making. J. Infect. Public Health (2019)
[3] J.H. Epstein *et al.* Nipah virus ecology and infection dynamics in its bat reservoir, Pteropus medius, in Bangladesh. Int. J. Infect. Dis. (2016)
[4] B. Thomas *et al.* Nipah virus infection in Kozhikode, Kerala, South India, in 2018: epidemiology of an outbreak of an emerging disease. Indian J. Community Med. (2019)
[5] Gaurav Sharma, Prashant Singh Rana and Seema Bawa, "Hybrid machine learning models for predicting types of Human T-cell Lymphotropic Virus", IEEE Transactions on Computational Biology and Informatics, pp.1-12, July 2019. [6] Raina K.Plowright, Daniel J.Becker, et al., "Prioritizing surveillance of Nipah virus in India", PLOS Neglected Tropical Diseases, pp.1- 17, June 2019.
[7] Shukla, V.; Rani, S.; Mohapatra, R.K. A New Approach for Leaf Disease Detection using Multilayered Convolutional Neural Network. In Proceedings of the 2023 3rd International Conference on Artificial Intelligence and Signal Processing, AISP 2023, Vijayawada, India, 18–20 March 2023.
[8] Wang, H.-Q.; Li, H.-L.; Han, J.-L.; Feng, Z.P.; Deng, H.X.; Han, X. MMDAE-HGSOC: A novel method for high-grade serous ovarian cancer molecular subtypes classification based on multi-modal deep autoencoder. Comput. Biol. Chem. 2023, 105, 107906. [CrossRef] [PubMed]
[9] Gholami, H.; Mohammadifar, A.; Golzari, S.; Song, Y.; Pradhan, B. Interpretability of simple RNN and GRU deep learning models used to map land susceptibility to gully erosion. Sci. Total. Environ. 2023, 904, 166960.
[10] Aditi and M.Shariff, "Nipah virus Infection: A review", Epidemiology and Infection 147, e95, pp. 1-6, Nov 2019.
[11] Rodolphe Pelissier, Mathieu lampietro and Branka Horvat, "Recent advances in the understanding of Nipah virus immunopathogenesis and anti-viral approaches", Flooo Research, pp. 1-10, 16 Oct 2019.
[12] Akanksha Rajput, Archit Kumar and Manoj Kumar, "Computational Identification of Inhibitors using QSAR Approach Against Nipah Virus", frontiers in Pharmacology, Volume 10, pp.1-9, February 2019
[13] Jorge D.Mello-Roman, Julio C.Mello-Roman, et al., "Predictive Models for the Medical Diagnosis of Dengue: A Case Study in Paraguay", Computational and Mathematical Methods in Medicine, Volume 2019, Article ID 7307803, 7 pages, Hindawi, 2019.
[14] Xiaolu, Yuantao Hao et al., "Using Machine Learning Algorithms to Predict Hepatitis B Surface Antigen Seroclearance", Computational and Mathematical Methods in Medicine, Volume 2019, Article ID 6915850, Hindawi, 2019
[15] Hassan, Md. Z. (2019). Genomic profiling of Nipah virus using NGS driven RNA-Seq expression data. *Bioinformation*, *15*(12), 853–862. https://doi.org/10.6026/97320630015853
[16] Babichev, S., Liakh, I., & Kalinina, I. (2023). Applying a Recurrent Neural Network-Based Deep Learning Model for

Gene Expression Data Classification. *Applied Sciences (Switzerland)*, *13*(21).
[17] Yin, R., Luusua, E., Dabrowski, J., Zhang, Y., & Kwoh, C. K. (2020). Tempel: Time-series mutation prediction of influenza A viruses via attention-based recurrent neural networks.*Bioinformatics*,*36*(9),2697–2704.
[18] Ma, L., Chen, Z., Guan, W., Chen, Q., & Liu, D. (2019). Rapid and specific detection of all known Nipah virus strains' sequences with reverse transcription-loop-mediated isothermal amplification. *Frontiers in Microbiology*, *10*(MAR). https://doi.org/10.3389/fmicb.2019.00418
[19] Muhammad, S. A., Guo, J., Noor, K., Mustafa, A., Amjad, A., & Bai, B. (2023). Pangenomic and immunoinformatics based analysis of Nipah virus revealed CD4+ and CD8+ T-Cell epitopes as potential vaccine candidates. *Frontiers in Pharmacology*,*14*.
[20] Rahman, M. Z., Islam, M. M., Hossain, M. E., Rahman, M. M., Islam, A., Siddika, A., Hossain, M. S. S., Sultana, S., Rahman, M., Klena, J. D., Flora, M. S., Daszak, P., Epstein, J. H., Luby, S. P., & Gurley, E. S. (2021). Genetic diversity of Nipah virus in Bangladesh. *International Journal of Infectious Diseases*,*102*,144–151.
[21] Pelissier, R., Iampietro, M., & Horvat, B. (2019). Recent advances in the understanding of Nipah virus immunopathogenesis and anti-viral approaches. In *F1000Research* (Vol. 8). F1000 Research Ltd.
[22] X. Xiao, Z.-C. Xu, W.-R. Qiu, P. Wang, H.-T. Ge, and K.-C. Chou, ''IPSW(2L)-PseKNC: A two-layer predictor for identifying promoters and their strength by hybrid features via pseudo K-tuple nucleotide composition,'' Genomics, vol. 111, no. 6, pp. 1785–1793, Dec. 2019