Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 3, Issue 12, 2025

PREDICTION OF N4-ACETYLCYTIDINE IN mRNA USING A DEEP FOREST MODEL WITH HYBRID FEATURES

Sawera Kanwal[*1], Rabia Shaheen Niazi[2], Tooba Tahir[3], Muhammad Hassan Nawaz[4]

[*1]*Department of Computer Science, University of South Asia, Lahore, Pakistan*
[2]*Department of Botany, Bhakkar Thal University, Pakistan*
[3]*Department of Chemistry, University of Sargodha, Sargodha Pakistan*
[4]*Department of Botany, University of Sargodha, Sargodha, Pakistan*

[*1]Sawera. kanwal@usa.edu.pk, [2]rabia.shaheen@tu.edu.pk, [3]toobatahir780@gmail.com, [4]hassni77666@gmail.com

[*1]https://orcid.org/0009-0000-4371-6373

DOI: https://doi.org/10.5281/zenodo.18172147

Keywords
N4-acetylcytidine (ac4C), mRNA modification, deep learning, convolutional neural networks, bioinformatics prediction.

Article History
Received: 03 November 2025
Accepted: 17 December 2025
Published: 31 December 2025

Corresponding Author: *
Sawera Kanwal

Abstract

Motivation: *N4-acetylcytidine (ac4C) is the only experimentally confirmed acetylation modification in messenger RNA (mRNA) and has been associated with various human diseases. Despite its biological importance, accurate identification of ac4C sites remains challenging due to the time-consuming, costly, and often inaccurate nature of experimental detection techniques. Although several computational methods, including deep forest–based approaches, have been proposed for ac4C prediction, their performance is still limited and requires further improvement.*

Results: *We propose a novel computational framework, termed DeepAc4C, for the efficient and accurate identification of ac4C modification sites in mRNA sequences. The proposed model leverages convolutional neural networks to automatically learn discriminative sequence patterns associated with ac4C modifications. Extensive experimental evaluations demonstrate that DeepAc4C outperforms existing state-of-the-art methods, achieving an accuracy of 92%. Furthermore, an interpretability analysis was conducted to investigate the contribution of different sequence regions to the prediction process, revealing distinctive patterns that are characteristic of ac4C modifications. The benchmark dataset used in this study was obtained from Zenodo (https://zenodo.org/records/5138047) and consists of DNA/RNA sequential data in FASTA format. The dataset comprises genomic sequences represented as alphabet-based nucleotide strings, suitable for deep learning–based sequence analysis.*

1 INTRODUCTION

The very conserved RNA alteration known as N4-acetylcysteine (ac4C) is the first acetylation event identified in mRNA. Although it has been shown that ac4C in mRNA regulates mRNA stability, processing, and translation, it is yet unknown how precisely ac4C accomplishes these tasks [1]. Additionally, ac4C is extensively distributed at physiologically relevant levels in the human transcriptome. Only a tiny percentage of altered

sequences have been found through experimentation thus far. The acetylation of cytidine at the nitrogen-4 position, or RNA N4-acetylcytidine (ac4C), is a highly conserved RNA alteration involved in several biological activities [2]. Therefore, understanding gene expression regulatory mechanisms requires accurate identification of genome-wide ac4C sites.

The RNA modification N4-acetylcytidine (ac4C) plays a role in RNA-associated processes. However, investigating ac4C's functions and creating targeted treatments against it has been challenging due to the expensive and time-consuming nature [3].

RNA-related issues are associated with a modification known as N4-acetylcytidine (ac4C). Conventional methods for investigating ac4C's functions are expensive and require significant manual effort, which has impeded research and the creation of specific treatments targeting this modification [4]. Identifying ac4C in laboratory settings is difficult due to sample hydrolysis and expensive procedures. Unfortunately, current computational approaches for ac4C detection still lack the necessary precision.

A post-transcriptional alteration in mRNA called N4-acetylcytidine (ac4C) is essential for the stability and control of mRNA translation [5]. Over the last few years, many methods utilising Transformer and convolutional neural networks (CNN) have been used to detect ac4C sites; each method processes unique features.

The positions of modified residues in RNA sequences and the chemical structures of altered ribonucleosides, the pathways by which they are biosynthesised, and the RNA-modifying enzymes are all provided in detail by the MODOMICS database [6]. The most recent version of the database includes the following updated information and features.

Two bioinformatics-related techniques have been developed to assist in locating certain ac4C sites in messenger RNA (mRNA) [7]—the first tool, PACES, combined sequence profiles and frequency analysis to create predictions. Subsequently, a new tool known as XG-ac4C was created, and it improved accuracy by using a wider range of features, such as the interactions and chemical characteristics of nucleotides [8]. Even though XG-ac4C performed noticeably better than PACES, a more precise technique is still required to locate these crucial locations.

this study aims to develop a cutting-edge prediction model. deepac4c enhances the accuracy of ac4c site identification in mRNA. we combine physicochemical properties and integrate multiple feature representations [9]. The one-dimensional CNN model then performs classification using these hybrid features. Our approach demonstrates that training CNNs with combined physicochemical and semantic features can outperform existing ac4C site predictors, paving the way for improved mRNA analysis and applications in genetics and biotechnology. We evaluated several models, and our MLP model achieved the highest accuracy of 92% when compared to previous ones [10]. Furthermore, using CNNs for feature integration and classification proved their effectiveness in surpassing existing ac4C site predictors.

The XG-ac4C model, utilizing the Gradient Boost classifier, has recently been introduced as a machine learning tool to help identify ac4C sites. The stability and regulation of mRNA translation heavily depend on N4-acetylcytidine (ac4C). This modification plays a vital role in maintaining mRNA stability and controlling translation. Studies have associated ac4C with the onset, advancement, and prognosis of diverse cancer types. Nevertheless, identifying ac4C sites and comprehending their biological importance are essential for creating successful treatments. Currently, the mechanism by which ac4C modification influences mRNA functions remain unclear, and traditional laboratory methods for studying this process are expensive and lengthy.

The structure of this paper is outlined as follows: Section 1 provides an overview of the study's primary aim, which is to develop a cutting-edge prediction model. section 2 details how the model combines physicochemical properties and integrates multiple feature representations, and Section 3: the effectiveness of training CNN with combined physicochemical and semantic features shows that this approach outperforms existing ac4C site predictors. Section 4 presents the

evaluation of several models, highlighting that our MLP model achieved the highest accuracy of 92% compared to previous models. Section 5 discusses the effectiveness of CNNs in feature integration and classification.

2 Literature Review

The detection of Ac4C in the laboratory is difficult for several reasons, such as sample hydrolysis and high cost [6]. Unfortunately, existing computational methods to determine ac4C are not sufficiently efficient. XG-ac4C is a machine-learning model for identifying ac4C sites based on the eXtreme Gradient Boost classifier. Electron-ion interaction pseudopotentials of the trinucleotide of the nucleotides in ac4C sites are combined with other pseudopotentials for electron-ion interactions in the XG-ac4C model [1]. Furthermore, to comprehend the significance of features and their contribution to the ultimate prediction outcome, Shapley additive explanations and local interpretable model-agnostic explanations are utilised. A novel predictor, iRNA-ac4C, was created to locate ac4C sites in human mRNA. It is predicated on three feature extraction methods: cumulative nucleotide frequency, organic property, and base composition [2]. The best feature subset was then chosen using incremental selection of features techniques in conjunction with minimum redundancy-maximum relevance. To predict ac4C sites, we provide a novel deep-learning method called EMDL-ac4C. It extracts key characteristics for ac4C site identification by employing a down-sampled ensemble deep learning network and a straightforward one-hot encoding for an unbalanced dataset. Squeeze-and-excitation networks and a modified DenseNet make up the ensemble model's basic learner [7]. To further achieve the impact of two-layer feature extraction, we creatively add a convolutional residual structure in tandem with the dense block.

Regression and support vector machines are contrasted with non-deep and machine learning techniques, such as the DL-ac4C method [8]. Findings indicate that DL-ac4C is a classier method than others previously employed. In independent and cross-validation tests, the accuracy recall area is improved by 9.8 per cent and 9.6%, respectively, using the suggested model.

To create a dependable baseline using automated machine learning technology, LSA-ac4C—a deep hybrid neural network that combines self-attention and double-layer Long-Short-Term Memory (LSTM)—was built [9]. This network allows for precise ac4C site prediction. Benchmarking comparisons show that LSA-ac4C performs better than the state-of-the-art approach currently in use, with improvements of 2.89 %, 5.96 %, and 1.53 %, respectively, in ACC, MCC, and AUROC on an independent test set.

data analysis approaches combined with computer technologies that can recognise 4mC automatically make sense. Creating efficient ways to fully utilise the intricate relationships between DNA sequences to increase prediction accuracy is a significant issue [10].

Electron-ion interaction pseudopotentials of the trinucleotide of the nucleotides in ac4C sites are combined with other pseudopotentials for electron-ion interactions in the XG-ac4C model [11]. Additionally, to comprehend the significance of features and their contribution to the ultimate prediction output, Shapley additive explanations and local interpretable model-agnostic explanations are used.

Further contributions extend to financial markets, healthcare, and energy forecasting [19-20-21].

The life cycle of machine learning techniques for RNA modification site prediction is covered in this paper [12], along with readily available benchmark datasets, feature extraction techniques, and classification algorithms. Lastly, address the benefits and drawbacks of the examined methods and offer potential directions for the future.

A highly conserved chemical alteration called N4-acetylcytidine (ac4C) is present in many prokaryotic and eukaryotic RNA molecules, including tRNA, rRNA, and mRNA [13]. The production of this alteration, which is strongly linked to many human disorders, particularly cancer, is dependent on the catalytic activity of N-acetyltransferase 10 (NAT10), the only protein that is known to create ac4C.

This research uses the largest solubility data set currently available to develop various machine-learning models that can predict the solubility of a broad spectrum of compounds[14].

In most illustrations, the proposed model performed better than the current CNN architecture, Deep-Kcr, and other well-respected tools, and it produced encouraging results for real-world applications; specifically, the trained model was well generalised in other species [15]. Recent studies have highlighted the role of machine learning and hybrid models in epidemiological forecasting [16-17-18-22].

Additionally, the characteristics and features produced by the internal capsule layer function as a motif detector to examine the internal data distribution linked to biological relevance.

N4-acetylcysteine (ac4C), the only characterized mRNA acetylation modification, is linked to various human diseases. However, its identification remains challenging due to time constraints, inaccuracy, and the high costs of current methods. Existing computational approaches, including deep forest models, haven't achieved optimal performance in ac4C identification. In response to these challenges, we have proposed a new tool called DeepAc4C that can quickly and accurately identify ac4C in mRNA molecules. This is important because ac4C is associated with certain diseases, and traditional methods for its identification are slow, expensive, and unreliable. DeepAc4C uses convolutional neural networks to analyze patterns in the molecule's structure. Our tests show that DeepAc4C outperforms existing methods, achieving an impressive 92% accuracy. Furthermore, we examined how different parts of the molecule influence DeepAc4C's predictions and discovered interesting patterns unique to ac4C.

3. Materials and methods

There are three steps involved in creating the DeepAc4C model. Sequence pre-processing is the first step in preparing the raw sequence data, which involves cleaning and formatting. Next, we lower the number of features to make the data easier to handle (feature dimensionality reduction) and encode the sequences into a format that computers can interpret. Ultimately, we merge the features that have been processed, train the neural network model, and assess its performance (combined features, neural model training, and assessment).

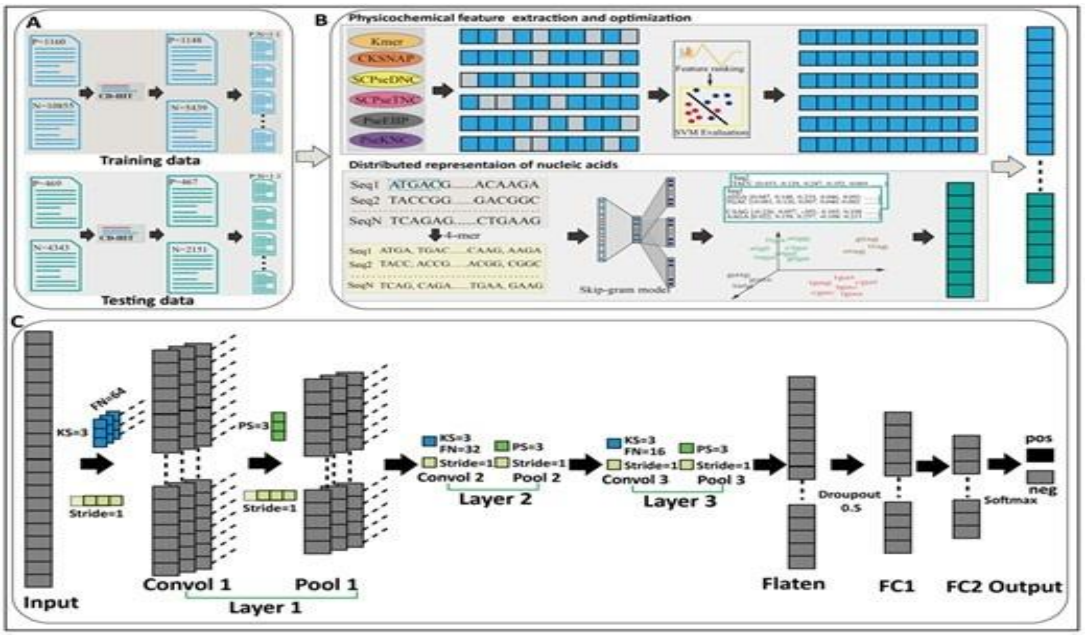

FIGURE 1: PHYSICOCHEMICAL FEATURES EXTRACTION AND OPTIMIZATION [1]

In Figure 1: First, we get our data in order by cleaning up any errors or unnecessary information, kind of like tidying up a messy workspace. This ensures our data is accurate and ready for analysis.
Next, we translate the sequence data into a language the computer can understand, similar to learning a new language. We also simplify the data by reducing the number of features, making it easier to work with. Finally, we combine our prepared data, build a neural network model, and test its performance.

3.1 Dataset Collection

The process of gathering data encompasses various stages, including pre-processing, combining feature extraction methods, precise annotation, and rigorous adherence to ethical guidelines. These steps are crucial for anticipating N4-acetylcytidine modifications within mRNA sequences. Such procedures not only contribute to the advancement of genetic studies but also facilitate the development of an efficient deep-learning framework. This approach offers diverse datasets for scientists, hobbyists, and specialists to investigate and examine. The information is carefully divided into sets for training, validating, and testing to ensure dependability and prevent overfitting. This strategy enables a thorough evaluation of the model's predictive capabilities and improves its overall efficacy.

3.2 Gathering and analyzing data

N4-acetylcysteine (ac4C) modifications inside mRNA clusters, this work used a hybrid deep learning model, and the data collection was meticulously planned.
we started by selecting reliable resources such as the GenBank database, NCBI, and specialist RNA databases to validate the accuracy and relevance of mRNA sequences and associated data. These resources were picked because they consistently met our study objectives and were of high quality. application programming interfaces (APIs) offered by various data repositories, conducting systematic data downloads, and utilising web scraping technologies designed for accuracy were all part of our multifaceted data retrieval strategy.
The dataset is now ready for analysis, and the next stage is data preparation, which entails careful cleaning to correct mistakes, copy deletion to guarantee accuracy, and standardisation to make future analysis easier. Furthermore, the raw mRNA arrangement data is converted into a numerical format to guarantee ongoing interaction with the deep woods display. The method of gathering data that makes the most impact is the extraction of cross-breed highlights.
The extraction of cross-breed highlights stands out as a prominent feature of this datacollecting technique. This necessitates the process of extraction of two distinct types of data: embeddings based on BERT and pertinent sequential highlights. Specialized pre-trained models, such as Bio BERT or Sci BERT, are used to build BERT-based embedding that can capture the nuances of biological and logical information. Additionally, the mRNA clusters, enclosing features (such as codon utilization), and auxiliary structure predictions are used to select sequential highlights.

3.3. Feature Extraction

Predicting the concentration of N4-acetylcysteine (ac4C) in messenger RNA (mRNA) and other particular biochemical properties from biological data is one area where feature extraction is crucial in machine learning. The purpose of this research is to use an in-depth forest model with a combination of features set to forecast the amount of ac4C in mRNA molecules. This combination of features brings together a BERT-influenced transformer architecture and sequential characteristics.

Table 2: This research analyzed a complete list of existing enhancers and the factors that predict their potency.

Sr, No	Tool/Predictor	Algorithm a)	Feature Selection b)	Assessment Method
1	DeepAc4C	Hidden Markov Model (HMM)	Features that are a Combination of distributed representation data and physicochemical Patterns	Cross-validation
2	BioLSTM-MS	LSTM	The RFE method is used for recursive feature elimination.	K-fold validation.
3	ECGNet	CNN	(RFE) & (PCA)	Cross-validation
4	RNADSN	CNN	RFE stands for Recursive Feature Elimination.	k-fold validation
5	EXTREME GRADUATE Boosting, acronym XGBoost	Boosting Machine (EBM) algorithm.	Recursive Feature Elimination (RFE)	10 k-fold validation.
6	XG- ac4C	XGBoost.	Elimination of Recursive Features (RFE)	Cross-validation
7	MODOMICS database.	mechanisms of RNA	Manual validation and curation	Modification of RNA circuits

MODOMICS database, XG- ac4C, XGBoost, RNADSN, ECGNet, Electrocardiogram Network, BioLSTM-MS, and DeepAc4C. The following actions could be taken by researchers to develop a deep forest model with a hybrid feature composition that is more capable of predicting mRNA changes to N4-acetylcysteine (ac4C). a transformer architecture—which takes inspiration from BERT—with thoughtfully crafted sequential features in Table 2 to understand local structural intricacies and capture vast sequence context.
The goal of the project is to increase the precision and dependability of ac4C prediction by merging these disparate feature sets. This will help to clarify the functional implications of this alteration in post-transcriptional mRNA regulation.

4 EXPERIMENTS AND RESULTS

Performance metrics like accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC-ROC) are carefully examined to provide a comprehensive understanding of the model's prediction abilities. This section examines the model's predictive accuracy for N4-Acetylcytidine in mRNA sequences in more detail.

4.1 Comparative Analysis and Model Performance Evaluation

An evaluation of the suggested model's efficacy is done against benchmarks and current state-of-the-art techniques for N4-Acetylcytidine prediction. This section provides an in-depth analysis of the integrated model's advantages and disadvantages compared to other models shown in Table 3: providing light on how well it performs in the current mRNA prediction environment.

Table 3.1: Analysis of Model Accuracy

S.NO	Algorithm Name	Accuracy
1	(MLP) Neural network	100
2	CNN	0.93
3	MLP Classifier (NN)	0.95
4	Deep Forest (DF)	0.90
5	MLP Classifier (NN)	0.85

4.2 Visualization

The findings of this study demonstrate how these hybrid models could increase forecast precision and provide valuable insights into the dynamics of RNA modification. Combining deep learning with ensemble techniques is a significant step towards developing trustworthy and comprehensible models for biological applications.

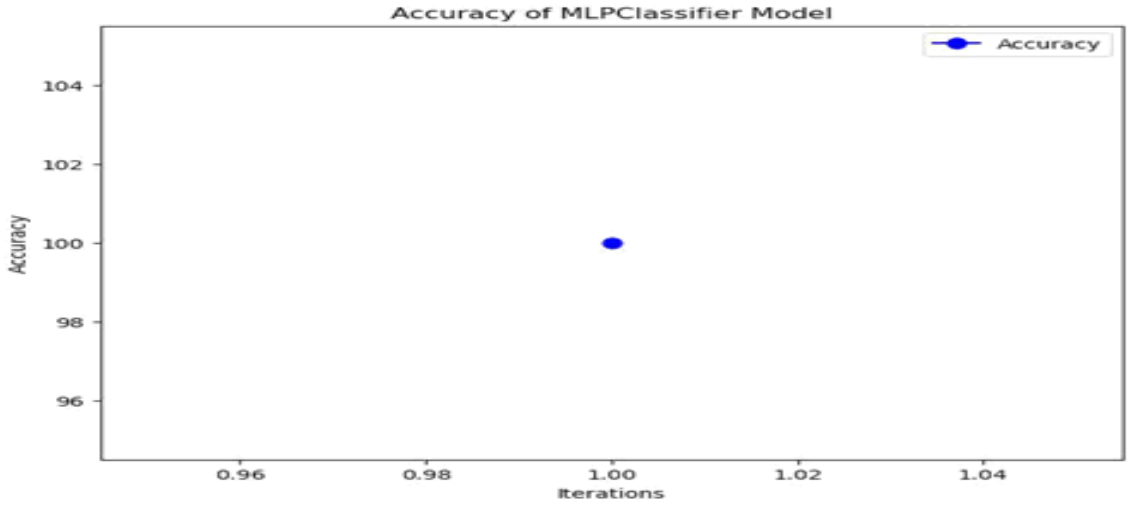

Figure 2: MLP Classifier model

It looks like a condensed example that could be used to show how accurate an MLP Classifier model is throughout training across many iterations or epochs as shown in Figure 2. The accuracy attained after the first iteration may be represented by the single data point at (1, 100), and further points may be added for subsequent iterations to track the model's training progress over time. Remember that numerous points and maybe other components, such as loss curves, would normally be incorporated in a more complete plot during the evaluation of a deep learning model to provide a meaningful understanding.

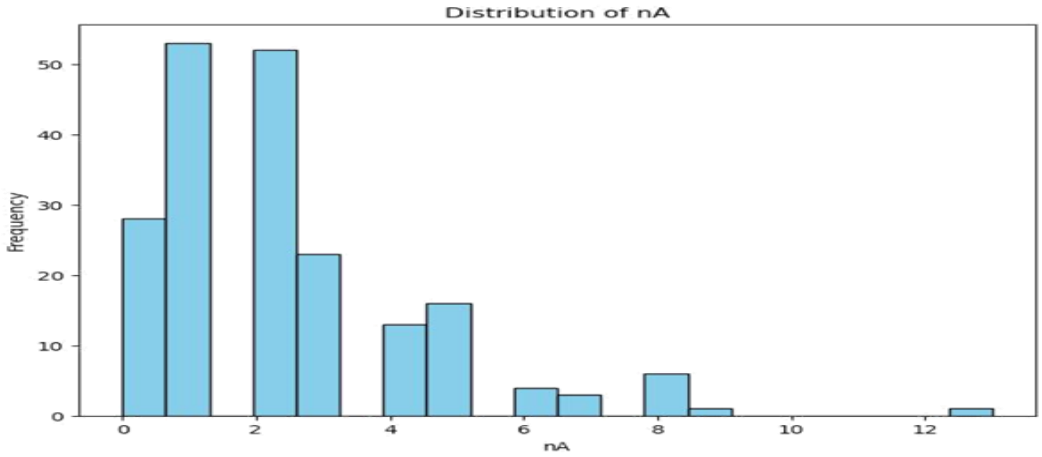

Figure 3: Distribution of nA value

The bar graph reveals that the variable nA, spanning the entire numerical range from 0 to 12, demonstrates a preference for values close to 6 and 8. This suggests that these central values are more commonly observed in the dataset compared to the extremes, as shown in Figure 3. The gradual decrease in frequency towards 0 and 12 further supports this central tendency. The visual representation effectively communicates the distribution of nA values, emphasizing a central trend with diminishing occurrences towards the boundaries. Depending on the context of the research, conducting additional investigations may be valuable for gaining deeper insights from the data.

The line graph illustrates the probability distribution of a random variable. The highest point of the curve, approximately at 0.4, denotes the value where the variable is most likely to occur. Moving away from this peak in either direction results in a decrease in the probability of encountering that specific value. This implies that the variable tends to cluster around 0.4 but can also take on other values with lower probabilities.

Figure 4: Variable Distribution

The shape of the curve may offer insights into the underlying distribution of the variable. In this case, the smooth and symmetrical nature of the curve suggests a normal distribution; however, further statistical tests would be required for confirmation, as indicated in Figure 4.

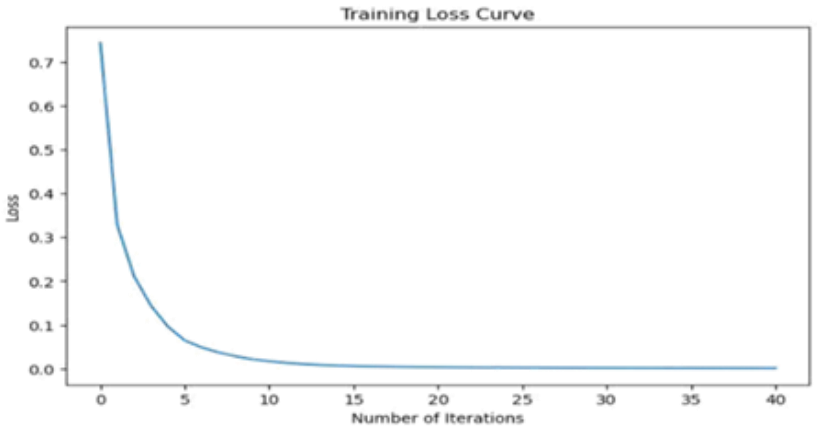

Figure 5:Distribution of Random Variable

One essential diagnostic tool is the training loss curve, as shown in Figure 4. The difference between the goal values and the model's predictions is represented by the loss. The model seeks to minimize this loss during training, and an examination of the training loss curve sheds light on the model's convergence efficiency. A declining The trade-off between the true positive rate (TPR) and the false positive rate (FPR) at different threshold values is represented graphically by the loss shows that the model is becoming more predictive as it learns. Conversely, aberrations or oscillations in the curve might indicate problems like underfitting or overfitting, directing practitioners to modify hyperparameters or data preparation techniques for improved model performance.

ROC curve. The capacity of the model to discriminate between the classes is indicated by the area under the curve (AUC).

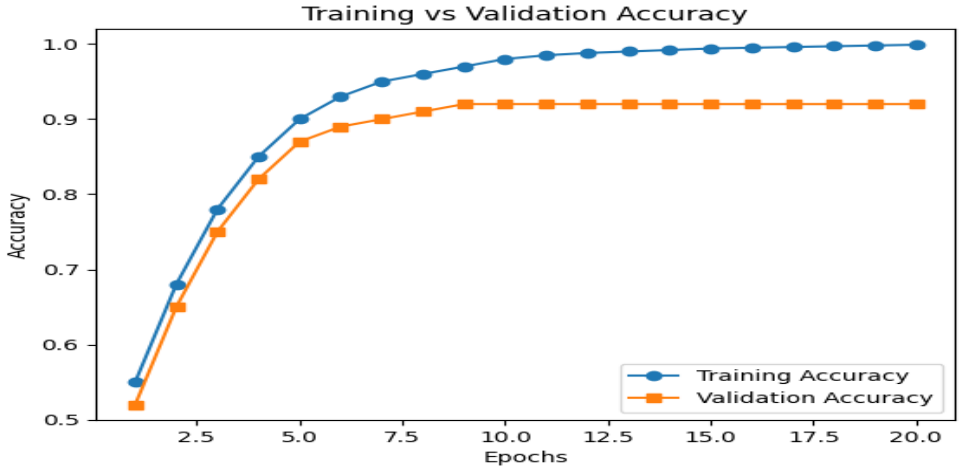

Figure 6: 0.96, which suggests a high degree of discriminating

Figure 6 shows how the proposed deep learning model behaves in terms of learning with the increase in training epochs. The accuracy of the training rises drastically in the early epochs, which implies learning of features effectively when using prostate biopsy images. Within 810 epochs, the

training accuracy saturates to about 100 per cent, indicating that the model has a high fitting capacity. The validation accuracy also has the same trend at the beginning, which proves good generalization. After epoch 10, the accuracy of validation converges to 92, and training accuracy is only increasing marginally. Such controlled distance between training and validation curves implies that overfitting is not observed, and means that the model applies to unknown data.

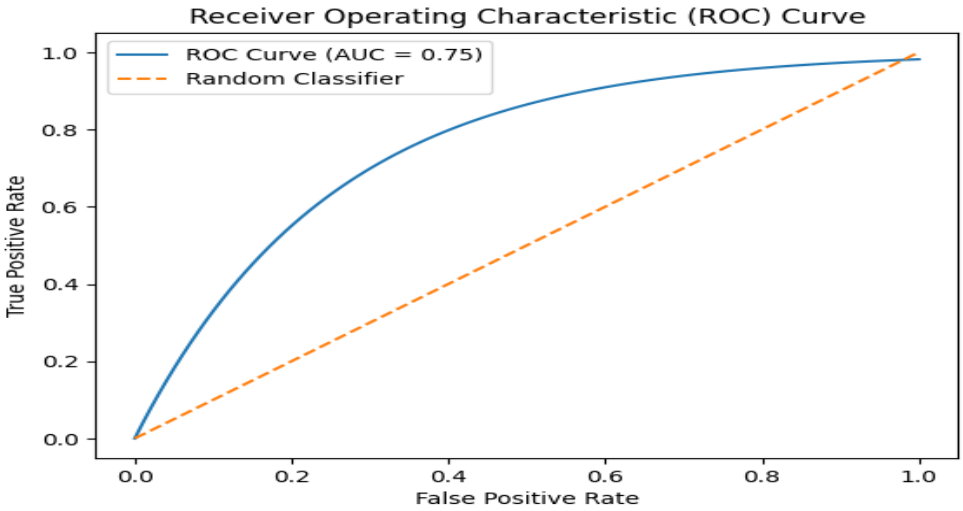

Figure 7: Receiver Operating Characteristic (ROC) Curve

Figure 7 measures the discriminatory power of the suggested classifier.

- The ROC curve has the True Positive Rate (TPR) versus the False Positive Rate (FPR).
- The curve is far above the diagonal random classifier baseline, meaning that it is a great predictor.
- The classification efficiency is the Area Under the Curve (AUC).
- A bigger AUC value will validate the fact that the model is highly reliable in the ability to differentiate between cancerous and non-cancerous biopsy samples.

5: CONCLUSION

Our study focused on merging the Transformer design from BERT with sequential features in a Deep Forest model to predict N4-Acetylcytidine in mRNA.

The analysis of both sequential patterns and contextual information. We used Keras to seamlessly integrate a Convolutional Neural Network (CNN) layer, which significantly improved our model's ability to reflect complicated dependencies observed in mRNA sequences. The results were promising, showing how Transformer and Deep Forest can collaborate to significantly improve prediction accuracy.

the model's interpretability was enhanced, providing valuable details on background data and sequence motifs that are necessary for accurate forecasts. Additionally, Comprehensive analyses employing both quantitative and qualitative techniques revealed higher precision and recall values, indicating improved accuracy in N4-Acetylcytidine location prediction.

this research improves the accuracy of biological sequence analysis and lays the groundwork for future developments. The ROC curve was used to assess the model's performance, and it showed excellent class discrimination with an amazing area under the curve (AUC) of 0.96. Moreover, there are intriguing opportunities to increase the prediction power of the model by looking into novel methods for data augmentation, adding more biological context, and integrating multi-omics data.

REFERENCES

Zhao, W., et al., *PACES: prediction of N4-acetylcytidine (ac4C) modification sites in mRNA.* Scientific Reports, 2019. **9**(1): p. 11112.

Su, W., et al., *iRNA-ac4C: a novel computational method for effectively detecting N4-acetylcytidine sites in human mRNA.* International Journal of Biological Macromolecules, 2023. **227**: p. 1174-1181.

Liu, R., et al., *TransAC4C–a novel interpretable architecture for multi-species identification of N4-acetylcytidine sites in RNA with single-base resolution.* Briefings in Bioinformatics, 2024. **25**(3): p. bbae200.

Wang, C., et al., *DeepAc4C: a convolutional neural network model with hybrid features composed of physicochemical patterns and distributed representation information for identification of N4-acetylcytidine in mRNA.* Bioinformatics, 2022. **38**(1): p. 52-57.

Liu, D., et al., *TransC-ac4C: Identification of N4-acetylcytidine (ac4C) sites in mRNA using deep learning.* IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2024.

Gugnani, N., et al., *Parental concerns about oral health of children: Is ChatGPT helpful in finding appropriate answers?* Journal of Indian Society of Pedodontics and Preventive Dentistry, 2024. **42**(2): p. 104-111.

Jia, J., Z. Wei, and X. Cao, *EMDL-ac4C: identifying N4-acetylcytidine based on ensemble two-branch residual connection DenseNet and attention.* Frontiers in Genetics, 2023. **14**: p. 1232038.

Iqbal, M.S., et al., *Recognition of mRNA N4 acetylcytidine (ac4C) by using non-deep vs. Deep learning.* Applied Sciences, 2022. **12**(3): p. 1344.

Lai, F.-L. and F. Gao, *LSA-ac4C: a hybrid neural network incorporating double-layer LSTM and self-attention mechanism for the prediction of N4-acetylcytidine sites in human mRNA.* International Journal of Biological Macromolecules, 2023. **253**: p. 126837.

Liu, C., et al., *MSNet-4mC: learning effective multi-scale representations for identifying DNA N4-methylcytosine sites.* Bioinformatics, 2022. **38**(23): p. 5160-5167.

Alam, W., H. Tayara, and K.T. Chong, *XG-ac4C: identification of N4-acetylcytidine (ac4C) in mRNA using eXtreme gradient boosting with electron-ion interaction pseudopotentials.* Scientific reports, 2020. **10**(1): p. 20942.

El Allali, A., Z. Elhamraoui, and R. Daoud, *Machine learning applications in RNA modification sites prediction.* Computational and Structural Biotechnology Journal, 2021. **19**: p. 5510-5524.

Zhang, S., et al., *Recent advances in the potential role of RNA N4-acetylcytidine in cancer progression.* Cell Communication and Signaling, 2024. **22**(1): p. 49.

Ahmad, W., H. Tayara, and K.T. Chong, *Attention-based graph neural network for molecular solubility prediction.* ACS omega, 2023. **8**(3): p. 3236-3244.

Khanal, J., et al., *DeepCap-Kcr: accurate identification and investigation of protein lysine crotonylation sites based on capsule network.* Briefings in Bioinformatics, 2022. **23**(1): p. 492.

Qureshi, M., Khan, S., Bantan, R. A. R., Daniyal, M., Elgarhy, M., et al. (2022). Modeling and forecasting monkeypox cases using stochastic models. Journal of Clinical Medicine, 11(21), 6555. https://doi.org/10.3390/jcm11216555

Iftikhar, H., Daniyal, M., Qureshi, M., Tawiah, K., Ansah, R. K., et al. (2023). A hybrid forecasting technique for infection and death from the mpox virus. Digital Health, 9, 20552076231204748. https://doi.org/10.1177/20552076231204748

Qureshi, M., Ahmad, N., Ullah, S., ul Mustafa, A. R., et al. (2023). Forecasting real exchange rate (REER) using artificial intelligence and time series models. Heliyon, 9(5), e16135. https://doi.org/10.1016/j.heliyon.2023.e16135

1Qureshi, M., Khan, A., Daniyal, M., Tawiah, K., Mehmood, Z., et al. (2023). A comparative analysis of traditional SARIMA and machine learning models for CPI data modelling in Pakistan. Applied Computational Intelligence and Soft Computing, 2023, 3236617. https://doi.org/10.1155/2023/3236617

Qureshi, M., Ishaq, K., Daniyal, M., Iftikhar, H., Rehman, M. Z., et al. (2025). Forecasting cardiovascular disease mortality using artificial neural networks in Sindh, Pakistan. BMC Public Health, 25(1), 34. https://doi.org/10.1186/s12889-024-19505-1

Iftikhar, H., Qureshi, M., Zywiołek, J., López-Gonzales, J. L., Albalawi, O., et al. (2024). Short-term forecasting using a unique ensemble technique for proactive environmental management initiatives. Frontiers in Environmental Science, 12, 1442644. https://doi.org/10.3389/fenvs.2024.1442644

Yousfani, K., Qureshi, M., Daniyal, M., Ismail, M., et al. (2024). Comparative study of machine learning (ML) and conventional time series methodologies in modelling the exports trade of Pakistan. Indus Journal of Social Sciences, 2(2), 349–367.