RESEARCH ARTICLE

A Deep Learning Approach for Long Non-coding RNA Identification in Plants: DeepPlnc V2.0

Siddhi Pandey[1] **and Rahul Sharma**[2,*]

[1] School of Bio Science & Bio Engineering, D. Y. Patil International University, Pune 411044, India
[2] School of Computer Science, Engineering & Applications, D. Y. Patil International University, Pune 411044, India

Abstract

Long non-coding RNAs (lncRNAs) are important for plant growth, how plants respond to stress, and their overall development. However, it can be difficult to identify them accurately because they come in many structures and can look similar to coding RNAs. In this study, we introduce DeepPInc V2.0, a new tool that uses deep learning to analyze both the sequence and the secondary structure of RNAs, combining them in a DenseNet-CNN hybrid model. DeepPInc V2.0 outperforms existing tools on various plant datasets, achieving an accuracy of 94.2%, an F1-score of 0.93, and a Matthews Correlation Coefficient (MCC) of 0.88. It consistently outperforms seven leading tools in this area. Importantly, the model still works well even with incomplete or shortened sequences, which is a common issue in transcriptome studies. When we applied DeepPInc V2.0 to the wheat (Triticum aestivum) transcriptome under heat stress, we found over 27,000 possible lncRNAs. Of these, 1,830 were expressed differently, suggesting they may play a role in helping plants adapt to stress. These results demonstrate that DeepPInc V2.0 is a reliable and accurate platform for identifying lncRNAs in plants, facilitating the study of large RNA collections and understanding the functions of non-coding elements.

Keywords: long non-coding RNAs, ensembl, pncStress, PIncDB, DeepPInc.

1 Introduction

Broadly, RNAs are categorized into coding RNAs and noncoding RNAs (ncRNAs). Coding RNAs, primarily messenger RNAs (mRNAs), serve as templates for protein synthesis and are indispensable for cellular and organismal functions [1]. In contrast, ncRNAs, though lacking protein-coding potential, exert essential regulatory roles in gene expression, development, and stress responses. These include microRNAs (miRNAs) that mediate mRNA degradation or translational repression, long noncoding RNAs (lncRNAs) that regulate chromatin architecture and transcriptional programs, small nucleolar RNAs (snoRNAs) that direct chemical modifications of ribosomal and transfer RNAs, and circular RNAs (circRNAs) that function as molecular sponges or regulators of gene expression [2, 3] (Figure 1).

In the 18th century, Felice Fontana's observations of a

Submitted: 19 May 2025
Accepted: 08 September 2025
Published: 30 September 2025

Vol. 1, **No.** 2, 2025.
10.62762/BISH.2025.421075

***Corresponding author:**
✉ Rahul Sharma
prof.rahuls@gmail.com)

Citation
Pandey, S., & Sharma, R. (2025). A Deep Learning Approach for Long Non-coding RNA Identification in Plants: DeepPlnc V2.0. *Biomedical Informatics and Smart Healthcare*, 1(2), 79–88.

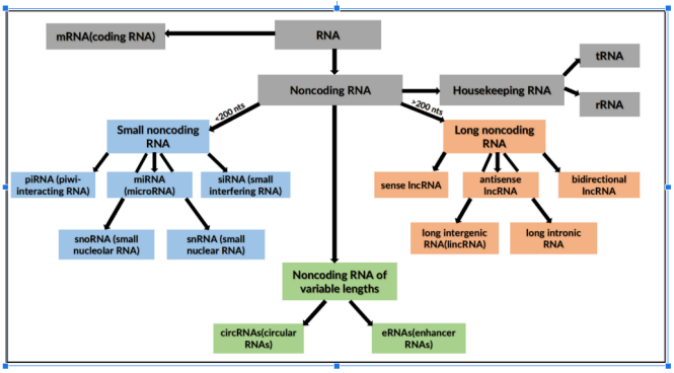


Figure 1. RNA along with its types and further classification spaces.

cellular structure in eel epidermis containing ncRNA laid the groundwork for our understanding of the nucleus and nucleolus. LncRNAs, initially dismissed as transcriptional noise, gained biological relevance following the discovery of the H19 gene during mouse embryogenesis [4, 5]. Sharing structural features with mRNAs, lncRNAs are transcribed by RNA polymerase II in animals and by Pol II, IV, and V in plants, and encompass diverse subtypes including intergenic lincRNAs, antisense transcripts, and intronic lncRNAs [6–9]. Collectively, these discoveries underscore RNA's expanded functional repertoire beyond protein coding, establishing its central role in gene regulation, development, and disease biology. Modern molecular biology, aided by technologies like microarrays and high-throughput sequencing, has since revealed that vast portions of eukaryotic genomes are transcribed, leading to the discovery of a myriad of previously unknown ncRNAs. Initially, many of these were dismissed as "transcriptional noise" due to a lack of sequence conservation and observable phenotypic changes upon their disruption.

However, systematic functional studies over the last decade have brought attention to the crucial roles of lncRNAs, which are non-coding transcripts longer than 200 nucleotides (Figure 2).

LncRNAs are now recognized as key players in diverse regulatory functions, including chromatin modification, transcriptional regulation, RNA processing, and post-transcriptional gene regulation (Figure 2). For example, in Arabidopsis thaliana, the lncRNA COOLAIR recruits the PRC2 complex to repress flowering [10]. Similarly, in maize, the lncRNA IPS1 regulates phosphate homeostasis by sequestering miR399 [11].

The identification of plant-specific lncRNAs remains a challenge, with existing tools often lacking

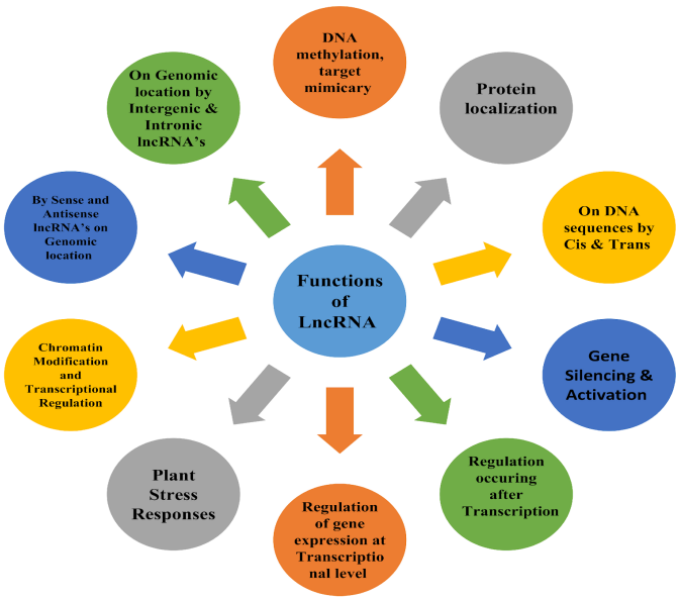


Figure 2. Functions of lncrna.

reliability and consistency due to their reliance on properties identified in animal datasets. While some plant-specific tools like PinePRO, RNAplone, PreLnc, and CNIT have been developed, they often underperform when confronted with incomplete and unannotated sequences, which are common in real-world transcriptomic data (Figure 3).

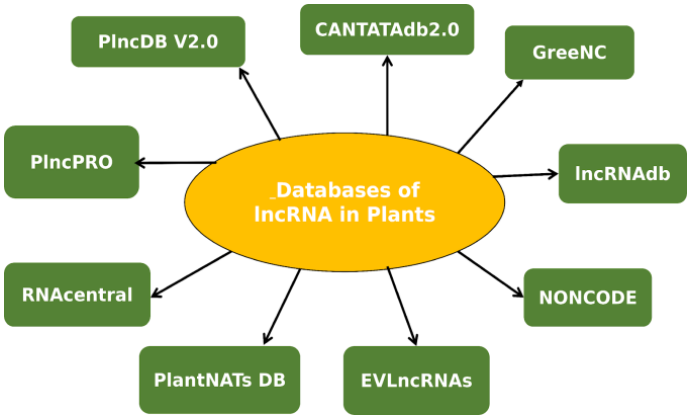


Figure 3. Databases of lncRNAs in plants.

This gap highlights the need for advanced methods, such as deep learning, which can capture the subtle, transient features that traditional machine learning approaches often miss.

2 Materials and Methods

2.1 DeepPInc V2.0 Model

The study introduces a novel deep learning model, DeepPInc V2.0, which combines a DenseNet-based model with an evolved CNN

technique. The architecture is designed to handle both nucleotide sequences and their secondary structural representations simultaneously. One component of the DenseNet comprises 121 layers, including convolutional layers, pooling layers, and dense blocks, which process the nucleotide sequences (Figure 4).

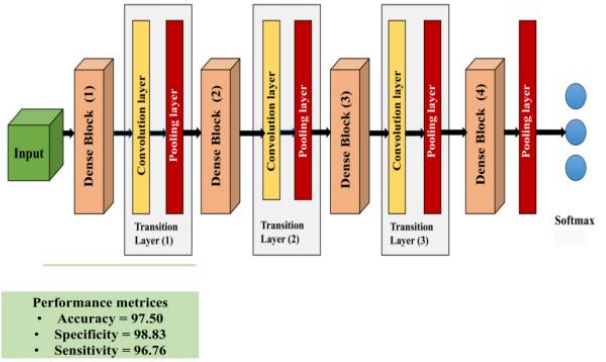


Figure 4. Detailed pipeline of the DenseNet V2.0.

The other component processes the dot-bracket secondary structural representations of the RNAs. This structural data is obtained using RNAfold [12], which provides the Dot-Bracket representation of RNA secondary structures (Figure 5). The Dot-Bracket notation is then encoded using "one-hot" encoding. Each character in the Dot-Bracket representation ('.', '(', and ')') is converted into a binary vector, representing the presence of unpaired bases ('.'), left-hand paired bases ('('), and right-hand paired bases (')'). By integrating both sequence and structural data, DeepPInc V2.0 ensures a comprehensive representation of RNA molecules, which is essential for accurate classification of lncRNAs in plants. The model's architecture undergoes continuous refinement to enhance its ability to characterize plant lncRNAs. The features generated by both components are concatenated and processed through additional layers, including batch normalization and a final output layer, to yield a confidence probability for lncRNA identification in 400-base windows.

2.2 Hyper-parameter Optimization

Hyper-parameter optimization is critical in achieving the best performance for DeepPInc V2.0. This process involves Bayesian optimization methods to ensure a thorough and efficient exploration of the hyper-parameter space. Bayesian optimization builds on the results of the random search by creating a probabilistic model of the objective function. This model is used to predict the performance of different hyper-parameter sets, allowing for

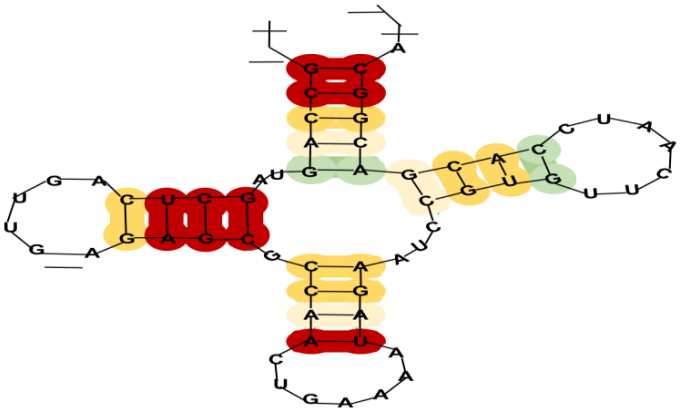


Figure 5. RNA secondary structure (Folding) prediction using RNAfold.

a more focused and efficient search. Bayesian optimization iteratively refines the hyper-parameters by balancing exploration (testing new configurations) and exploitation (focusing on configurations known to perform well), ultimately converging on the optimal set [13].

The hyper-parameters optimized for DenseNet 2.0 approach include:

- Number of Layers: The total number of layers in the DenseNet architecture.
- Filter lengths: The size of the convolutional filters used in the dense blocks.
- Kernel sizes: The dimensions of the filter or window used in convolutional operations in CNNs.
- Strides: The step size for convolutional operations.
- Activation Functions: Functions such as ReLU (Rectified Linear Unit) and Leaky ReLU that introduce non-linearity into the network.
- Dense Block Configuration: The architecture and arrangement of layers within a dense block in a DenseNet.
- Transition Layers: Transition layers are inserted between consecutive dense blocks. These transition layers serve to down sample feature maps and reduce the number of channels, effectively compressing information before passing it to the next dense block.
- Growth Rate: The growth rate in DenseNet refers to the number of additional feature maps (channels) that are added to the output of each layer within a dense block.

- Compression Factor: The factor by which the number of feature maps is reduced between dense blocks.

Each hyper-parameter plays a crucial role in determining the performance and efficiency of the DeepPInc 2.0 model. By optimizing these hyper-parameters using Bayesian optimization algorithm, we ensured that the model achieved the best possible performance while maintaining efficiency and generalizability [14]. The optimization process involved adjusting settings such as activation functions, loss functions, optimizers, and the number of channels in the CNN (Figure 6) and max-pooling layers. Automated search methods, including grid search, random search, and Bayesian hyper-parameter search, were utilized to overcome the limitations of manual search.

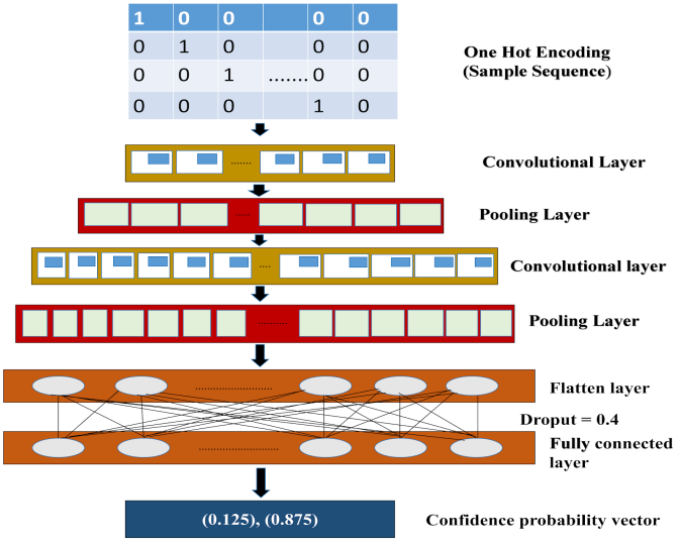


Figure 6. Architecture of CNN (Convolution Neural Network).

2.3 Performance Evaluation Criteria

The model's performance was assessed on test sets using standard evaluation metrics. Confusion matrices were constructed to differentiate between correctly and incorrectly identified instances in the test set. Sensitivity indicates the proportion of positive instances correctly identified, while Specificity denotes the proportion of negative instances correctly identified. Precision assesses the ratio of correctly identified positives to the total of true and false positives [15]. F1-score, which balances precision and recall, was also calculated. Additionally, Matthews Correlation Coefficient (MCC) was used, considering all four confusion matrix classes. A higher MCC score indicates a more robust and balanced model with greater consistency in performance.

Performance metrics were computed using the following formulas:

$$Acc = \frac{TN + TP}{TN + TP + FN + FP} \tag{1}$$

$$Specificity(Sp) = \frac{TN}{TN + FP} \tag{2}$$

$$Sensitivity(Sn) = Recall = \frac{TP}{TP + FN} \tag{3}$$

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

$$F1 - Score = 2 \times \left(\frac{Precision \times Recall}{Precision + Recall} \right) \tag{5}$$

$$AUC = \int_0^1 Pr[TP](v)dv \tag{6}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{7}$$

Where: TP = True Positive, TN = True Negatives, FP = False Positives, FN = False Negatives, Acc = Accuracy, AUC = Area Under Curve (AUC) from receiver operating characteristic (ROC) curves, F1-score is a harmonic average of sensitivity and precision. MCC indicates a correlation coefficient between the true classes and the predicted classes.

2.4 Dataset and Benchmarking Criteria

DeepPInc V2.0 was trained and validated on a meticulously curated dataset (Dataset "A") (Figure 7) containing experimentally validated lncRNA and non-lncRNA sequences sourced from comprehensive databases like Ensembl Plants, PncStress, and PIncDB V2.0. (Table 1) The model was benchmarked against seven existing tools.

Table 1. List of databases for lncRNAs.

S.No.	Databases	Data reported (lncRNA)	Years
1	PNRD	Computationally predicted	2014
2	PLNlncRbase	Experimentally validated	2015
3	Ensembl plants	Computationally predicted and Experimentally validated	2016
4	CANTATAdb V2.0	Computationally Predicted	2019
5	PreStress	Experimentally validated	2020
6	GreeNC 2.0	Computationally Predicted	2022

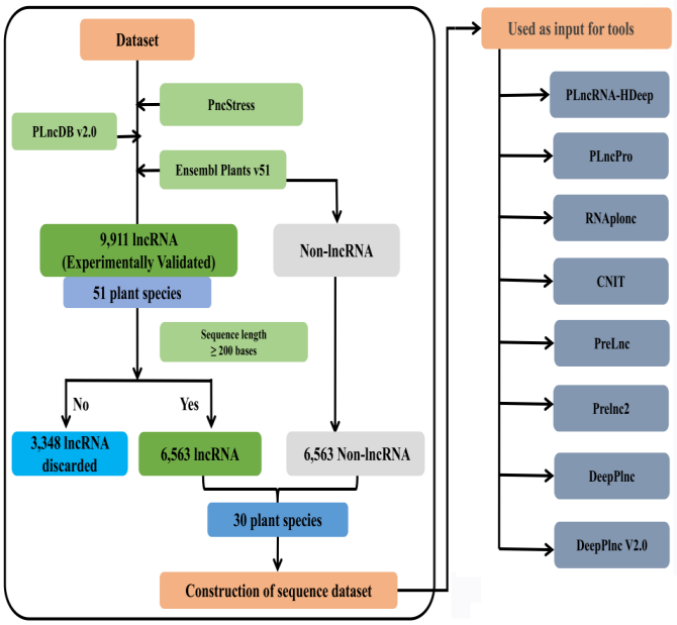


Figure 7. Construction of Dataset 'A' for training and evaluation.

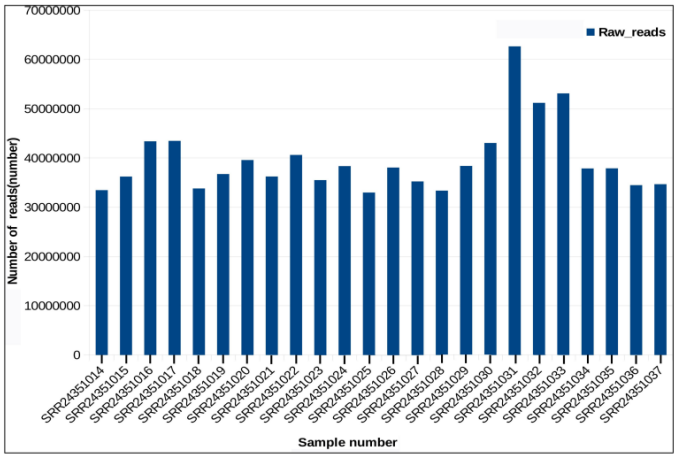


Figure 8. Number of raw reads in each sample.

2.5 Application on Wheat Transcriptomes

To demonstrate its practical application, DeepPInc V2.0 was used to identify lncRNAs from the transcriptome of winter wheat (Triticum aestivum), a species with no existing information regarding lncRNAs. Transcriptomic data was downloaded from the ENA Browser (Project ID: PRJNA963171), consisting of 24 samples (Figure 8) from the winter wheat cultivar "Shiluan02-1" at various stages of vernalization (Table 2).

The NGS pipeline (Figure 9) involved quality checks with FastQC and Multiqc. (Figures 10A and 10B) This step helps in identifying any potential issues such as adapter contamination, over-represented sequences, or low-quality reads that may affect downstream analyses. FastQC generates detailed reports for each sample, highlighting areas that require attention. Multiqc [16] is being done after this step for getting a single report of the separated reports generated by fastqc.

lncRNA candidates were identified. After filtering for sequences with a minimum length of 200 bases, 34,597 sequences were used as input for DeepPInc V2.0.

Proceeding further we did trimming with Trimmomatic, (Figure 11) which employs various parameters to filter out reads below a certain quality threshold and trim any residual adapter sequences, ensuring that only high-quality reads (quality score >30) are retained for further analysis.

Adding on we performed alignment with Hisat2. (Figure 12) HISAT2 aligns the reads to the reference sequences, allowing for accurate mapping and quantification of gene expression levels. The aligned reads were assembled into transcripts using Strawberry. From the assembled transcriptome, which contained 81,879 transcript sequences, 40,430 potential lncRNAs were identified.

Lastly, Transcriptome profiling during vernalization in winter wheat identified 80,442 expressed genes, with 41,449 ubiquitously expressed across all samples. Differential expression analysis revealed extensive transcriptional reprogramming, with 37,615 DEGs after vs before vernalization, highlighting major shifts from vegetative to reproductive states. Stage-specific

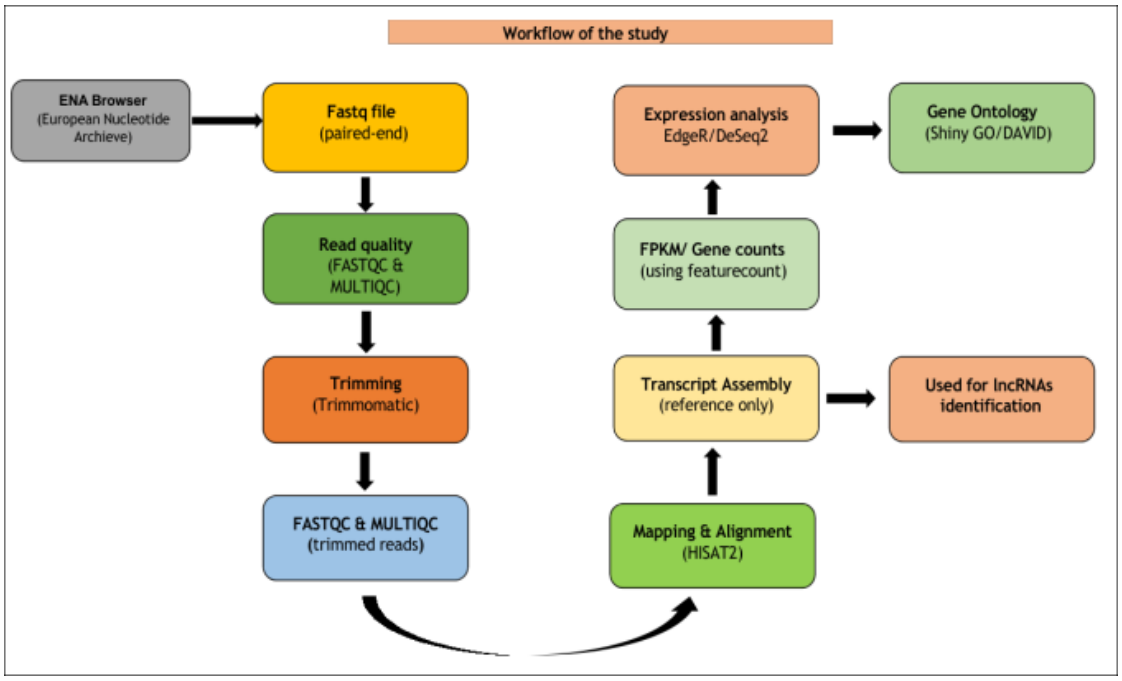


Figure 9. Flowchart of NGS Pipeline for transcriptomic data.

Table 2. List of the samples considered at 8 timepoints in the vernalization phase.

Run accession	Description
SRR24351037	Before vernalization
SRR24351036	Before vernalization replicate 1
SRR24351025	Before vernalization replicate 2
SRR24351023	After vernalization
SRR24351022	After vernalization replicate 1
SRR24351020	After vernalization replicate 2
SRR24351019	Vernalization 7 days
SRR24351018	Vernalization 7 days replicate 1
SRR24351021	Vernalization 7 days replicate 2
SRR24351017	Vernalization 14 days
SRR24351016	Vernalization 14 days replicate 1
SRR24351015	Vernalization 14 days replicate 2
SRR24351014	Vernalization 21 days
SRR24351035	Vernalization 21 days replicate 1
SRR24351034	Vernalization 21 days replicate 2
SRR24351033	Vernalization 28 days
SRR24351032	Vernalization 28 days replicate 1
SRR24351031	Vernalization 28 days replicate 2
SRR24351030	Vernalization 35 days
SRR24351029	Vernalization 35 days replicate 1
SRR24351028	Vernalization 35 days replicate 2
SRR24351027	Vernalization 42 days
SRR24351026	Vernalization 42 days replicate 1
SRR24351024	Vernalization 42 days replicate 2

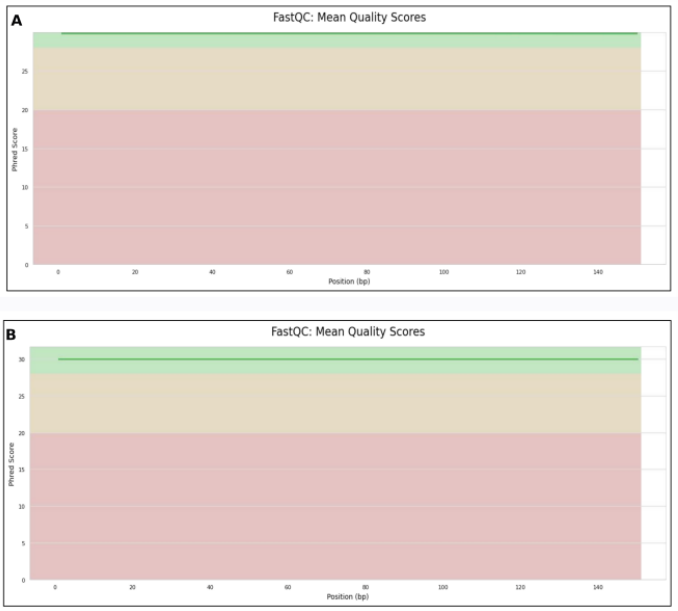


Figure 10. A) MultiQC before trimming, B) MultiQC after trimming.

responses included up to 41,774 DEGs during mid-vernalization and dynamic changes across early (7–14 h), middle (21–35 h), and late (42 h) stages. Comparisons showed both upregulated (6,707) and downregulated (5,335) genes, reflecting robust regulation. Overall, results underscore dynamic, stage-specific transcriptional adjustments enabling adaptation to prolonged cold and transition toward flowering (Table 3) (Figure 13).

These volcano plots display differentially expressed genes (DEGs) with $\log_2$ fold change on the X-axis

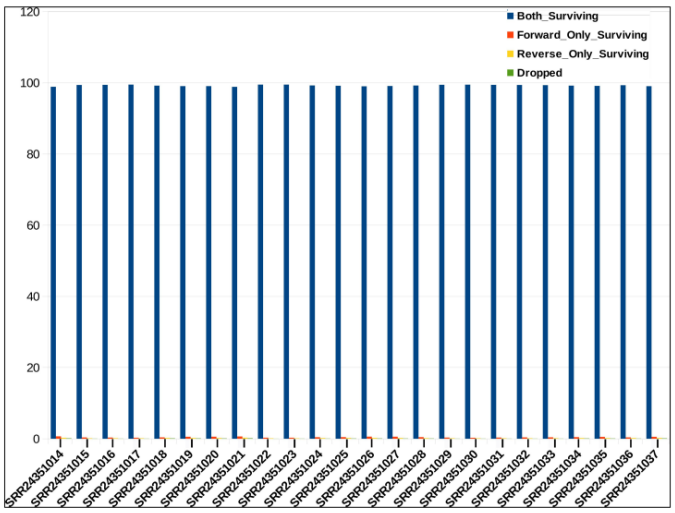


Figure 11. Graph of trimmed output. Dark blue bar indicates both surviving paired end reads.

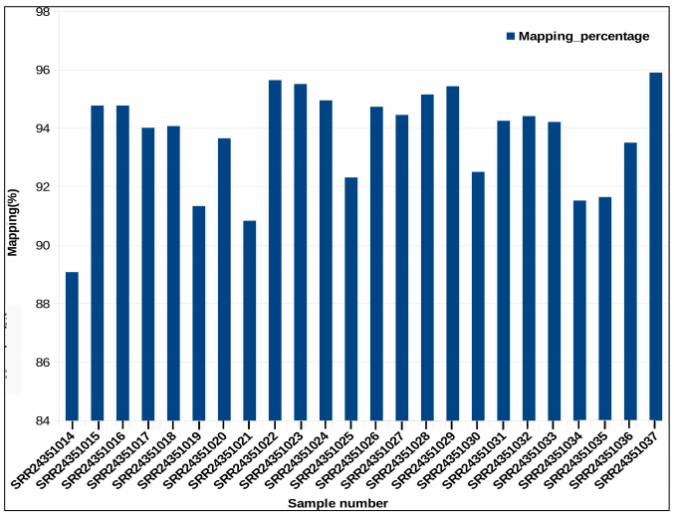
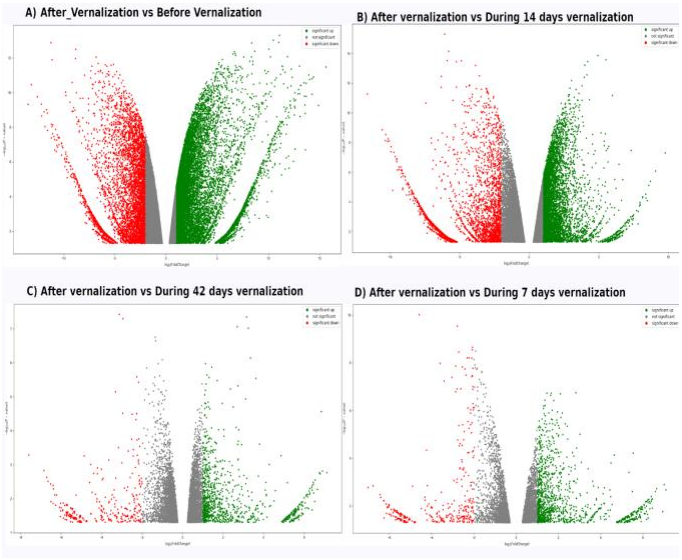
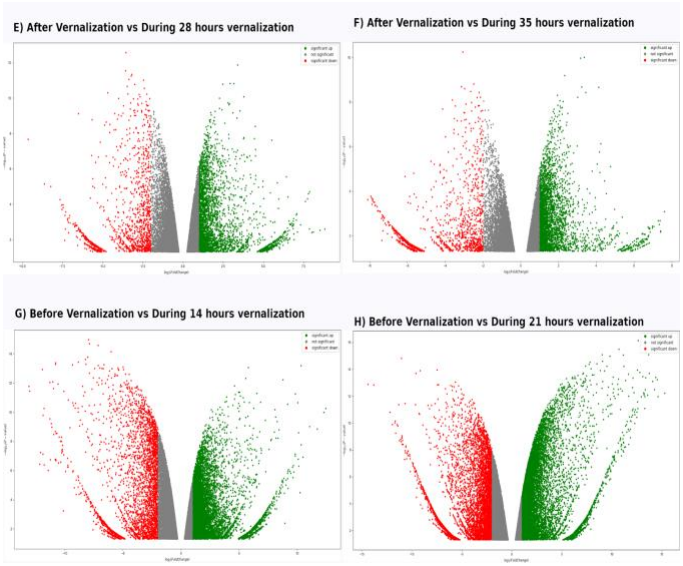


Figure 12. Graph of mapping samples output indicating overall alignment rate.



(a)



(b)

Figure 13. Volcano plots of differentially expressed genes (DEGs).

and $-\log_{10}$ adjusted P-value on the Y-axis. Gray dots represent non-significant genes, while red and blue dots denote significantly upregulated and downregulated genes, respectively, based on thresholds of $\log_2$ fold change ≥ 2 or ≤ -2 and adjusted P-value < 0.05. Threshold lines highlight significance, and key genes may be labeled for emphasis on their biological relevance.

To clearly represent the stepwise workflow of our approach, the pseudo-code of DeepPInc V2.0 is presented in Algorithm 1.

3 Results

DeepPInc V2.0 consistently outperformed seven leading lncRNA prediction tools, achieving 92.4% accuracy and 0.88 MCC, compared to the best baseline accuracy of 84.7% (MCC = 0.71) (Figures 14 and 15). Notably, its performance remained stable ($> 80\%$ accuracy) even on truncated transcripts, where competing methods showed significant drops, confirming its robustness [17, 18].

The benchmarking results demonstrated that DeepPInc V2.0 consistently outperformed seven existing tools, with a superior accuracy and Matthews Correlation Coefficient (MCC). (Figures 14 and 15) Its ability to handle truncated sequences highlighted its robustness and precision. The application of DeepPInc V2.0 to the wheat transcriptome successfully identified lncRNAs in a species where no such information was previously available. From the 34,597

Table 3. List of the DEGs within different pairs of conditions.

Conditions	Total DEGs	Upregulated	Downregulated
after_vs_before	6707	5335	12042
after_vs_during14	1127	1250	2377
after_vs_during28	394	323	717
after_vs_during35	452	345	797
after_vs_during42	3307	3435	6742
after_vs_during7	2045	2963	5008
before_vs_during14	6215	5433	11648
before_vs_during21	6094	6176	12270
before_vs_during28	7013	5334	12347
before_vs_during35	7554	5489	13043
before_vs_during42	3109	2282	5391
before_vs_during7	5596	6211	11807
during14_vs_21	735	1683	2418
during14_vs_28	906	798	1704
during14_vs_35	1952	1253	3205
during14_vs_42	4204	4070	8274
during21_vs_28	2064	748	2812
during21_vs_35	2079	923	3002
during21_vs_42	4767	4117	8884
during28_vs_35	766	541	1307
during28_vs_42	3643	3731	7374
during35_vs_42	3342	3962	7304
during7_vs_14	1932	639	2571
during7_vs_21	839	786	1625
during7_vs_28	3633	2079	5712
during7_vs_35	4101	2228	6329
during7_vs_42	4102	4588	8690

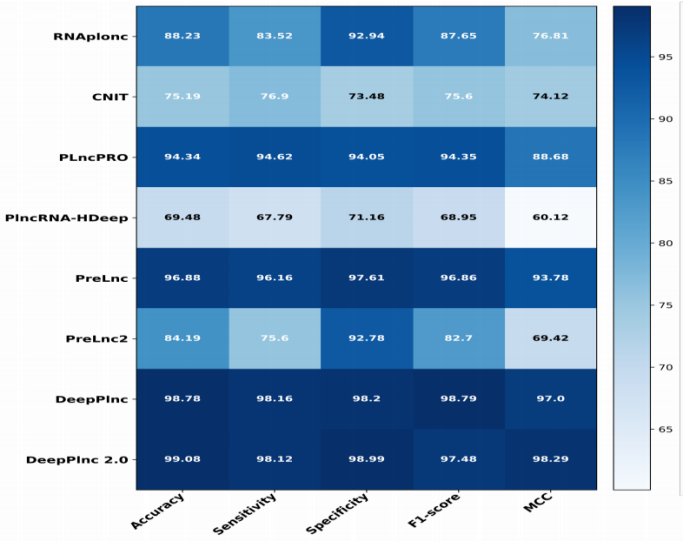


Figure 14. Comparative benchmarking on Dataset "A" test datasets.

filtered sequences, DeepPInc V2.0 identified 29,061 as lncRNAs and 5,536 as non-lncRNAs. Approximately 84% of the identified lncRNA transcripts were correctly classified with a confidence probability of ≥ 0.50. This demonstrated a significantly lower false characterization rate compared to other tools.

The study successfully developed DeepPInc V2.0, a significant advancement in the computational

Algorithm 1: Pseudo-code of DeepPInc V2.0

Data: RNA sequences (FASTA), RNA secondary structure (Dot-Bracket)
Result: Classification label: lncRNA / non-lncRNA
Step 1: Preprocessing;
for *each RNA sequence* **do**
  Encode nucleotides as one-hot vectors;
  Predict secondary structure using RNAfold;
  Convert dot-bracket notation into one-hot encoding;
end
Step 2: Model Architecture;
Sequence branch: Input → DenseNet-121 → Feature vector;
Structure branch: Input → CNN (Conv + Pool + Dense) → Feature vector;
Fusion: Concatenate features (sequence + structure) → Fully connected layers → Softmax classifier;
Step 3: Training;
Use Dataset A (plant lncRNA/non-lncRNA);
Optimize hyperparameters via Bayesian optimization;
Loss function: Binary Cross-Entropy;
Optimizer: Adam;
Step 4: Evaluation;
Compute Accuracy, Precision, Recall, F1-score, MCC, AUC;
Compare with baseline tools;
Step 5: Application (Case Study: Wheat Transcriptome);
Input: Filtered transcripts from RNA-seq pipeline;
Predict lncRNA/non-lncRNA labels;
Identify differentially expressed lncRNAs;

identification of plant lncRNAs. By leveraging a DenseNet architecture and incorporating both sequence and secondary structure information, the model provides a powerful and reliable resource for the scientific community. The successful application to winter wheat transcriptomes underscores the model's potential for genome annotation and the study of developmental processes in plants, offering a foundation for future advancements in plant lncRNA research and functional genomics.

When applied to the wheat transcriptome, DeepPInc V2.0 classified 29,061 out of 34,597 transcripts as lncRNAs, with ~84% predictions assigned a probability ≥ 0.50, reducing false characterizations

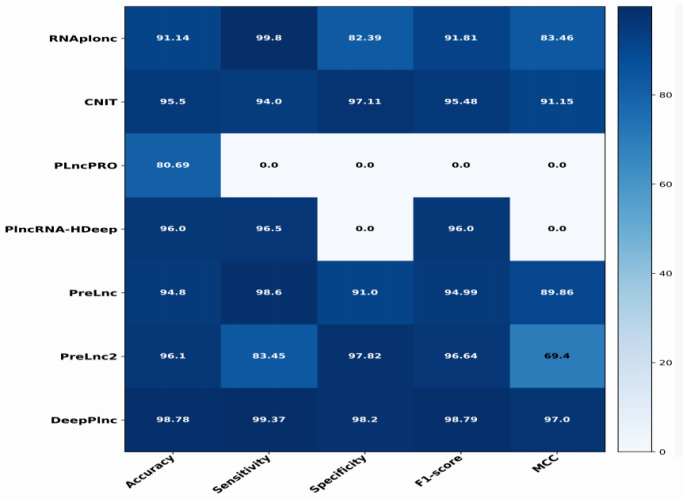


Figure 15. Comparative tools benchmarking on their respective dataset.

compared to existing tools. Importantly, several identified lncRNAs were differentially expressed under heat stress, highlighting their potential regulatory roles in stress adaptation.

Overall, these results demonstrate that DeepPInc V2.0 is more accurate, resilient to incomplete data, and biologically informative, providing a valuable resource for plant transcriptomics and functional genomics.

4 Conclusion and Key Contributions

lncRNAs have become crucial elements of eukaryotic transcriptomes and regulatory systems. Increasingly, lncRNAs are being discovered to play key roles in plant growth, development, and maintenance. Although several tools have been developed to identify lncRNAs in plants, their inconsistency and performance shortcomings highlight the need for improved methods. In this study, we developed DeepPInc V2.0 which represents a significant advancement in the identification of plants through deep learning techniques. Leveraging the DenseNet architecture, the model was trained and validated on Dataset "A," which includes experimentally validated lncRNA and non-lncRNA sequences from comprehensive databases such as Ensembl Plants, PncStress, and PIncDB V2.0. The results demonstrated that DeepPInc V2.0 outperformed seven existing state-of-the-art tools, exhibiting superior accuracy and MCC. The model's robustness was evident as it maintained high classification performance even when applied to truncated sequences, highlighting its versatility and precision.

The application of DeepPInc V2.0 in real-world biological research, particularly in studying transcriptomic changes during the vernalization process in winter wheat (Triticum aestivum), showcased its practical utility. The model effectively identified DEGs at various time points, providing valuable insights into the gene expression dynamics involved in plant adaptation to prolonged cold exposure. This study underscores the potential of DeepPInc V2.0 not only as an advanced tool for lncRNA identification but also as a significant contributor to the broader understanding of plant genomics and developmental processes. The successful deployment of DeepPInc V2.0 paves the way for future advancements in plant lncRNA research and functional genomics, offering a powerful resource for the scientific community.

In this work, we presented DeepPInc V2.0, a deep learning framework designed to address key challenges in plant lncRNA identification. By fusing sequence and structural information through a bi-modal encoding strategy and employing a DenseNet-CNN hybrid architecture, the tool delivers accurate and robust predictions, even when transcripts are incomplete—a frequent scenario in de novo assemblies.

Our benchmarking demonstrated that DeepPInc V2.0 outperforms seven state-of-the-art predictors across multiple plant datasets, consistently achieving higher accuracy, F1-scores, and MCC values. To showcase its real-world applicability, we applied the model to the wheat transcriptome under heat stress, identifying over 27,000 candidate lncRNAs and pinpointing stress-responsive differentially expressed ones, underscoring its biological relevance.

Key Contributions of DeepPInc V2.0

- Bi-modal encoding with biological depth – Integrates sequence and secondary structure to capture richer biological signals.
- Smarter deep learning design – DenseNet-CNN hybrid efficiently learns complex features for superior performance.
- Dependable on incomplete data – Maintains accuracy even on truncated transcripts common in plant assemblies.
- Proven benchmarking superiority – Outperforms seven existing tools in accuracy, F1, and MCC.
- Validated on real data – Revealed $> 27,000$ wheat lncRNAs, including heat stress–responsive ones.

Taken together, these advances make DeepPInc V2.0 a powerful and practical tool for the plant research community. Beyond lncRNA identification, its robustness and design principles pave the way for applications in functional genomics, stress biology, and crop improvement, where understanding noncoding regulation is critical.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Santosh, B., Varshney, A., & Yadava, P. K. (2015). Non-coding RNAs: biological functions and applications. *Cell biochemistry and function, 33*(1), 14-22. [Crossref]

[2] Gallart, A. P., Pulido, A. H., de Lagrán, I. A. M., Sanseverino, W., & Cigliano, R. A. (2015). GREENC: a Wiki-based database of plant lncRNAs. *Nucleic acids research, 44*(Database issue), D1161. [Crossref]

[3] Bak, R. O., & Mikkelsen, J. G. (2014). miRNA sponges: soaking up miRNAs for regulation of gene expression. *Wiley interdisciplinary reviews: RNA, 5*(3), 317-333. [Crossref]

[4] Baek, J., Lee, B., Kwon, S., & Yoon, S. (2018). LncRNAnet: long non-coding RNA identification using deep learning. *Bioinformatics, 34*(22), 3889-3897. [Crossref]

[5] Wang, Y., Fan, X., Lin, F., Chen, X., Mou, S., Wu, S., & Kang, Z. (2022). Long Non-Coding RNAs: New Players in Plants. *International Journal of Molecular Sciences, 23*(14), 9218. [Crossref]

[6] Wierzbicki, A. T., Blevins, T., & Swiezewski, S. (2021). Long noncoding RNAs in plants. *Annual Review of Plant Biology, 72*(1), 245-271. [Crossref]

[7] Saha, C., Saha, S., & Bhattacharyya, N. P. (2025). LncRNAOmics: A Comprehensive Review of Long Non-Coding RNAs in Plants. *Genes, 16*(7), 765. [Crossref]

[8] Chekanova, J. A. (2015). Long non-coding RNAs and their functions in plants. *Current opinion in plant biology, 27*, 207-216. [Crossref]

[9] Beermann, J., Piccoli, M. T., Viereck, J., & Thum, T. (2016). Non-coding RNAs in development and disease: background, mechanisms, and therapeutic approaches. *Physiological reviews, 96*(4), 1297-1325. [Crossref]

[10] Jampala, P., Garhewal, A., & Lodha, M. (2021). Functions of long non-coding RNA in Arabidopsis thaliana. *Plant Signaling & Behavior, 16*(9), 1925440. [Crossref]

[11] Wang, Y., Wang, Z., Du, Q., Wang, K., Zou, C., & Li, W. X. (2023). The long non-coding RNA PILNCR2 increases low phosphate tolerance in maize by interfering with miRNA399-guided cleavage of ZmPHT1s. *Molecular Plant, 16*(7), 1146-1159. [Crossref]

[12] Lorenz, R., Bernhart, S. H., Höner zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P. F., & Hofacker, I. L. (2011). ViennaRNA Package 2.0. *Algorithms for Molecular Biology, 6*(1), 26. [Crossref]

[13] Quitadadmo, A., Johnson, J., & Shi, X. (2017, August). Bayesian hyperparameter optimization for machine learning based eQTL analysis. In *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics* (pp. 98-106). [Crossref]

[14] Franke, J. K., Runge, F., Köksal, R., Backofen, R., & Hutter, F. (2024). Rnaformer: A simple yet effective deep learning model for rna secondary structure prediction. *BioRxiv*, 2024-02.

[15] Miller, J. R., Yi, W., & Adjeroh, D. A. (2024). Evaluation of machine learning models that predict lncRNA subcellular localization. *NAR Genomics and Bioinformatics, 6*(3), lqae125. [Crossref]

[16] Hesketh, A. R. (2019). RNA sequencing best practices: experimental protocol and data analysis. In *Yeast Systems Biology: Methods and Protocols* (pp. 113-129). New York, NY: Springer New York. [Crossref]

[17] Szcześniak, M. W., Rosikiewicz, W., & Makałowska, I. (2016). CANTATAdb: a collection of plant long non-coding RNAs. *Plant and Cell Physiology, 57*(1), e8-e8. [Crossref]

[18] Yang, C., Yang, L., Zhou, M., Xie, H., Zhang, C., Wang, M. D., & Zhu, H. (2018). LncADeep: an ab initio lncRNA identification and functional annotation tool based on deep learning. *Bioinformatics, 34*(22), 3825-3834. [Crossref]