iDNAProt-ES

General Claimed:

(A) proposed a DNA-binding protein prediction method that utilizes both sequence based evolutionary and structure based features of proteins to identify their DNA-binding functionality

(B)used SVM-RFE (Recursive Feature Elimination) method to extract an optimal set of features.

(C) trained a prediction model using SVM with linear kernel.

(D) achieves accuracy: jackknife test: **90.18%**, 10-fold cross-validation test: **88.87%**, Independent dataset: **80.64%**

Features Used:

1. Features generated from PSSM file:

Feature Name	Feature Vector Size
1. Amino acid composition	20
2. Dubchak features	105
3. PSSM Bigram(represents the transition probabilities of two adjacent amino acid residue positions)	400
4. PSSM 1-lead Bigram (transition probabilities of the amino acid residue positions at 1 distance or separation)	400
5. PSSM Composition	20
6. PSSM Auto-Covariance	200
7. PSSM Segmented Distribution (used Fp = 5, 10, 25.)	200

Total = 1,345

2. SPIDER (used Spider2) based features:

Feature Name	Feature Vector Size
1. Secondary Structure Occurrence(count or frequency of three types of motifs structural motifs in proteins: α-helix (H)	3
2. Secondary Structure Composition(secondary structure motif occurrence normalized by the length of the phage protein length)	3
3. Accessible Surface Area Composition, Torsional Angles Composition, Structural Probabilities Composition	12
4. Torsional Angles Bigram	64
5. Structural Probablities Bigram	9
6. Torsional Angles Auto-Covariance	80
7. Structural Probablities Auto-Covariance	30

Total = 201

Overall Feature Vector Size: 1345 + 201 = 1546

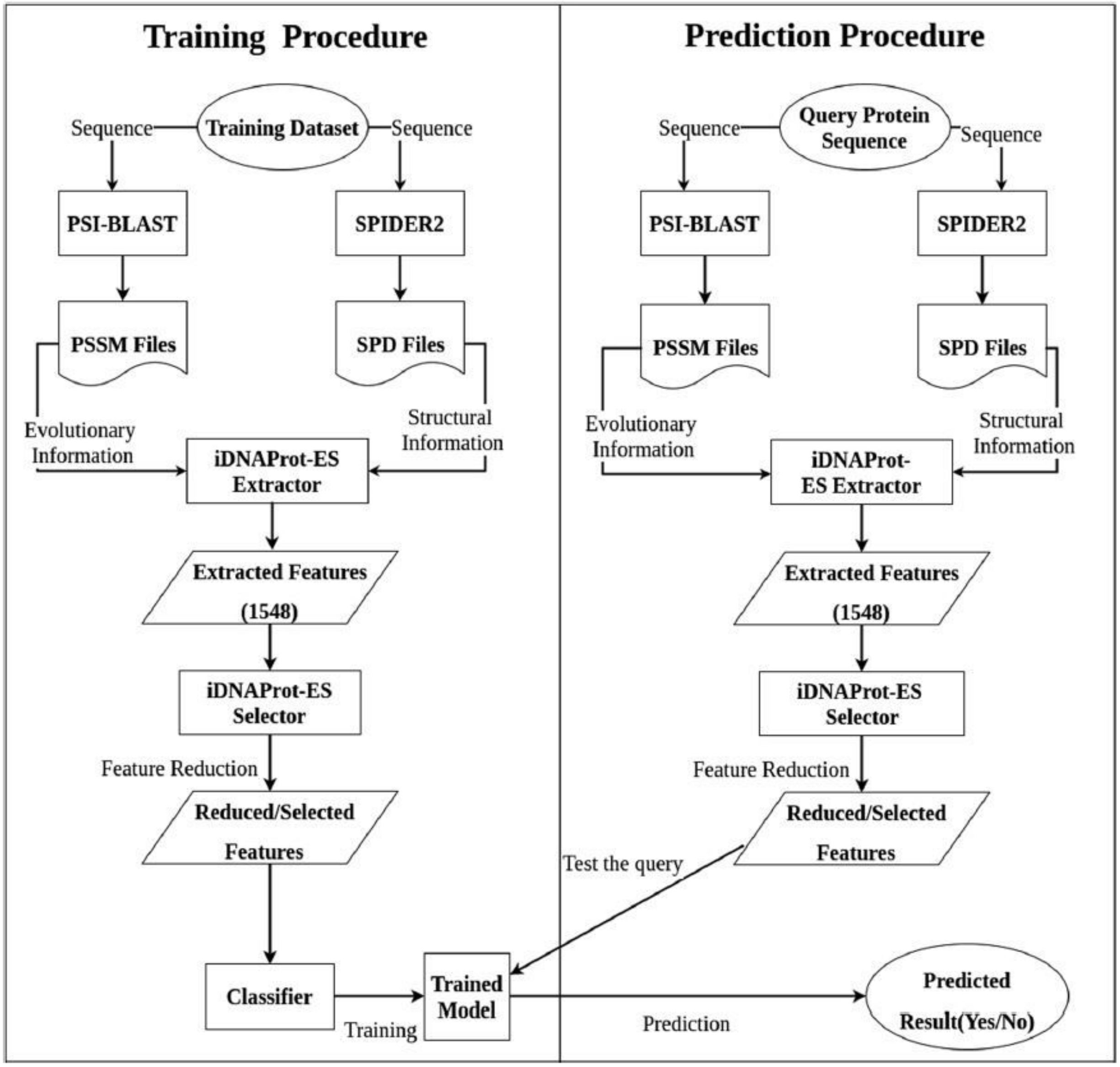

Figure: System flow diagram of iDNAProt-ES showing the training and prediction procedure.

Performance Measures:

(A) Jackknife on the benchmark dataset:

Accuracy (90.18%), SE (0.9038), SP(0.9000), MCC (0.8036), auROC(0.9412)

(B) 10-fold cross validation on the benchmark dataset:

ACC(88.87%), SE(0.8945), SP(0.8826), MCC(0.7788), auROC(0.9391), auPR(0.8828)

(C) Independent dataset:

Accuracy (80.64%), SE (0.8131), SP(0.8000), MCC (0.6130), auROC(0.8434)