RESEARCH ARTICLE

# DeepFI: A hybrid network based on convolutional neural network for predicting plant functional introns

Yifei Wang[1], Jiming Hu[2], Jinshan Li[1], Shutian He[1], Xiaoyong Sun[1,*]

[1]College of Information Science and Engineering, Shandong Agricultural University, Tai' an, Shandong, China. [2]China Mobile Communications Group Jiangsu Company Limited Taizhou Branch, Taizhou, Jiangsu, China.

**Received:** June 27, 2025; **accepted:** October 30, 2025.

**Introns are non-coding sequences within the coding region of genes that were originally believed to lack biological function after being spliced out. However, recent studies have revealed that introns in fact play critical regulatory roles in eukaryotes. Because of these findings, accurate prediction of functional introns has become a key challenge in genomic research. Deep learning has been widely applied in gene prediction. However, development of cross-species applicable models for the prediction of functional introns in plants remains difficult. This research proposed a hybrid deep neural network model (DeepFI) based on a convolutional neural network (CNN) for the precise prediction of functional introns in plants by incorporating dual path encoding, multi-scale feature extraction, and an attention fusion mechanism along with a dynamic adversarial training method (EnhancedFGM) to improve prediction accuracy. In addition, a cross-species mixed dataset was constructed, containing 76,945 pairs of positive and negative samples from monocot (Arabidopsis, rice, maize) and dicot plants (soybean, cotton, rapeseed). The ability of DeepFI to analyze this cross-species mixed dataset was compared with three benchmark models including BiLSTM, PDLLM, and DanQ. The results showed that DeepFI significantly outperformed other benchmark models across all tests. Varying degrees of improvement were observed for single-species datasets, particularly for the cotton dataset where accuracy increased by 4.63%. For the cross-species mixed dataset, DeepFI achieved an accuracy of 94.58%, surpassing the best-performing benchmark model by 1.44%. Moreover, DeepFI demonstrated excellent zero-shot prediction ability with accuracies of 87.60% and 82.63% for prediction of functional introns in the distantly related species grape and sorghum, respectively. This work not only filled the gap in cross-species prediction tools for plant functional introns, but also provided a reliable computational method for functional annotation of plant genomes, which had important implications for evolutionary biology, genetic crop improvements, and the exploration of non-coding DNA functions.**

**Keywords:** deep learning; functional introns; prediction; plants; cross-species.

[*]**Corresponding author:** Xiaoyong Sun, College of Information Science and Engineering, Shandong Agricultural University, Tai' an, Shandong 271018, China. Email: sunx1@sdau.edu.cn.

## Introduction

Introns are non-coding regions between exons in gene sequences. During processing of precursor mRNA into mature mRNA, introns are removed through splicing, while exons are joined together to form transcripts that are eventually translated into proteins [1]. Historically, introns were originally considered to lack genetic function and were not believed to participate in essential

cellular processes such as protein synthesis. However, advances in molecular biology have revealed that introns are not “junk DNA” but possess specific biological functions [2-4]. In 2019, two breakthrough studies confirmed a key role for introns in stress responses. Parenteau *et al*. knocked out over 200 introns in *Saccharomyces cerevisiae* and found that their presence was critical for yeast survival under nutrient deficiency conditions [5], while Morgan *et al*. confirmed that 34 introns in yeast cells regulated cell growth rates by modulating the mTOR signaling pathway, thereby enhancing yeast adaptability and survival fitness [6]. More recent studies extend these results and show that introns also play important roles in plants, which demonstrated that proper splicing of group II introns was essential for mitochondrial gene expression and plant development [7, 8]. In *Arabidopsis thaliana*, introns have been shown to contain specific functional elements that are indispensable for the regulation of various fundamental biological processes [9, 10].

Recently, deep learning has become an important tool in the field of bioinformatics, where significant progress has been achieved in gene prediction tasks. Cheng *et al*. proposed the MiRTDL algorithm for miRNA target prediction using convolutional neural network (CNN) and achieved a sensitivity, specificity, and accuracy of 88.43%, 96.44%, and 89.98%, respectively [11]. The iDeepE model successfully predicted RNA-binding proteins by combining local and global RNA sequence features using CNN and achieved the best average area under the curve (AUC) of 93.10% [12]. Wang *et al*. constructed the DeepCirCode model for circular RNA prediction based on a CNN algorithm, which achieved 85.24% accuracy on the test set [13]. Zhang *et al*. combined a stacked ensemble classifier with a transformer-based bi-directional encoder in BERT-m7G model to predict N7-methylguanosine sites, resulting in 95.48% accuracy and a Matthews correlation coefficient (MCC) of 91% [14]. Ritu *et al*. employed a bi-CNN algorithm in their DeepPlnc model, which enabled high precision identification of plant lncRNAs with 98% accuracy [15]. Using RNA sequence features extracted *via* a transformer model, Xiang *et al*. were able to predict N6-methyladenosine modification sites with 84.43% accuracy and an MCC of 83.72% [16]. In addition, deep learning models have also been generalized for cross-species analyses and maintain good performance in the absence of species-specific data [17, 18]. Despite these advances, applications of deep learning for functional intron prediction remain limited. Previous study developed an intron-capture RNA sequencing (RNA-seq) technique that successfully identified multiple introns with functional elements in *Arabidopsis thaliana* [19], and based on that, a functional intron database, PlantIntronDB, was established, which included five plant species [20]. However, current methods for the identification and extraction of functional introns have several limitations, which include that the experimental procedures involved are complex and require multiple technical steps, while the data processing phase is time-consuming. These technical limitations present significant challenges for extending the methodology to additional species. More critically, there is currently a lack of computational tools that can efficiently and accurately predict functional introns with cross-species transferability and generalization capabilities. This significant methodological gap severely hinders large-scale and systematic identification of functional introns across different plant species and limits research applications in genome annotation and functional exploration.

This research aimed to address this gap in functional intron prediction capabilities in plants by developing an accurate, efficient, and cross-species applicable prediction model. A hybrid network model, DeepFI, based on CNN was proposed for predicting plant functional introns, which integrated dual path encoding, multi-scale feature extraction, and an attention fusion mechanism along with an enhanced fast gradient method (EnhancedFGM) dynamic adversarial training technique to improve model performance. To support model training and

validation, large-scale RNA-seq data from multiple plant species were integrated and analyzed, creating a cross-species mixed dataset. The proposed model was compared with several benchmark models under unified experimental conditions. This research not only filled the methodological gap in plant functional intron prediction but also provided technical support for the expansion of functional intron databases. Furthermore, it offered a reliable computational tool for large-scale, cross-species identification of functional non-coding elements, which was important for advancing research in plant genomics and evolutionary biology and could be used to guide potential improvements in crop genetics.

## Materials and methods

### Data acquisition and dataset construction

RNA-seq data for eight plant species were downloaded from the NCBI SRA database (https://www.ncbi.nlm.nih.gov/sra) including *Arabidopsis thaliana* (Arabidopsis), *Gossypium raimondii* (cotton), *Zea mays* (maize), *Brassica napus* (rapeseed), *Oryza sativa Japonica* (rice), *Glycine max* (soybean), *Vitis vinifera* (grape), and *Sorghum bicolor* (sorghum). A total of 622 Arabidopsis samples, 59 cotton samples, 185 maize samples, 210 rapeseed samples, 236 rice samples, 230 soybean samples, 433 grape samples, and 649 sorghum samples were obtained. The SRA Toolkit v3.0.2 (https://github.com/ncbi/sra-tools) was used to convert the downloaded SRA files into FASTQ format to obtain raw RNA-seq reads. Subsequently, HISAT2 v2.2.1 (https://daehwankimlab.github.io/hisat2/) was applied to align the reads to their corresponding reference genomes including Arabidopsis (TAIR10), cotton (Graimondii2_0_v6), maize (Zm-B73-REFERENCE-NAM-5.0), rapeseed (AST_PRJEB5043_V1), rice (IRGSP-1.0), soybean (Glycine_max_v2.1), grape (PN40024.v4), and sorghum (Sorghum_bicolor_NCBIv3). All reference genomes were obtained from the Ensembl Plants database (https://plants.ensembl.org/). To ensure data quality, samples with alignment rates below 80% or those that did not produce valid sequence data were discarded. Samtools v1.9 (https://github.com/samtools/samtools) was then used to sort and index the generated BAM files for further analysis. The gene annotation files in GTF format were processed using R v4.3.1 (https://www.r-project.org/) and Bioconductor package v3.18 (https://www.bioconductor.org/). Gene and exon annotation information were extracted, and all exon regions for the same gene were merged. Then, exon intervals were subtracted from the gene regions to precisely define the intronic regions of each gene. The extracted intron information included the genomic start and end positions, chromosome IDs, strand direction, gene IDs, and other relevant information. To ensure the reliability and accuracy of the data, generated BED files were visualized using Integrative Genomics Viewer (IGV) v2.19.2 (https://igv.org/). To ensure the quality of the identified functional introns, sequencing reads for each sample were filtered according to the criteria described in previous study [20], which included that the reads must be located entirely within introns, strand-specific, and not be spliced. The number of reads within each intron was counted. To eliminate artifacts, all reads that overlapped with exons and introns were removed. To build a high-quality functional intron dataset, intron sequences detected in at least six biological repeated samples (reads number ≥ 5) were defined as functional introns (positive samples), while intron sequences undetected in all samples (reads number = 0) were defined as non-functional introns (negative samples). Chromosome IDs, start and end positions, and strand orientation were extracted from CSV files to generate BED files, which were then visualized using IGV. Combined with whole genome sequence files, the corresponding gene sequences were extracted using Bedtools v2.30.0 (https://github.com/arq5x/bedtools2) to obtain all positive and negative samples. To ensure the accuracy and robustness of the model, and to avoid bias associated with imbalanced data distribution, the study applied a random under-

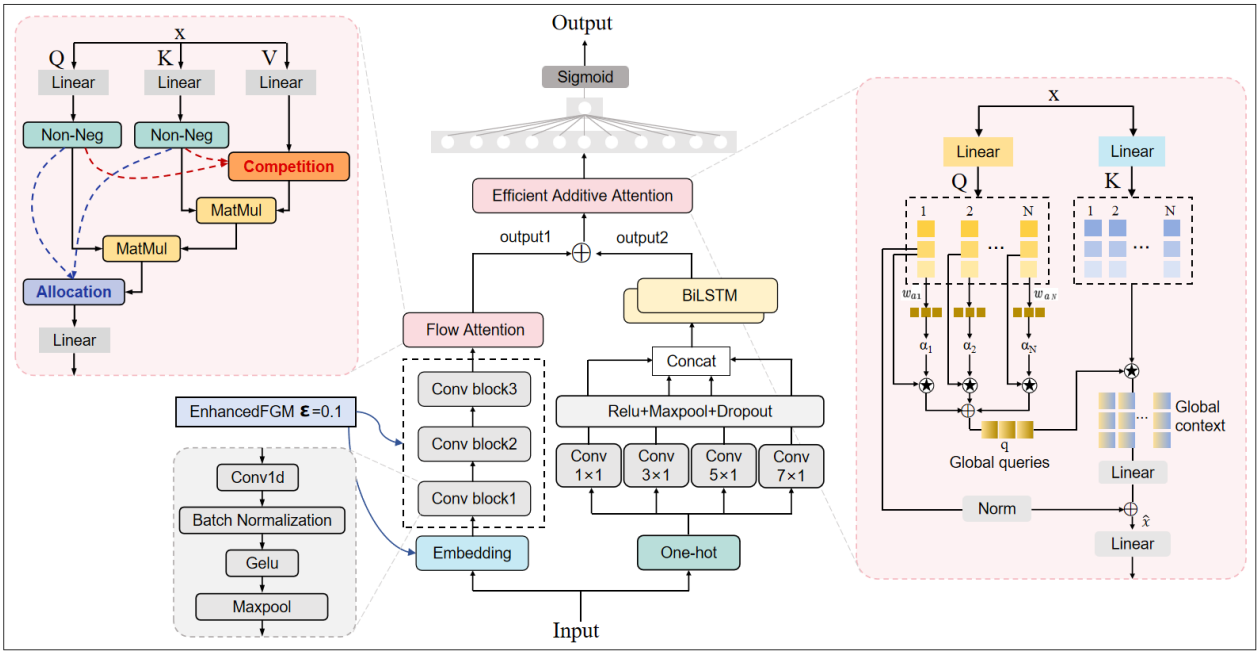


**Figure 1.** DeepFI model architecture.

sampling method to balance the ratio of positive to negative samples at 1:1. In the dataset division, a stratified sampling strategy was used to maintain a consistent ratio of positive and negative samples across the training, validation, and test sets. The data was split into an approximately 8:1:1 ratio. For the cross-species prediction experiments, grape and sorghum were set as independent test species, while the remaining six species (Arabidopsis, cotton, maize, rapeseed, rice, and soybean) were used for model training and validation. For each of these six species, balanced single-species datasets were constructed, and these were further combined into a cross-species mixed dataset containing 76,945 positive and 76,945 negative samples.

**Data preprocessing**

Before training the model, all nucleotide sequences were processed through a standardized preprocessing procedure to ensure consistency in input dimensions. All sequences were duplicated to eliminate the influence of redundant data. The input length of all sequences was standardized to a fixed length of 1,000 base pairs (bp). For sequences shorter than 1,000 bp, gaps were filled by random insertion of "N" bases at insertion positions that were randomly selected from position 1 to L, where L was the original sequence length. This strategy could effectively simulate the presence of unknown bases that might occur in real sequencing and avoid the deviation that might be introduced by fixed position padding. For sequences longer than 1,000 bp, the first 1,000 bp were retained for further analysis. Sequences that were exactly 1,000 bp in length were directly used as input for the model. This preprocessing workflow ensured that all input sequences adhered to a uniform length requirement, while maintaining biological relevance.

**Model architecture**

The proposed DeepFI model incorporated several core innovations including dual path encoding, multi-scale feature extraction, and attention fusion architecture (Figure 1). In the coding layer, the original sequence was processed in parallel *via* two pathways of word embedding encoding and one-hot encoding. Word embedding encoding could effectively capture the high-level

semantic features of intron sequences by modeling them in continuous vector space, while one-hot encoding accurately retained base composition information from the original sequence, enabling precise identification of key functional sites. The combination of these two pathways allowed for full characterization of the multi-dimensional characteristics of functional introns. For the word embedding encoding pathway, three convolutional blocks were designed for hierarchical feature extraction. An EnhancedFGM dynamic adversarial training algorithm was incorporated into both the word embedding and convolutional layers, which enabled simulation of natural sequence variation by directionally disturbing the gradient of the embedding layer and convolution layer and significantly improved the adaptability and transferability of the model to cross-species introns. Each convolution block consisted of a one-dimensional convolutional layer, a batch normalization layer, a Gaussian Error Linear Unit (GELU) activation function, and a max pooling layer. The number of channels was progressively set to 16, 32, and 64. Subsequently, the model employed a flow-attention mechanism, which effectively captured long-range dependencies in the sequence by dynamically adjusting feature weights, where output was recorded as output1. The one-hot encoding pathway employed a multi-scale feature extraction strategy, which utilized parallel one-dimensional convolutional layers with kernel sizes of 1, 3, 5, and 7 to capture sequence features at different scales including single-nucleotide features, short motifs, medium-range regulatory elements, and long-range functional domains. Each convolutional branch applied a Rectified Linear Unit (ReLU) activation function and a max pooling layer for nonlinear transformation and feature dimension reduction, and with Dropout being introduced to prevent overfitting. After multi-scale features were fused along the feature dimension by a Concatenation (CONCAT) function, the model was passed through a two-layer bi-directional long short-term memory (BiLSTM) network to further extract the time series features, and the final output was recorded as output2. Features from output1 and output2 were concatenated and fed into an Efficient Additional Attention (EAA) mechanism. This EAA mechanism significantly improved computational efficiency and reduced memory consumption, while maintaining high prediction accuracy because it simplified the calculation process and adopted additive attention. The output layer employed a Sigmoid activation function to map the final prediction value to the [0, 1] interval, representing the probability of functional intron prediction. In this study, the binary cross-entropy loss function was used as the loss function for the DeepFI model.

**EnhancedFGM dynamic adversarial training algorithm**

Although FGM usually introduced perturbations to all model parameters, it lacked the ability to customize perturbations for specific layers. The EnhancedFGM was specifically designed for targeted adversarial training, which enhanced the robustness of the model by incorporating two key improvements including the application of layer-selective perturbations by adding gradient-based noise only to the word embedding and convolutional layers to avoid interference from irrelevant parameters and the introduction of a gradient safety mechanism to automatically filter outliers and normalize the perturbation amplitude. During training, EnhancedFGM first backed up the original parameters, then injected controllable perturbations to generate adversarial samples (Figure 2). After computing adversarial loss, it immediately restored the parameters to ensure training stability. The perturbation was calculated as follows.

$$r_{\text{EnhancedFGM}} = \epsilon \cdot \frac{\nabla_\theta \mathcal{L}}{\|\nabla_\theta \mathcal{L}\|_2 + \delta} \cdot \mathbb{I}(\theta \in \{\theta_{emb}, \theta_{conv}\}) \quad (1)$$

where $\epsilon$ was the perturbation coefficient, which controlled the perturbation intensity. $\nabla_\theta \mathcal{L}$ was the gradient of loss function for parameter $\theta$. $\delta$ was a numerical stability term, set to $1e^{-8}$. $\mathbb{I}$ was an indicator function which ensured that perturbations were applied only to the parameters of the word embedding layer and

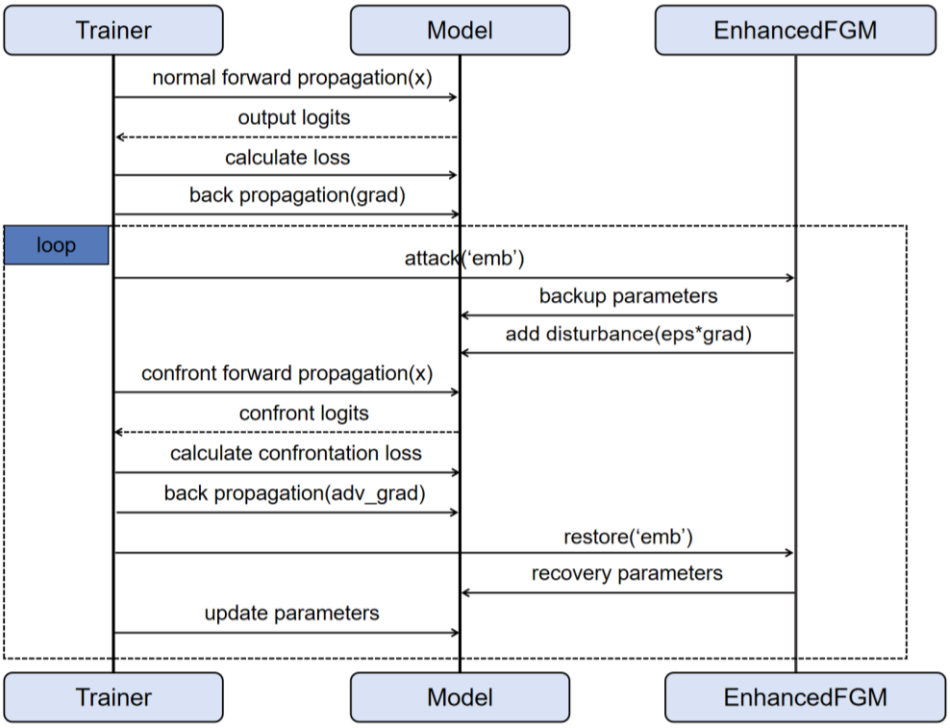


**Figure 2.** Flow chart of the EnhancedFGM dynamic adversarial training process.

convolutional layer.

**Benchmark model selection and training**

Because of the lack of specialized models published in the field of plant functional intron prediction, 18 related models were selected from the following three categories as benchmark references, which included 10 classic deep learning models that were widely used in sequence analysis including one-dimensional convolutional neural network (1D-CNN), two-dimensional convolutional neural network (2D-CNN), recurrent neural network (RNN), bidirectional recurrent neural network (BiRNN), long short-term memory network (LSTM), BiLSTM, gated recurrent unit (GRU), bidirectional gated recurrent unit (BiGRU), Transformer, and BERT; 5 models specifically designed for DNA sequence prediction including DeepSea [20] (http://deepsea.princeton.edu/), DanQ [21] (https://github.com/uci-cbcl/DanQ), TBiNet [22] (https://github.com/dmis-lab/tbinet), DeepATT [23] (https://github.com/cntx-gnnewton/deepatt), and DeepFormer [24] (https://github.com/YZ20211221/DeepFormer); and 3 large-scale models for DNA prediction including DNABERT [25] (https://github.com/Jerryji1993/DNABERT), DNABERT2 [26] (https://github.com/JackxTong/DNABERT2), and PDLLM [27] (https://github.com/zhangtaolab/Plant_DNA_LLMs). This multi-level, wide-coverage model selection strategy was used to ensure that performance evaluation of proposed DeepFI model was both comprehensive and objective. All benchmark models were trained and tested against the cross-species mixed dataset, strictly adhering to an 8:1:1 ratio for the training, validation, and test sets with the same data preprocessing process applied across all models. Each sequence was represented using one-hot encoding with bases A, C, G, and T being encoded as [1, 0, 0, 0], [0, 1, 0, 0], [0, 0, 1, 0], and [0, 0, 0, 1], respectively, while non-standard bases were encoded as [0, 0, 0, 0]. During model training, the batch size was set to 128 and the maximum number of training epochs was set to 50. To optimize the training process and prevent

overfitting, a combination of Early Stopping and ReduceLROnPlateau strategies was employed. The initial learning rate was set to $5e^{-4}$ and was reduced by a factor of 0.5 if the validation accuracy did not improve for three consecutive epochs. If no improvement was observed for five consecutive epochs, the training process was halted early.

**Model training and performance evaluation**

The DeepFI model was trained using the optimized configuration that the batch size was set to 128 and the number of training epochs was set to 50. Parameters were updated using the Adam optimizer with the ReduceLROnPlateau strategy being applied to dynamically adjust the learning rate. The initial learning rate was set to $5e^{-4}$ with a minimum learning rate threshold of $1e^{-6}$ and a weight decay coefficient of $1e^{-5}$. All Dropout layers were set to 0.3 to enhance regularization and prevent overfitting. For the single-species performance evaluation, a five-fold cross-validation strategy was adopted. For each species, introns from a complete chromosome were randomly selected as an independent test set (random seed = 42) to ensure complete independence between test and training data at the chromosome level. Intron data from the remaining chromosomes were trained and verified by five-fold cross-validation with data split into training and validation sets at an 8:2 ratio in each fold. During the cross-validation process, the model was independently trained on the training data of each fold, and model performance was monitored on the corresponding validation set to save the best-performing weights. Afterwards, the saved model weights were evaluated against the independent test set, and performance metrics were recorded. By aggregating evaluation results from all five folds on the test set, the mean and standard deviation for each metric were computed, which provided a comprehensive view of the model's overall performance and stability. This process effectively verified the cross-chromosome generalization capability of the model and avoided deviation caused by a single data division. For evaluation of the model against the cross-species mixed dataset, the data were randomly split into training, validation, and test sets at an 8:1:1 ratio. This mixed dataset was used to train and assess the performance of the model in multi-species integration scenarios.

**Evaluation metrics**

Model evaluation was conducted using six metrics including accuracy (Equation 2), precision (Equation 3), recall (Equation 4), F1 score (Equation 5), Matthews Correlation Coefficient (MCC) (Equation 6), and Area Under the Receiver Operating Characteristic (AUROC) (Equation 7) as follows.

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

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

$$F1-score = \frac{2\times Precision\times Recall}{Precision+Recall} \quad (5)$$

$$MCC = \frac{TP\times TN-FP\times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (6)$$

$$AUROC = \int_0^1 \mathrm{TPR}(\mathrm{FPR})\, d(\mathrm{FPR}) \quad (7)$$

where TP was true positive, representing the number of samples with positive real data and positive prediction results. TN was true negative, representing the number of samples with negative real data and negative prediction results. FP was false positive, representing the number of samples with negative real data and positive prediction results. FN was false negative, representing the number of samples with positive real data and negative prediction results. FPR was false positive rate, representing the proportion of actual negative samples that were incorrectly predicted as positive. TPR was true positive rate, indicating the proportion of actual positive samples that were correctly predicted as positive. Among these evaluation metrics, accuracy, precision, recall, F1 score, AUROC

ranged from [0, 1], while MCC ranged from [-1, 1].

**Ablation study design**

To evaluate the effectiveness of each core component of the DeepFI model, five model variants were designed for ablation experiments including systematic removal of individual modules for comparison. For the first model variant, the dual path encoding module was removed (w/o Dual-path), and only one-hot encoding was used as input, replacing the original dual path encoding strategy. In the second variant, the multi-scale feature extraction structure was removed (w/o Multi-scale), and a single convolutional layer was used instead of the original multi-kernel convolution layers. For the third model variant, the flow-attention mechanism was removed (w/o Flow-Attention) to assess the role of the attention mechanism in feature optimization. For the fourth model variant, the EAA mechanism was omitted (w/o EAA) to analyze the contribution of EAA to feature fusion. In the fifth model variant, the EnhancedFGM dynamic adversarial training algorithm was removed (w/o EnhancedFGM) during training to examine the impact of this algorithm on the performance of the model. All model variants were trained and tested using the same dataset, training configuration, and evaluation metrics, allowing for quantitative analysis of the contribution of each module to the overall performance of DeepFI.

**Cross-species prediction experimental design**

The cross-species generalization ability of the DeepFI model was evaluated using a cross-species zero-shot prediction experiment designed. The DeepFI model was trained on a mixed dataset, and two distantly related species (sorghum and grape) were selected as independent test sets for performance evaluation based on multiple scientific considerations. From an evolutionary development standpoint, grapes are dicotyledonous plants, while sorghum are monocotyledonous plants, and both belong to different evolutionary branches. The grape plants are known for their high genomic and structural heterozygosity, while sorghum has a more compact genome that is typical of Gramineae crops. Thus, the combination of these two species effectively captured a broad spectrum of intron structural diversity across plants. Furthermore, although both crops hold significant agricultural and economic importance, studies on the functionality of their introns remain limited. The experiment was conducted using standardized data preprocessing and a feature encoding method to ensure consistency in the input feature space. To evaluate the zero-shot learning ability of the model, independent forward propagation calculation of grape and sorghum test sets were carried out during the model verification stage by freezing the model parameters after training convergence. Performance evaluation also involved use of six metrics to comprehensively investigate the transfer learning effect of the model among distant species.

**Input sequence length experiment design**

To investigate the impact of input sequence length on model performance, a gradient comparison experiment was designed based on the upper quartile distribution of intron sequence lengths from six plant species. The upper quartile distribution for these species were Arabidopsis at 291 bp, cotton at 836 bp, maize at 723 bp, rapeseed at 454 bp, rice at 575 bp, and soybean at 723 bp. Based on this distribution, intron fragments of lengths 300 bp, 450 bp, 600 bp, 750 bp, 850 bp, and 1,000 bp were selected as inputs to the model. In the experiment, the DeepFI model architecture and training parameters remained unchanged to ensure that performance differences were solely attributable to variations in sequence length.

## Results

**Composition and length distribution of the intron dataset**

In the present study, a large-scale intron dataset was extracted from eight plant species with the

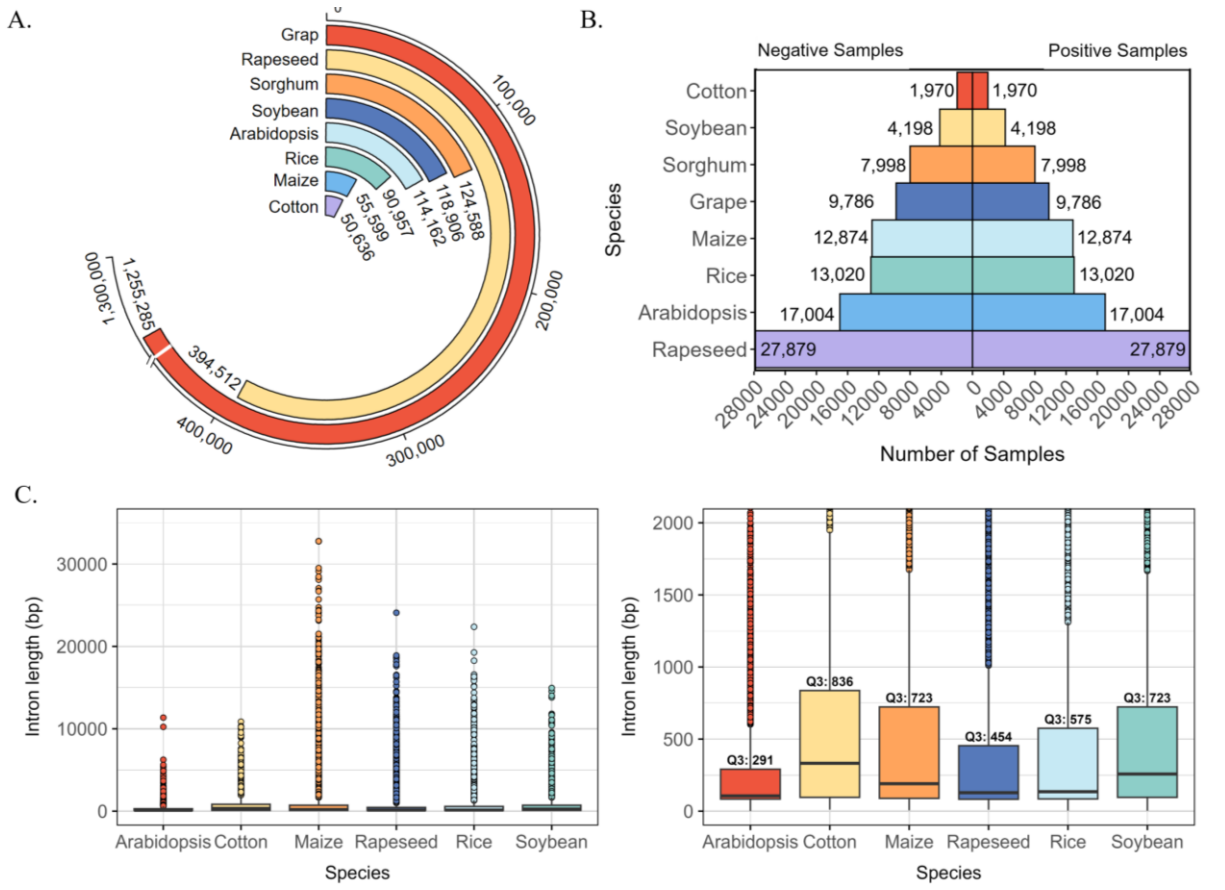


**Figure 3.** Composition and length distribution of intron dataset. **A.** Distribution of intron number of each species. **B.** Number of positive and negative samples of each species dataset after balance treatment. **C.** Distribution of intron sequence length of six species.

distribution of intron quantities shown in Figure 3A. Based on the experimentally selected functional introns (positive samples) and non-functional introns (negative samples), a positive-negative sample dataset was constructed. Analysis revealed that the original dataset exhibited a significant class imbalance between positive and negative samples, which could hinder model training and evaluation. To address this issue, a random under-sampling technique was applied to balance samples. The resulting balanced distribution of positive and negative samples was shown in Figure 3B. To determine the appropriate model input length, statistical analysis of the intron sequence length distribution for the six species used in training and validation was conducted. All intron sequences were less than 35,000 bp in length with over 86% of the sequences being shorter than 1,000 bp (Figure 3C). The result suggested that setting the input length to 1,000 bp was sufficient to cover the core functional regions of most introns. Therefore, a uniform length of 1,000 bp was chosen as the standard input length for subsequent model training and evaluation.

**Performance comparison of benchmark models**

The evaluation results of classical deep learning models showed that BiLSTM demonstrated the best overall performance with respect to functional intron prediction. It ranked first in several key metrics including accuracy, F1 score, and MCC. Although RNN achieved slightly higher precision and BERT showed better recall than BiLSTM, both models performed significantly worse in other metrics (Table 1). The superior performance of BiLSTM was mainly attributed to its bidirectional architecture, which effectively

**Table 1.** Performance evaluation of ten classical deep learning models for predicting functional introns.

| Model | Accuracy | Precision | Recall | F1-score | MCC | AUROC |
|---|---|---|---|---|---|---|
| 1D-CNN | 0.9232 | 0.9392 | 0.9052 | 0.9219 | 0.8474 | 0.9722 |
| 2D-CNN | 0.9215 | 0.9405 | 0.9008 | 0.9202 | 0.8448 | 0.9718 |
| RNN | 0.8810 | **0.9661** | 0.7892 | 0.8687 | 0.7750 | 0.8336 |
| BiRNN | 0.6761 | 0.7080 | 0.5968 | 0.6477 | 0.3564 | 0.7171 |
| LSTM | 0.9290 | 0.9469 | 0.9087 | 0.9274 | 0.8587 | 0.9708 |
| **BiLSTM** | **0.9303** | 0.9423 | 0.9163 | **0.9291** | **0.8609** | **0.9758** |
| GRU | 0.9264 | 0.9543 | 0.8953 | 0.9239 | 0.8544 | 0.9732 |
| BiGRU | 0.9293 | 0.9328 | 0.9249 | 0.9288 | 0.8586 | 0.9748 |
| Transformer | 0.9248 | 0.9455 | 0.9012 | 0.9228 | 0.8505 | 0.9711 |
| BERT | 0.9201 | 0.9207 | **0.9218** | 0.9212 | 0.8401 | 0.9595 |

**Table 2.** Performance evaluation of five DNA sequence prediction SOTA models for predicting functional introns.

| Model | Accuracy | Precision | Recall | F1-score | MCC | AUROC |
|---|---|---|---|---|---|---|
| DeepSea | 0.9278 | 0.9357 | 0.9209 | 0.9282 | 0.8522 | 0.9714 |
| **DanQ** | **0.9314** | **0.9395** | **0.9306** | **0.9350** | **0.8659** | 0.9738 |
| TBiNet | 0.9307 | 0.9366 | 0.9234 | 0.9300 | 0.8556 | 0.9732 |
| DeepATT | 0.9299 | 0.9372 | 0.9278 | 0.9325 | 0.8592 | **0.9742** |
| DeepFormer | 0.9292 | 0.9343 | 0.9232 | 0.9287 | 0.8573 | 0.9737 |

**Table 3.** Performance evaluation of three large-scale models in the field of genomics for predicting functional introns.

| Model | Accuracy | Precision | Recall | F1-score | MCC | AUROC |
|---|---|---|---|---|---|---|
| DNABERT | 0.9267 | 0.9365 | 0.9143 | 0.9253 | 0.8536 | 0.9721 |
| DNABERT2 | 0.9297 | **0.9371** | 0.9236 | 0.9303 | 0.8596 | 0.9733 |
| **PDLLM** | **0.9305** | 0.9368 | **0.9300** | **0.9334** | **0.8671** | **0.9782** |

captured long-range dependencies in DNA sequences and thereby enhanced prediction accuracy. Therefore, BiLSTM was selected as the benchmark model among the tested classical deep learning approaches. The evaluation results of five specialized DNA sequence prediction models showed that DanQ exhibited the strongest functional intron prediction performance, ranking first in five critical metrics with an AUROC slightly lower than that of DeepATT (Table 2). The outstanding performance of DanQ was mainly due to its hybrid architecture that integrated CNN with BiLSTM, enabling the model to capture both local features and long-range dependencies in DNA sequences. Based on its comprehensive performance, DanQ was chosen as the benchmark model within this category for subsequent comparison with the proposed DeepFI model. Three large-scale models for DNA sequence prediction were assessed (Table 3). The comparison results showed that PDLLM outperformed all other tested models across multiple metrics. Specifically, PDLLM achieved higher accuracy, F1 score, and AUROC than that of DNABERT and DNABERT2, and was particularly strong in recall. This improvement might be attributed to the fact that it had been specifically optimized for plant genomic data, which allowed it to better capture sequence patterns related to plant-specific functions. Consequently, PDLLM was selected as the benchmark model among large-scale approaches.

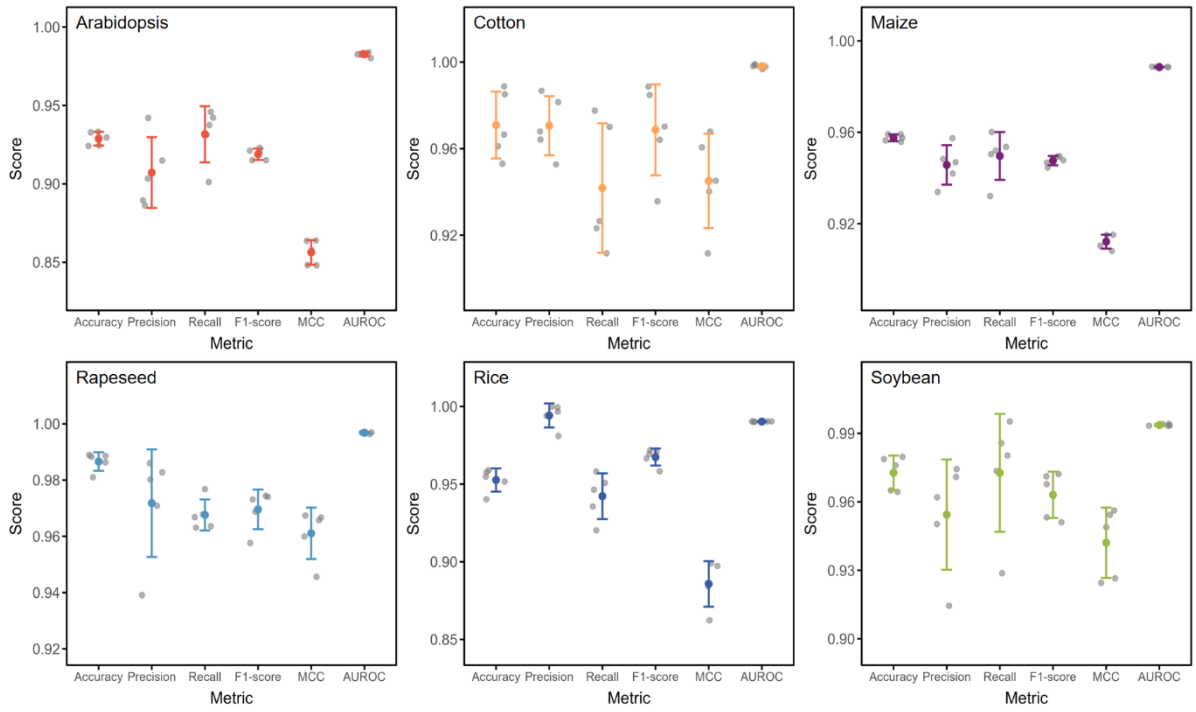


**Figure 4.** Evaluation results of each species on the test set after five-fold cross-validation.

**DeepFI performance**

The DeepFI model achieved outstanding performance in functional intron prediction across the six tested plant species with consistently low standard deviations across all evaluation metrics, indicating excellent stability under five-fold cross-validation. Overall, the average accuracy exceeded 92% and AUROC values were above 0.98 for all species. Among them, the best overall performance was achieved for rapeseed with an accuracy of 0.9866 and an AUROC of 0.9968, while comparatively lower performance was obtained for Arabidopsis with an accuracy of 0.9288 and an AUROC of 0.9826. Excellent results were also obtained for the other species tested. With respect to the other evaluation metrics, the average precision, recall, and F1 score for all species were above 0.95, while the average MCC reached 0.92, demonstrating the ability of the model to effectively balance between recognition of positive and negative samples. Notably, for the rapeseed dataset, DeepFI achieved the highest values across all metrics with precision of 0.9718, recall of 0.9676, F1 score of 0.9696, and MCC of 0.9611, demonstrating its exceptional classification performance and robustness for this species (Figure 4). For the cross-species mixed dataset, the DeepFI model achieved excellent results including an accuracy of 0.9458, a precision of 0.9549, a recall of 0.9363, a F1 score of 0.9455, a MCC of 0.8918, and AUROC of 0.9836 on the test set. The model was highly effective at discriminating between positive and negative samples (Figure 5A). The precision-recall (PR) curve further confirmed that the model maintained high precision while achieving high recall with an average precision (AP) of 0.9859 (Figure 5B). The proposed DeepFI displayed excellent performance and robustness for the prediction of functional introns in plants. The model consistently maintained a stable, high-level performance during both species-specific five-fold cross-validation and independent testing against a cross-species mixed dataset.

**Model comparison**

DeepFI was comprehensively compared with three benchmark models, BiLSTM, PDLLM, and DanQ, against both single- and mixed-species

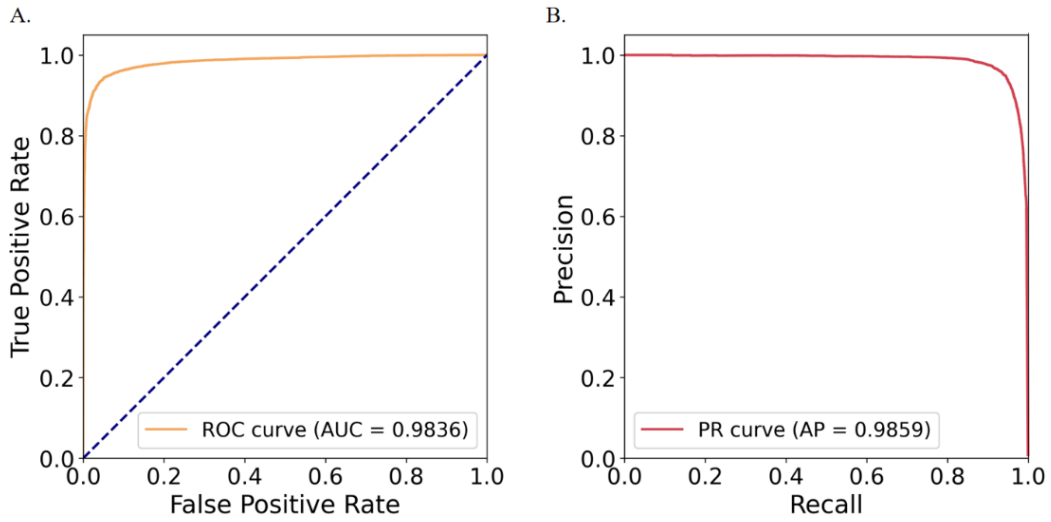


**Figure 5.** ROC curve (**A**) and PR curve (**B**) of DeepFI model on cross-species mixed dataset.

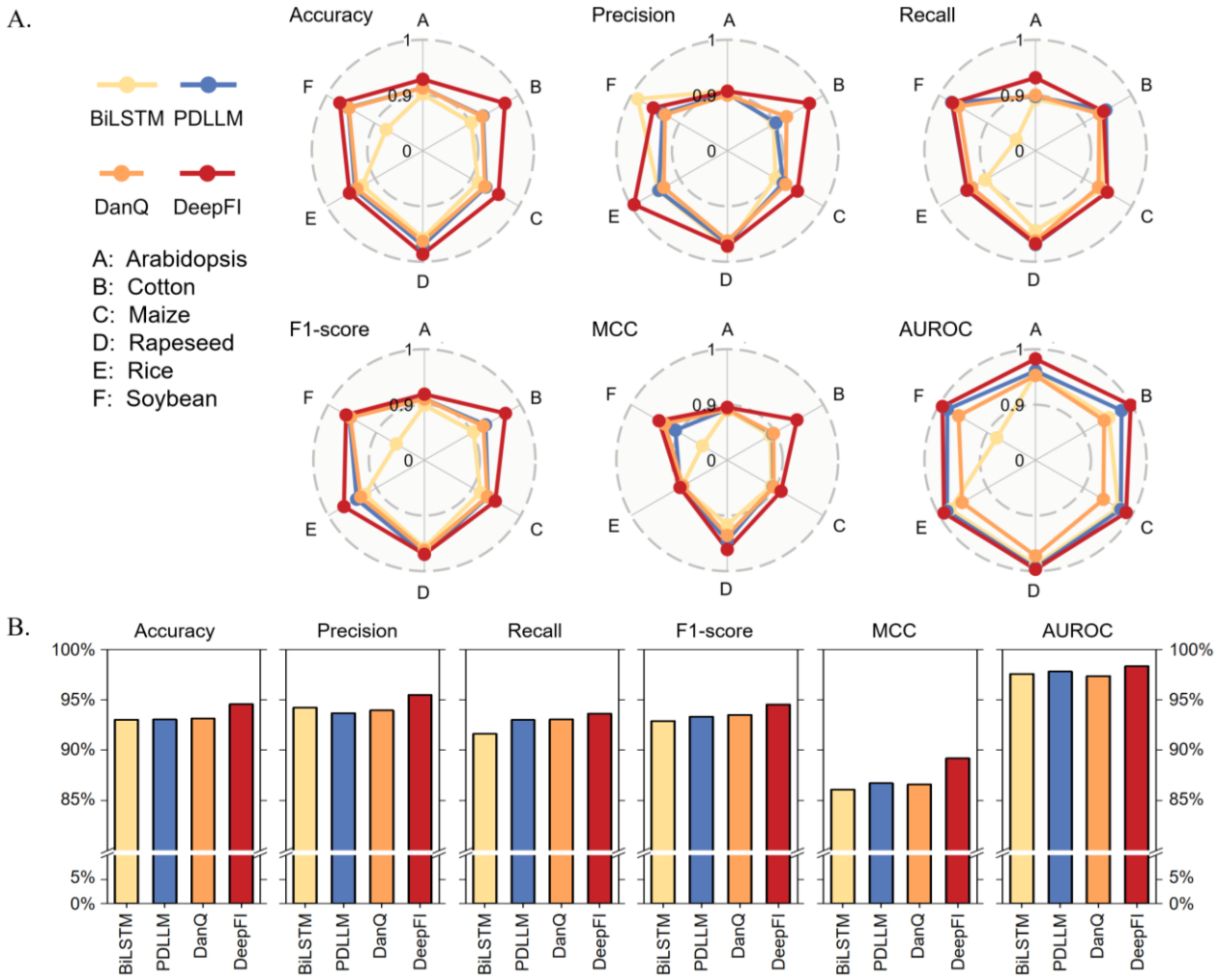


**Figure 6.** Model comparison. **A.** Comparison of the results of each evaluation index of the model on a single species dataset. **B.** Comparison of the results of each evaluation index of the model on the mixed dataset.

**Table 4.** Ablation study results.

| Model | Accuracy | Precision | Recall | F1-score | MCC | AUROC |
|---|---|---|---|---|---|---|
| **DeepFI** | **0.9458** | **0.9549** | **0.9363** | **0.9455** | **0.8918** | **0.9836** |
| *w*/o Dual-path encoding | 0.9333 | 0.9403 | 0.9253 | 0.9327 | 0.8772 | 0.9722 |
| *w*/o Multi-scale feature extraction | 0.9387 | 0.9474 | 0.9264 | 0.9368 | 0.8857 | 0.9761 |
| *w*/o Flow-Attention mechanism | 0.9390 | 0.9496 | 0.9319 | 0.9407 | 0.8862 | 0.9753 |
| *w*/o EAA mechanism | 0.9387 | 0.9486 | 0.9278 | 0.9381 | 0.8838 | 0.9737 |
| *w*/o EnhancedFGM | 0.9268 | 0.9324 | 0.9193 | 0.9258 | 0.8691 | 0.9674 |

datasets. DeepFI consistently demonstrated superior overall performance across all tested species. In terms of accuracy, DeepFI outperformed the best-performing benchmark model by 4.63% against the cotton dataset, and by 2.64% against the maize dataset. The DeepFI model also showed clear advantages with respect to precision during analysis of the cotton and rice datasets. With respect to F1 score, DeepFI achieved higher values across all six species with particularly notable improvements of 4.14% and 1.54% against cotton and maize datasets, respectively, compared to the strongest benchmark models. Moreover, in terms of the MCC metric, which measured classification balance, DeepFI surpassed the best benchmark model by 7.85% for the cotton dataset (Figure 6A). Model evaluation against the mixed dataset further confirmed that DeepFI consistently outperformed all benchmark models across all evaluation metrics. Specifically, DeepFI displayed improvements in both accuracy by 1.44% and precision by 1.26% over the best-performing benchmark model and also achieved higher recall and F1 score than that of BiLSTM, PDLLM, and DanQ. In addition, DeepFI achieved the highest MCC and AUROC values, where MCC improved by 2.47%, further demonstrating the excellent classification performance and robustness of this model (Figure 6B).

**Ablation study results**

The performance of DeepFI under different ablation settings demonstrated that, overall, the complete model achieved the best performance across all evaluation metrics, confirming the important role of each core module for enhancing model performance. Specifically, when the dual path encoding module was removed, all evaluation metrics decreased by more than 1%. Removal of the multi-scale feature extraction module led to the most notable drop in recall, demonstrating its key role in improving the model's ability to identify positive samples. In addition, removal of either the flow-attention mechanism or the EAA mechanism resulted in decreased overall performance, yielding reduced AUROC values of 0.9753 and 0.9737, respectively. Notably, omission of the EnhancedFGM dynamic adversarial training algorithm had the greatest impact on model performance with accuracy and MCC decreasing by 2.25% and 2.27%, respectively, strongly suggesting that this component was essential to overall model performance (Table 4). Based on ablation studies, each module was shown to substantially contribute to the overall performance of DeepFI. The integrated design of these components enabled the model to achieve state-of-the-art performance in plant intron prediction tasks.

**Cross-species prediction results**

To evaluate the cross-species generalization ability of DeepFI, zero-shot prediction experiments were conducted against two distantly related species, grape and sorghum. For the grape dataset, the model achieved an accuracy of 0.8760, precision of 0.8397, recall of 0.9385, F1 score of 0.8864, MCC of 0.7564, and AUROC of 0.9206, indicating that DeepFI could effectively predict functional introns in grape plants. For the sorghum dataset, model performance was slightly lower with an accuracy of 0.8263, precision of 0.8864, recall of 0.7528, F1 score of 0.8142, MCC of 0.6607, and AUROC of

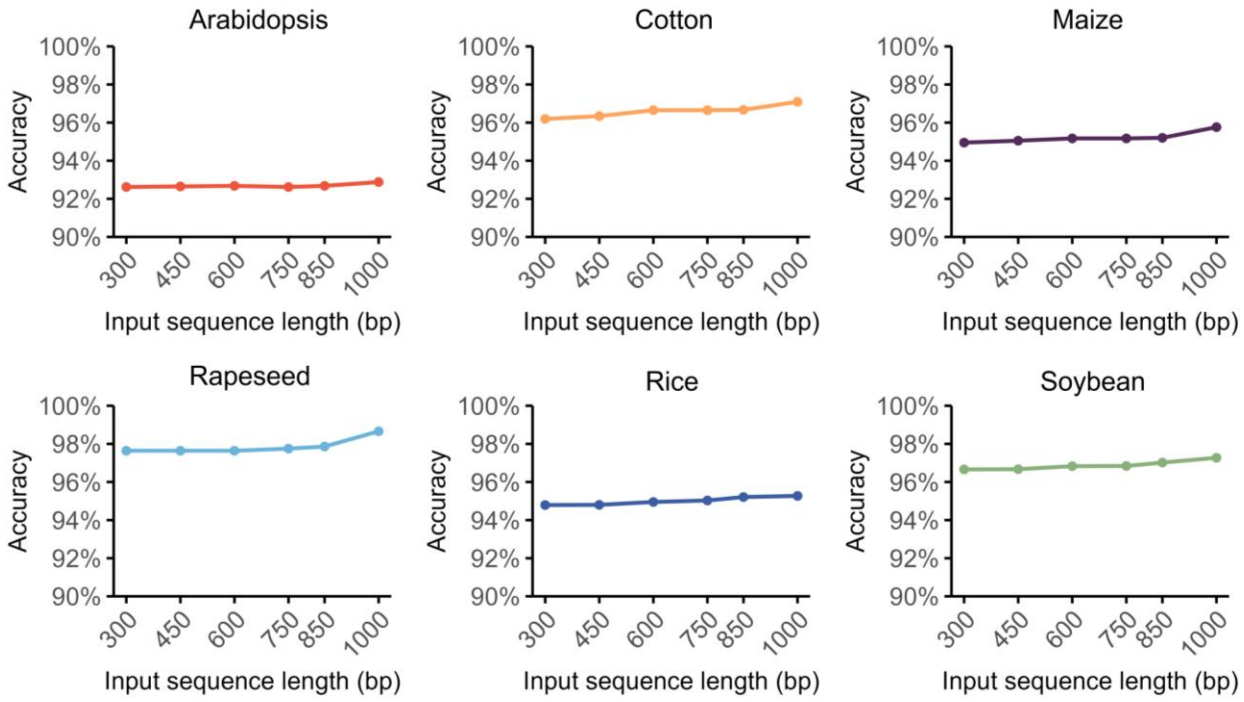


**Figure 7.** Model accuracy of different species with varying input sequence lengths.

0.8855. Notably, the MCC values for both species exceeded 0.65, suggesting that DeepFI maintained reliable discriminative power even across distantly related species. These results further validated the model's generalization ability and practical application potential for cross-species prediction of functional introns in plants.

**Effect of input sequence length**

Within a range of 300 bp to 1,000 bp, input sequence length had a relatively small impact on the performance of DeepFI, indicating that the model was well-adapted to intron sequences of varying lengths (Figure 7). Across all species, the highest prediction accuracy was achieved when the input sequence length was set to 1,000 bp, suggesting that longer sequences might help the model capture richer contextual information and thereby improve predictive performance. Overall, these results demonstrated the robustness and broad applicability of DeepFI for handling sequences of different lengths, providing strong support for its reliability in tasks requiring the prediction of functional introns across multi-species.

## Discussion

This study proposed a hybrid deep neural network model, DeepFI, which was designed for the accurate prediction of functional introns in plants. The model integrated dual path encoding, multi-scale feature extraction, and attention mechanisms, while incorporating an EnhancedFGM dynamic adversarial training strategy. This design enabled comprehensive improvements in feature representation, predictive accuracy, and robustness. For data construction, a total of 2,624 RNA-seq samples from eight plant species were collected from the NCBI SRA database and used to extract and identify 2,204,645 intron sequences. A cross-species mixed dataset containing 76,945 pairs of positive and negative samples was then established. For model evaluation, intron sequences from one complete chromosome of each species were reserved as an independent

test set, while introns from the remaining chromosomes were used for training and validation *via* a five-fold cross-validation strategy, which ensured that the training and test data were completely independent at the chromosome level. The results showed that DeepFI achieved excellent performance in functional intron prediction across six plant species with high scores across all evaluation metrics and small standard deviations. For the cross-species mixed dataset, the model achieved a test accuracy of 94.58% and an AUROC of 98.36%, underscoring its robustness and general applicability. Comparison with three categories of benchmark models further underscored the advantages of DeepFI. In contrast with the benchmark models, DeepFI leveraged the organic integration of multiple functional modules, making it particularly well-suited for functional intron prediction. Specifically, the dual path encoding module operated in parallel, enabling the model to learn both the functional semantics and structural characteristics of introns simultaneously. This design enhanced the capacity of the model to represent the multi-dimensional features of functional introns. The multi-scale feature extraction module captured sequence patterns at different levels, thereby improving the ability of the model to recognize diverse functional elements. To further enhance feature learning, multiple attention mechanisms were incorporated to address the challenges of local focus and global context fusion in sequence modeling. The EnhancedFGM dynamic adversarial training algorithm applied gradient perturbations to both the embedding and convolutional layers, which improved the anti-interference ability of the model. For single-species datasets, DeepFI consistently obtained the highest accuracies with improvements of 4.63% and 2.64% over the best-performing benchmark models against cotton and maize, respectively. With respect to the mixed dataset, DeepFI also outperformed all benchmark models, achieving the best results across all evaluation metrics. Cross-species prediction experiments further validated the generalization capacity of the DeepFI model. By using grape and sorghum, which were not included in training, as independent test sets, DeepFI achieved prediction accuracies of 87.60% and 82.63%, and AUROC values of 92.06% and 88.55%, respectively. Although performance was lower than that observed for species used during training, MCC values for both datasets exceeded 0.65, indicating that DeepFI maintained reliable discriminative power in zero-shot prediction scenarios. These results suggested that DeepFI possessed cross-species prediction potential, though its generalization ability remained constrained by the diversity of training data and interspecies differences [28]. Future research could incorporate more diverse and representative training samples to improve the generalizability of the model. In addition, leveraging transfer learning or domain adaptation techniques might further enhance the predictive accuracy of DeepFI in cross-species scenarios [29, 30]. Overall, DeepFI demonstrated outstanding accuracy, stability, and robustness in plant functional intron prediction. The model not only effectively handled intron sequences of varying lengths but also showed potential for cross-species prediction. In future research, the training dataset should be expanded to cover more representative plant species, particularly those of high agricultural value that currently lack sufficient data to improve the model's universality and prediction accuracy. In addition, a user-friendly online analysis platform needs to be developed to integrate DeepFI's prediction capabilities, offering tools for visualization and interactive analysis of results to support plant genomics research.

**Acknowledgements**

This work was supported by the National Natural Science Foundation of China (Grant No. 32070684 and 31571306). The authors would like to thank Supercomputing Center in Shandong Agricultural University for the technical support.

**Data availability**

All generated dataset and source code used in this study are available in the repository at https://github.com/Sophiefight123/DeepFI.

## References


1. Irimia M, Roy SW. 2014. Origin of spliceosomal introns and alternative splicing. Cold Spring Harb Perspect Biol. 6(6):a016071.
2. Chorev M, Carmel L. 2012. The function of introns. Front Genet. 3:55.
3. Jo BS, Choi SS. 2015. Introns: The functional benefits of introns in genomes. Genomics Inform. 13(4):112-118.
4. Parenteau J, Elela SA. 2019. Introns: Good day junk is bad day treasure. Trends Genet. 35(12):923-934.
5. Parenteau J, Maignon L, Berthoumieux M, Catala M, Gagnon V, Elela SA. 2019. Introns are mediators of cell response to starvation. Nature. 565(7741):612-617.
6. Morgan JT, Fink GR, Bartel DP. 2019. Excised linear introns regulate growth in yeast. Nature. 565(7741):606-611.
7. Zeng C, Jiao Q, Jia T, Hu X. 2022. Updated progress on group II intron splicing factors in plant chloroplasts. Curr Issues Mol Biol. 44(9):4229-4239.
8. Li X, Jiang Y. 2024. Research progress of group II intron splicing factors in land plant mitochondria. Genes. 15(2):176.
9. Li Z, Wang S, Cheng J, Su C, Zhong S, Liu Q, *et al*. 2016. Intron lariat RNA inhibits microRNA biogenesis by sequestering the dicing complex in Arabidopsis. PLoS Genet. 12(11):e1006422.
10. Wu HW, Deng S, Xu H, Mao HZ, Liu J, Niu QW, *et al*. 2018. A noncoding RNA transcribed from the AGAMOUS (AG) second intron binds to CURLY LEAF and represses AG expression in leaves. New Phytol. 219(4):1480-1491.
11. Cheng S, Guo M, Wang C, Liu X, Liu Y, Wu X. 2016. MiRTDL: A deep learning approach for miRNA target prediction. IEEE/ACM Trans Comput Biol Bioinform. 13(6):1161-1169.
12. Pan X, Shen HB. 2018. Predicting RNA-protein binding sites and motifs through combining local and global deep convolutional neural networks. Bioinformatics. 34(20):3427-3436.
13. Wang J, Wang L. 2019. Deep learning of the back-splicing code for circular RNA formation. Bioinformatics. 35(24):5235-5242.
14. Zhang L, Qin X, Liu M, Liu G, Ren Y. 2021. BERT-m7G: A transformer architecture based on BERT and stacking ensemble to identify RNA N7-methylguanosine sites from sequence information. Comput Math Methods Med. 2021:7764764.
15. Ritu, Gupta S, Sharma NK, Shankar R. 2022. DeepPlnc: Bi-modal deep learning for highly accurate plant lncRNA discovery. Genomics. 114(5):110443.
16. Xiang S, Zhang T, Wu M. 2023. M6ATMR: Identifying N6-methyladenosine sites through RNA sequence similarity matrix reconstruction guided by Transformer. Peer J. 11:e15899.
17. Yang X, Liu G, Feng G, Bu D, Wang P, Jiang J, *et al*. 2024. GeneCompass: Deciphering universal gene regulatory mechanisms with a knowledge-informed cross-species foundation model. Cell Res. 34(12):830-845.
18. Yang Q, Sun J, Wang X, Wang J, Liu Q, Ru J, *et al*. 2025. SVLearn: A dual-reference machine learning approach enables accurate cross-species genotyping of structural variants. Nat Commun. 16(1):2406.
19. Li H, Zhang Y, Bing J, Han J, Hu J, Zhao H, *et al*. 2023. Intron-capture RNA-seq reveals the landscape of intronic RNAs in Arabidopsis. Plant Physiol Biochem. 196:75-88.
20. Zhou J, Troyanskaya OG. 2015. Predicting effects of noncoding variants with deep learning-based sequence model. Nat Methods. 12(10):931-934.
21. Quang D, Xie X. 2016. DanQ: A hybrid convolutional and recurrent deep neural network for quantifying the function of DNA sequences. Nucleic Acids Res. 44(11):e107.
22. Park S, Koh Y, Jeon H, Kim H, Yeo Y, Kang J, *et al*. 2020. Enhancing the interpretability of transcription factor binding site prediction using attention mechanism. Sci Rep. 10:13413.
23. Li J, Pu Y, Tang J, Zou Q, Guo F. 2021. DeepATT: A hybrid category attention neural network for identifying functional effects of DNA sequences. Brief Bioinform. 22(3):bbaa159.
24. Yao Z, Zhang W, Song P, Hu Y, Liu J. 2023. DeepFormer: A hybrid network based on convolutional neural network and flow-attention mechanism for identifying the function of DNA sequences. Brief Bioinform. 24(4):bbad211.
25. Ji Y, Zhou Z, Liu H, Davuluri RV. 2021. DNABERT: Pre-trained bidirectional encoder representations from transformers model for DNA-language in genome. Bioinformatics. 37(15):2112-2120.
26. Zhou Z, Ji Y, Li W, Dutta P, Davuluri R, Liu H. 2023. DNABERT-2: Efficient foundation model and benchmark for multi-species genomes. Presented at 12th International Conference on Learning Representations, ICLR 2024, Hybrid, Vienna, Austria.
27. Liu G, Chen L, Wu Y, Han Y, Bao Y, Zhang T. 2025. PDLLMs: A group of tailored DNA large language models for analyzing plant genomes. Mol Plant. 18(2):175-178.
28. Michael TP. 2014. Plant genome size variation: Bloating and purging DNA. Brief Funct Genomics. 13(4):308-317.
29. Redko I, Morvant E, Habrard A, Sebban M, Bennani Y. 2020. A survey on domain adaptation theory: Learning bounds and theoretical guarantees. arXiv:2004.11829v6.
30. Garau-Luis JJ, Bordes P, Gonzalez L, Roller M, Almeida BPD, Hexemer L, *et al*. 2024. Multi-modal transfer learning between biological foundation models. arXiv:2406.14150v1.