1. Cascade Multi-Level Feature Subset Selection Analysis

Different threshold values have been adopted for the calculation of the optimum features. The selection process has been followed via a three-level feature selection. In the first level, a threshold value of 0.03 has been used to make the selection of top optimum features and this process has been repeated for a threshold value of 0.05 and 0.07 cascading. We have observed that in the third level the dimension of the feature decrease tremendously. By taking the topmost optimum features from the third level has been further made subject to a fused feature. these optimum features bring a great improvement in the model classification accuracy. Different classification algorithm has been evaluated to testify the outcome based on these optimum features. Every algorithm, perform consistently, better upon these features. For reference, the schematic diagram of the proposed model is given in Manuscript in Figure.1

Layer-1 Cascade Multi-Level Feature Selection

Different optimum representation of the baseline primitive feature of Layer-1 has been obtained via Cascade Multi-Level Feature selection algorithm. These features are described as follows.

Fkmer: Fkmer feature has been calculated by making a fuse feature vector of 3kmer, 4kmer, and 5 kmer.

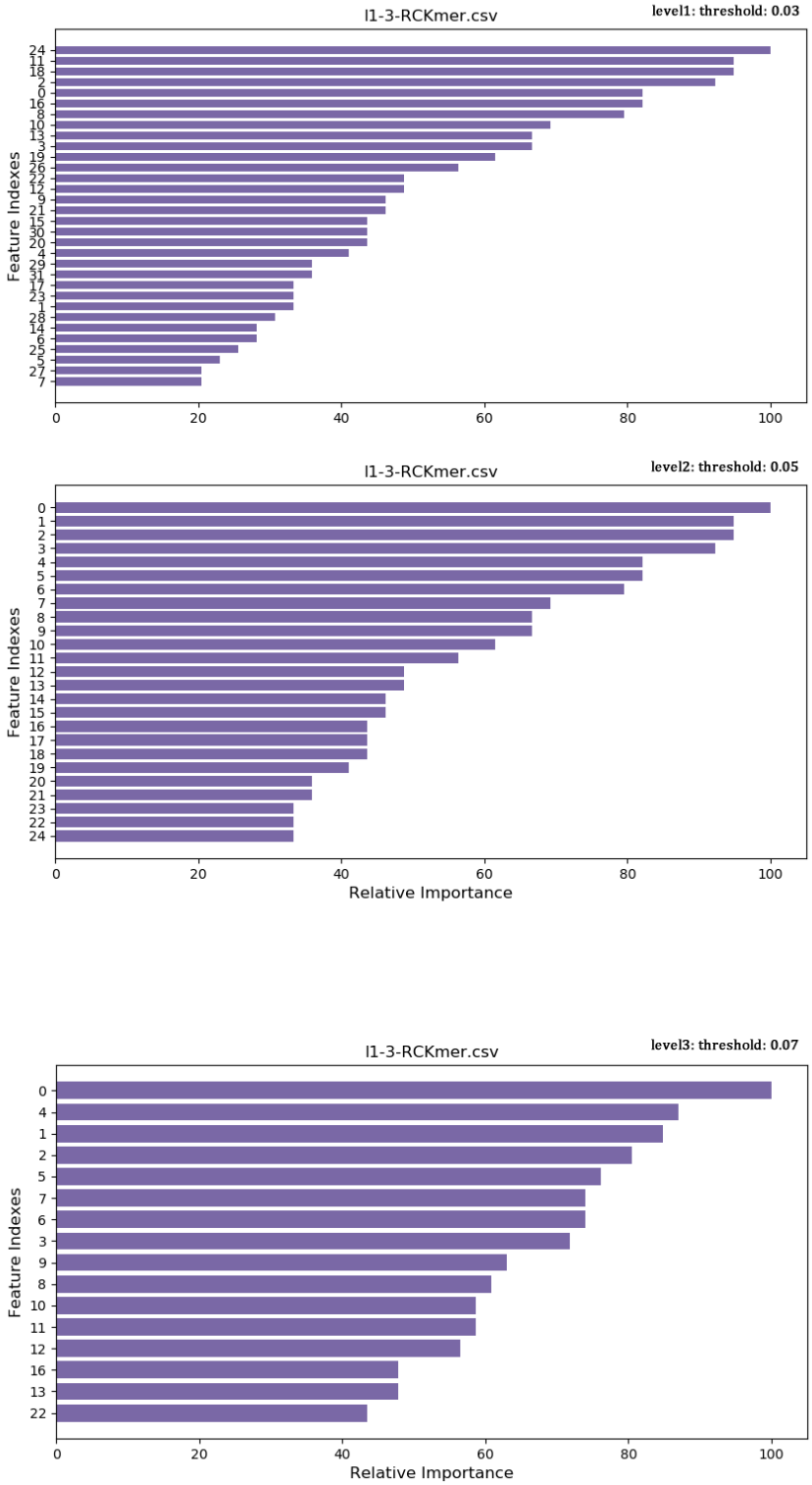


Figure.1. Cascade Level Feature Selection Plot of 3kmer over different threshold values

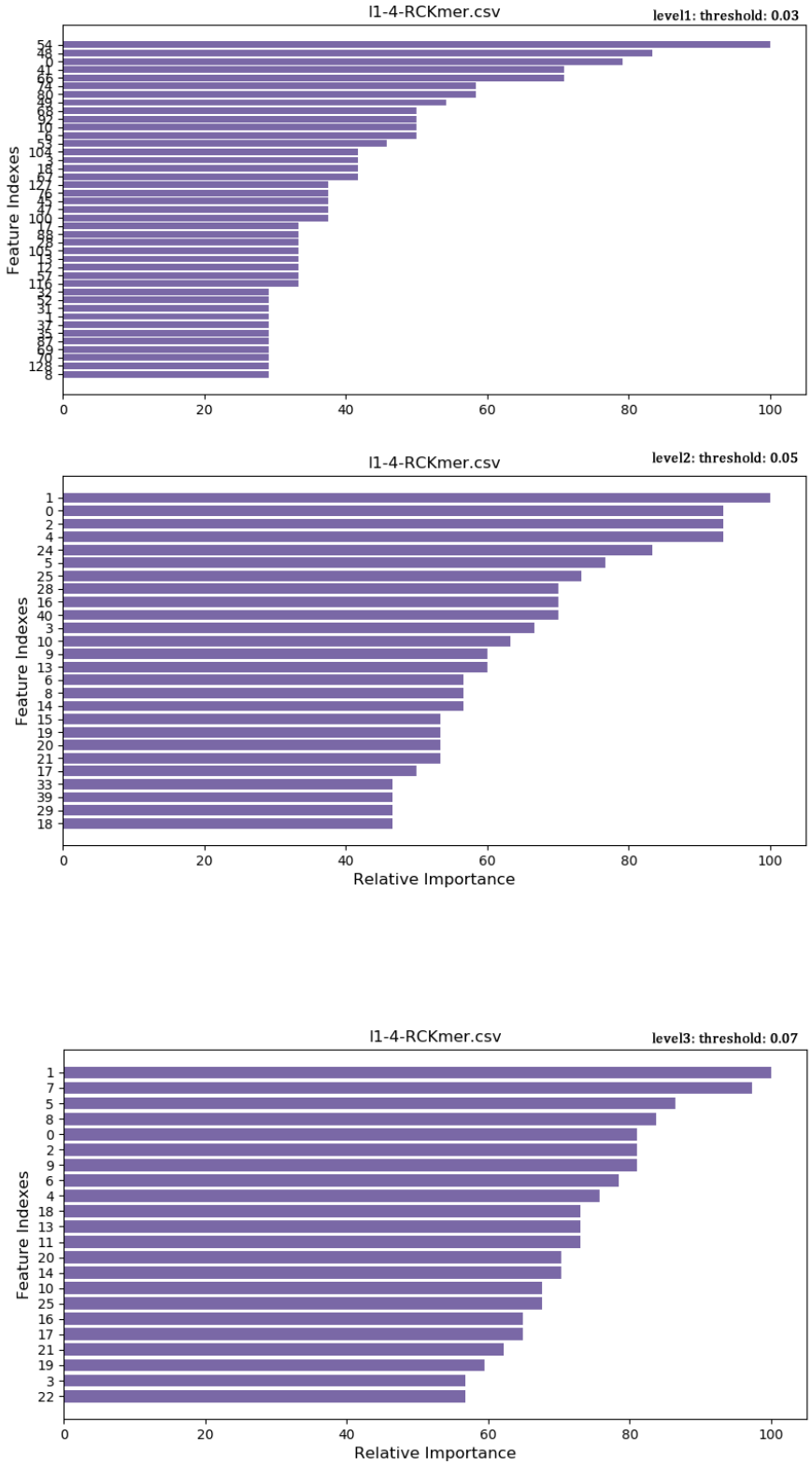


Figure.2. Cascade Level Feature Selection Plot of 4kmer over different threshold values

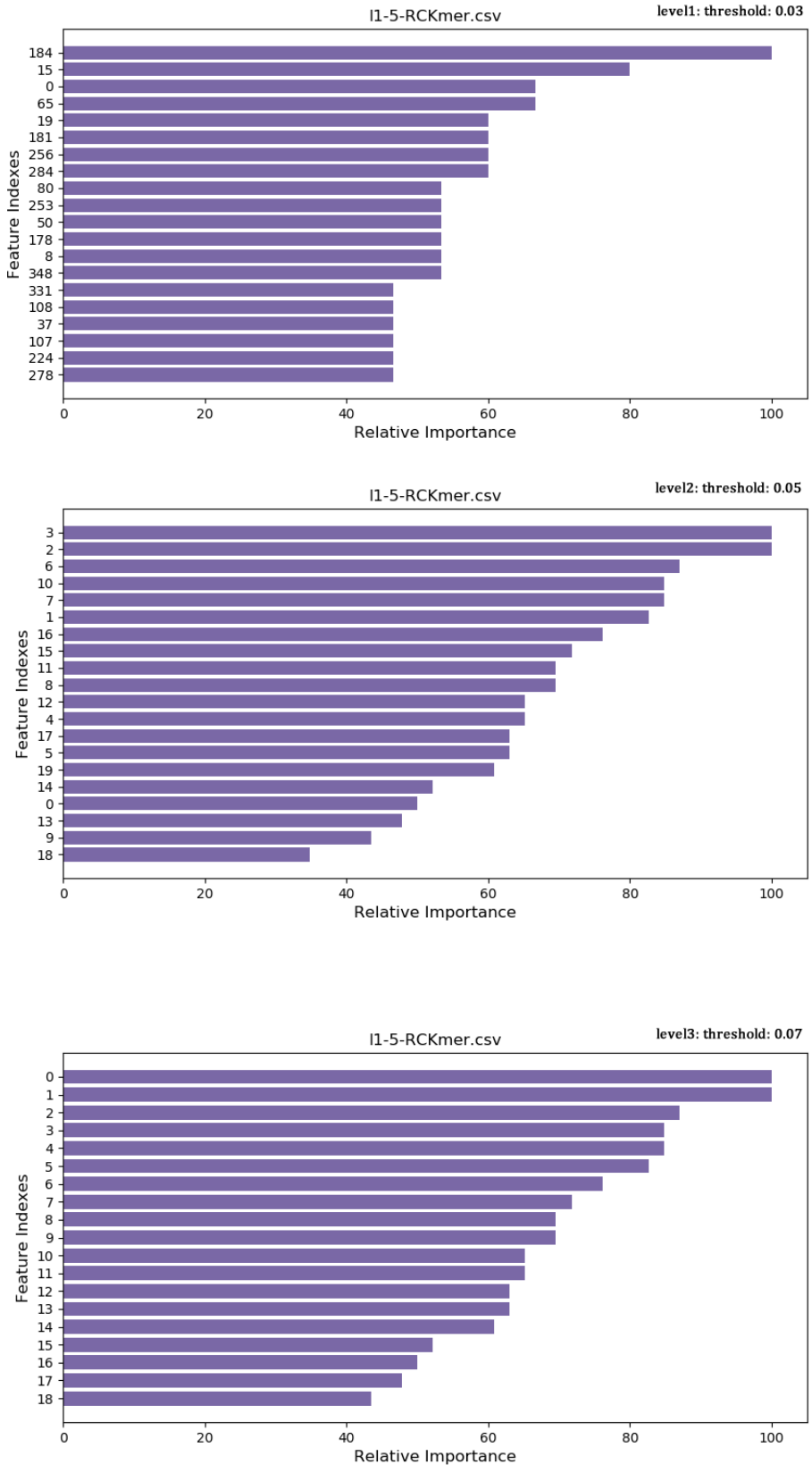
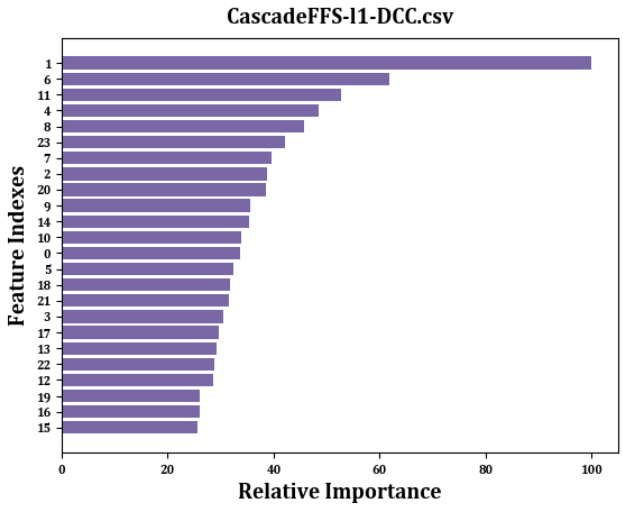


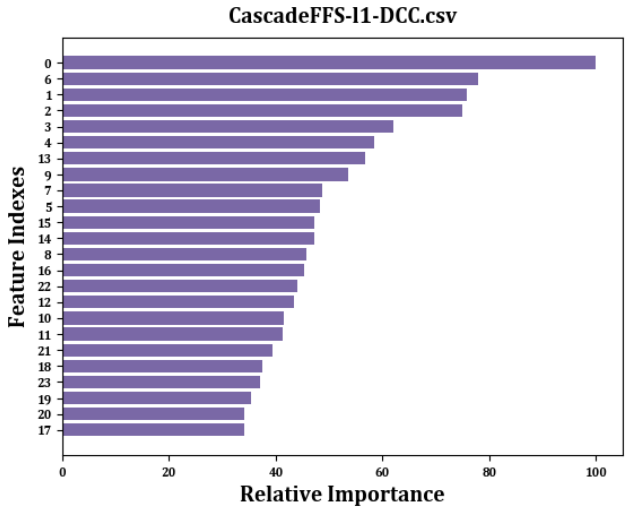
Figure.3. Cascade Level Feature Selection Plot of 5kmer over different threshold values

DCC(Di Nucleotide Cross Correlation Composition)

Threshold 0.03



Threshold 0.05



Threshold 0.03

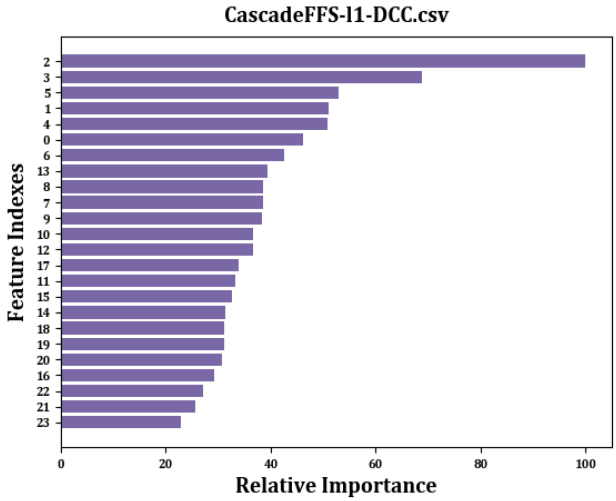
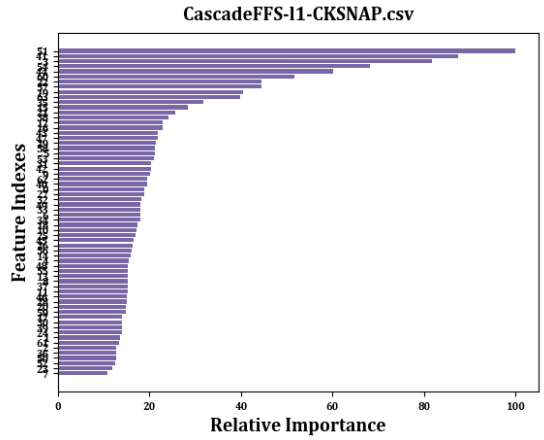


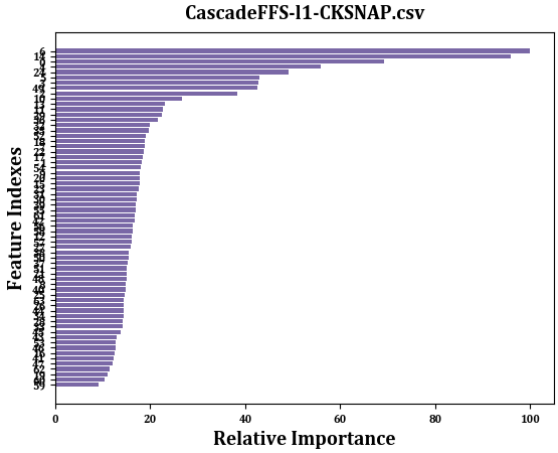
Figure.4. Cascade Level Feature Selection Plot DCC over different threshold values

CKSNAP (Composition of K spaced Nucleic Acid Pairs)

Threshold. 0.03



Threshold. 0.05



Threshold. 0.07

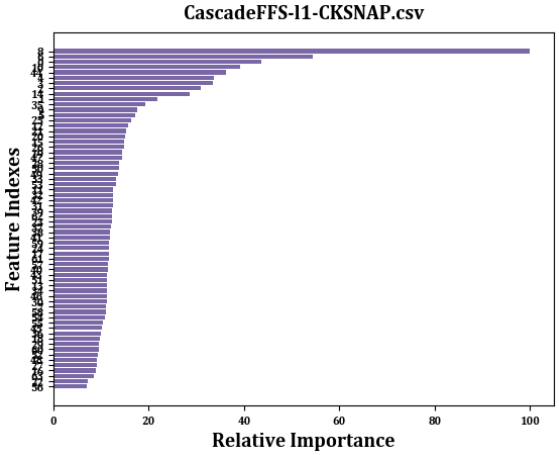
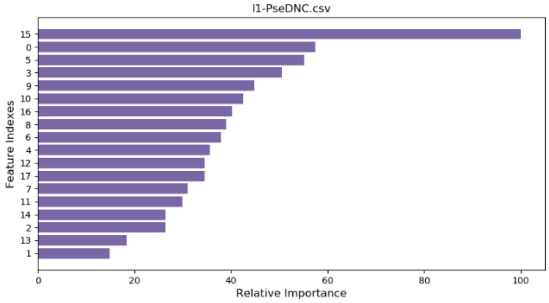


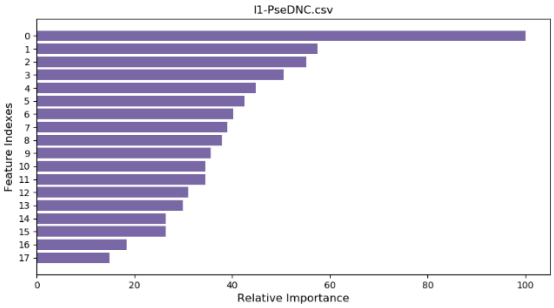
Figure.5. Cascade Level Feature Selection Plot of CKSNAP over different threshold values

PseDNC (Pseudo Dinucleotide Composition)

Threshold. 0.03



Threshold. 0.05



Threshold. 0.07

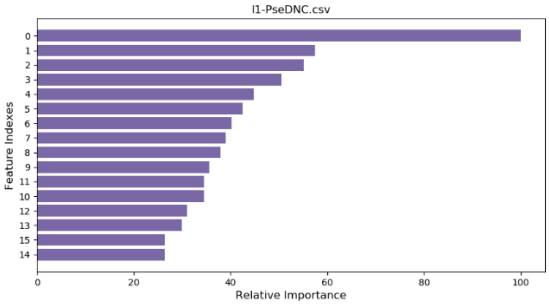
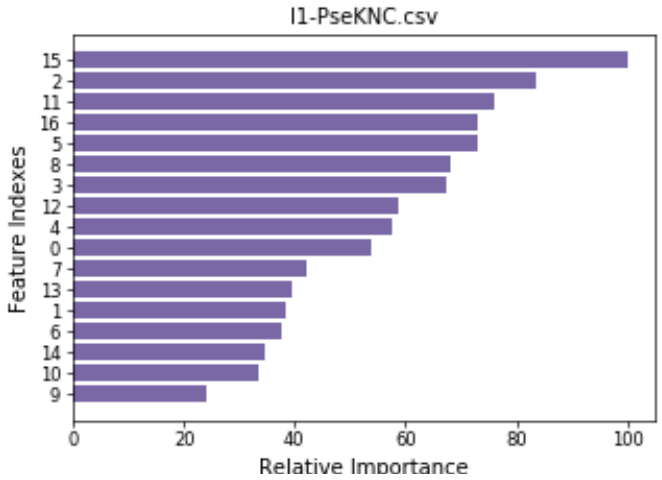


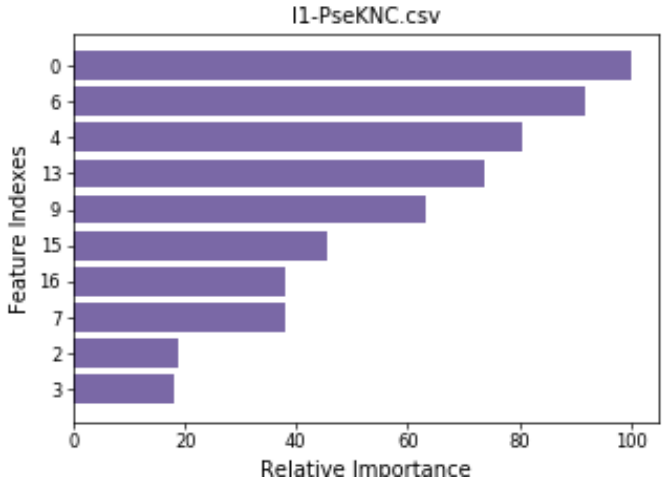
Figure.6. Cascade Level Feature Selection Plot of PseDNC over different threshold values

PseKNC (Pseudo K-nucleotide Composition)

Threshold. 0.03



Threshold. 0.05



Threshold. 0.07

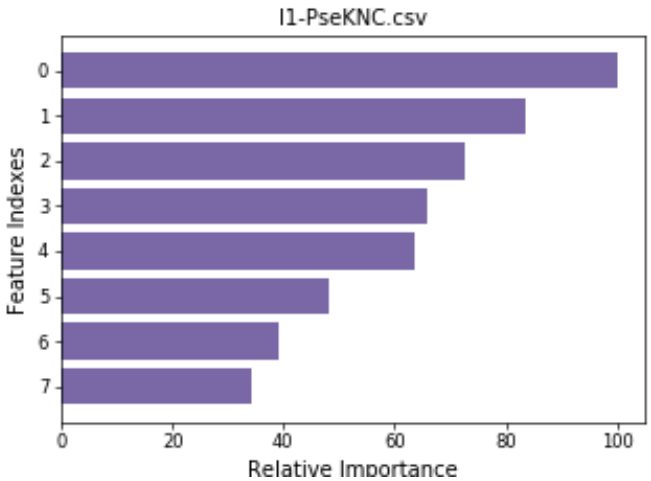


Figure.7. Cascade Level Feature Selection Plot of PseKNC over different threshold values

Layer-2 Cascade Multi-Level Feature Selection

Different optimum representation of the baseline primitive feature of layer-2 has been obtained via Cascade Multi-Level Feature selection algorithm. These features are described as follows.

Fkmer: Fkmer feature has been calculated by making a fuse feature vector of 3kmer, 4kmer, and 5 kmer.

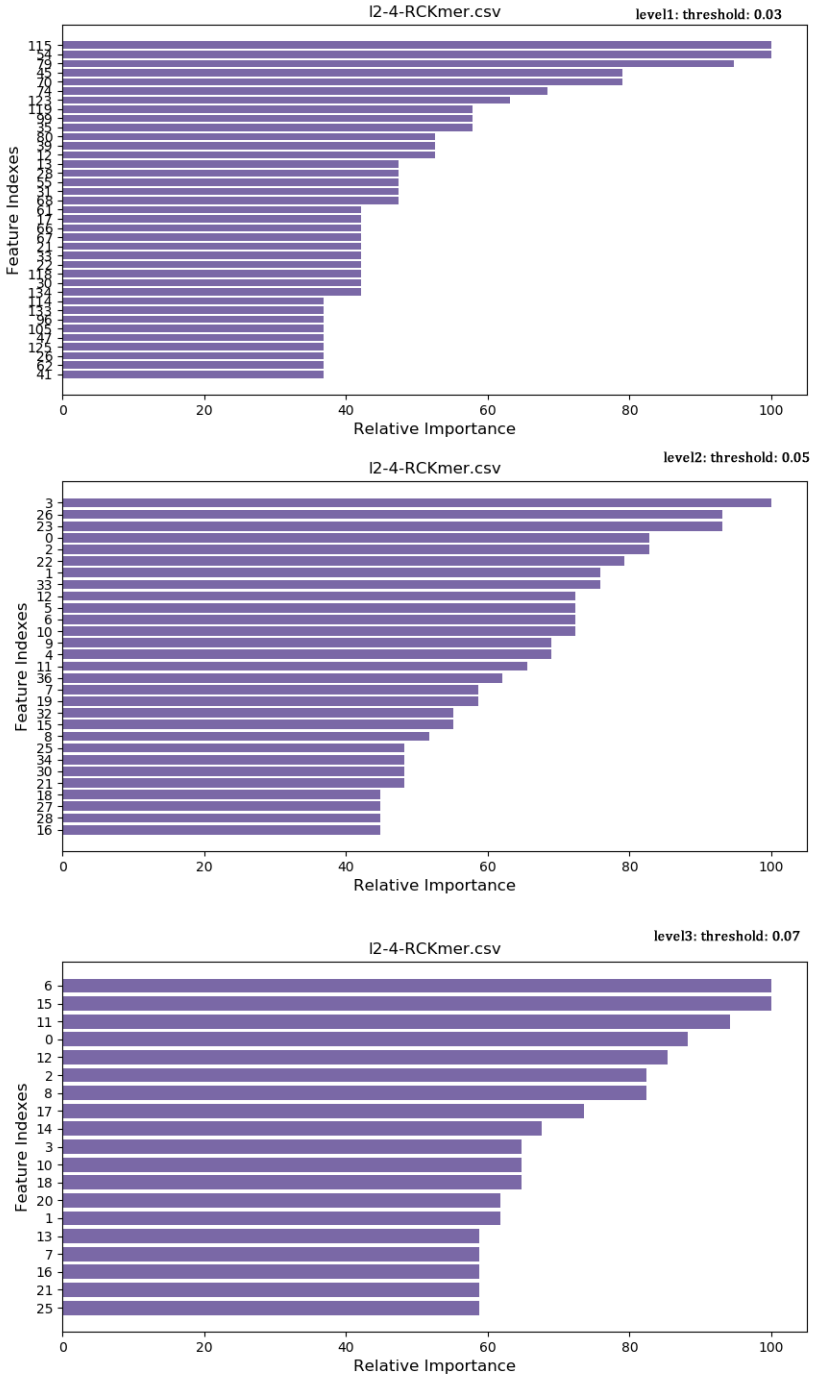


Figure.8 Cascade Level Feature Selection Plot of 3kmer over different threshold values

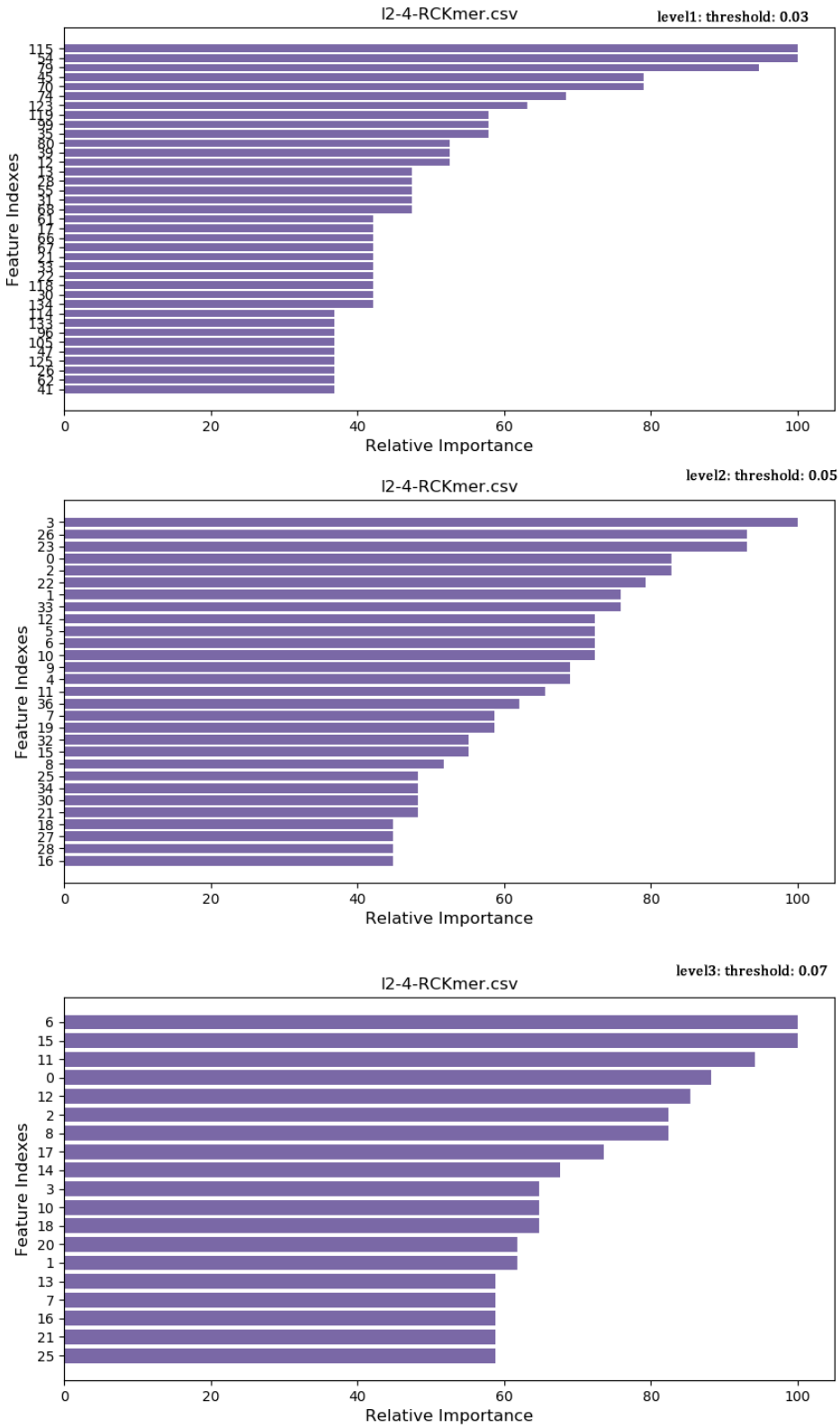


Figure.9. Cascade Level Feature Selection Plot of 4kmer over different threshold values

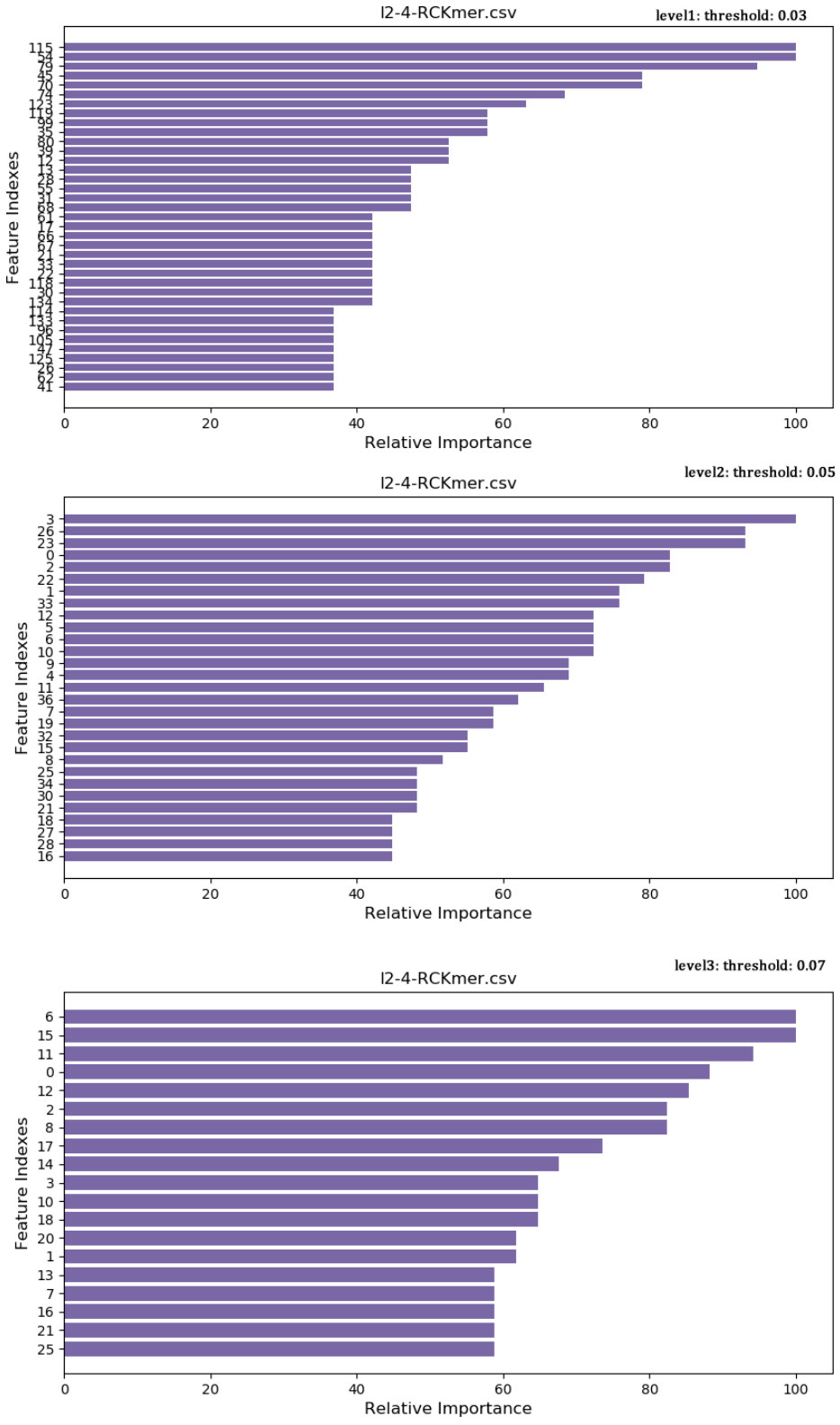
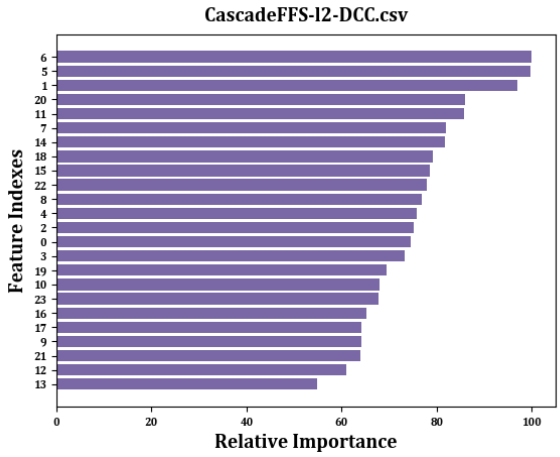


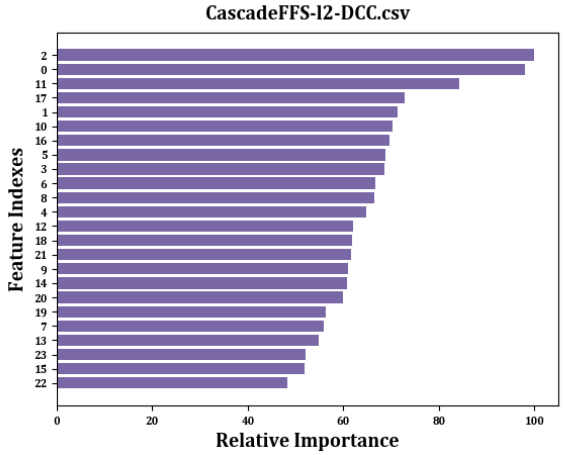
Figure.10. Cascade Level Feature Selection Plot of 5kmer over different threshold values

DCC(Di Nucleotide Cross Correlation Composition)

Threshold. 0.03



Threshold. 0.05



Threshold. 0.07

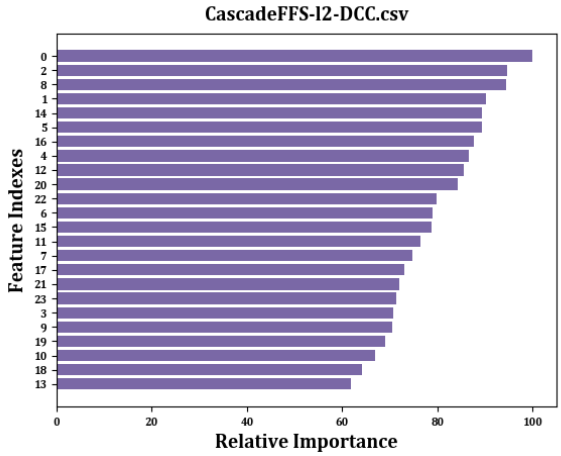
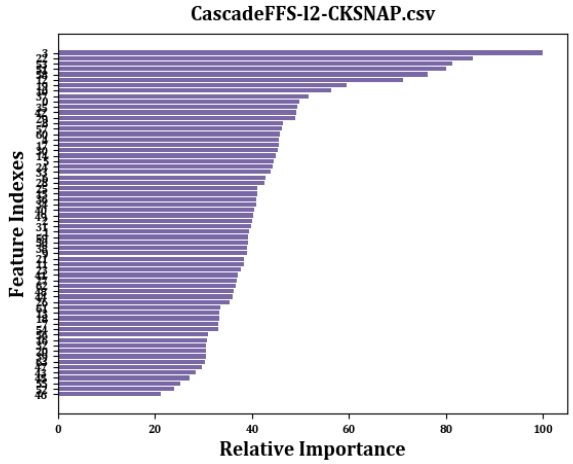


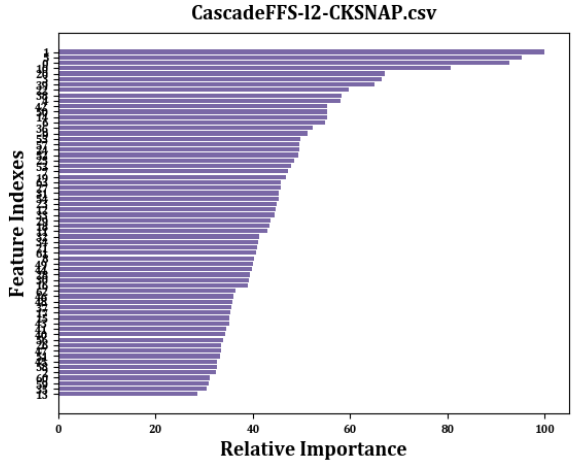
Figure.11. Cascade Level Feature Selection Plot of DCC over different threshold values

CKSNAP (Composition of K spaced Nucleic Acid Pairs)

Threshold. 0.03



Threshold. 0.05



Threshold. 0.07

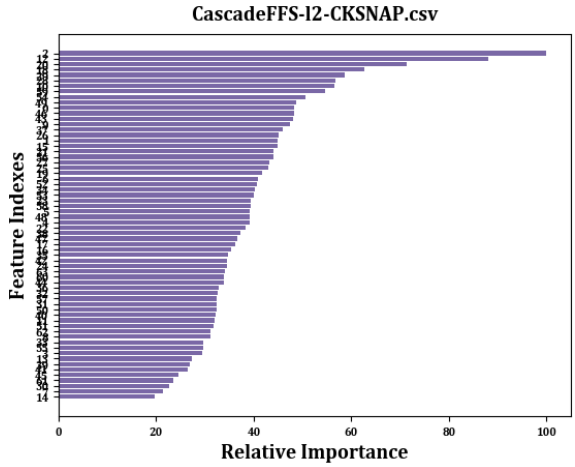
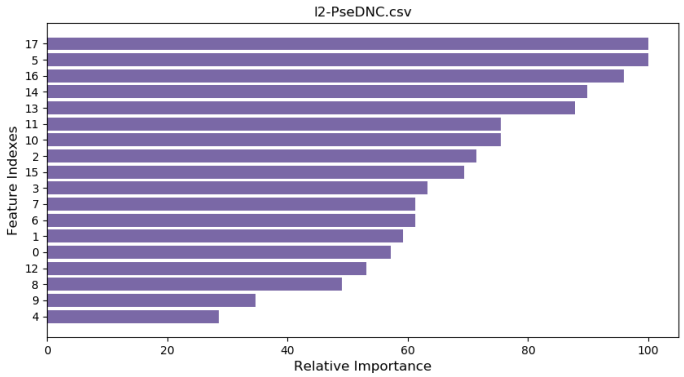


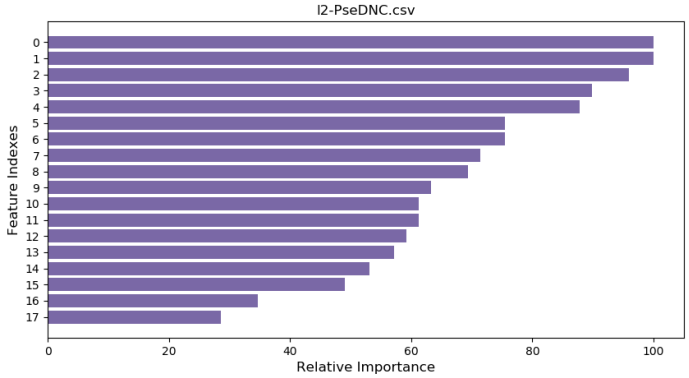
Figure.12. Cascade Level Feature Selection Plot of CKSNAP over different threshold values

PseDNC (Pseudo Dinucleotide Composition)

Threshold. 0.03



Threshold. 0.05



Threshold. 0.07

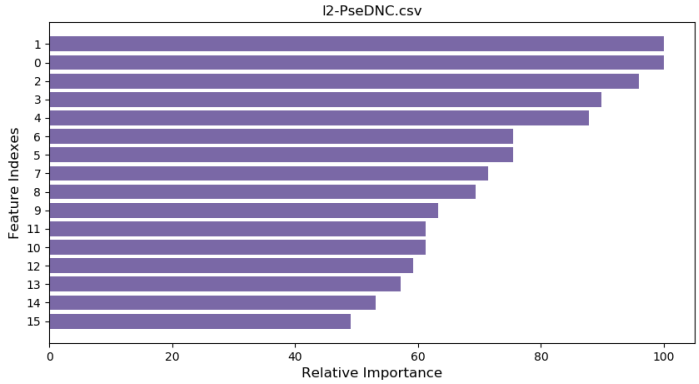
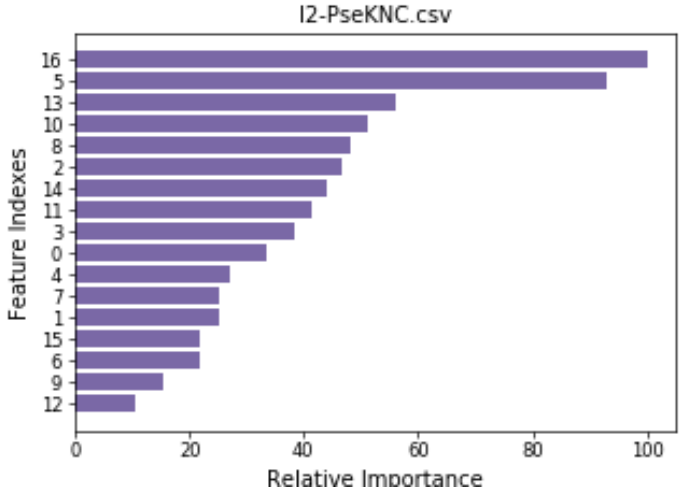


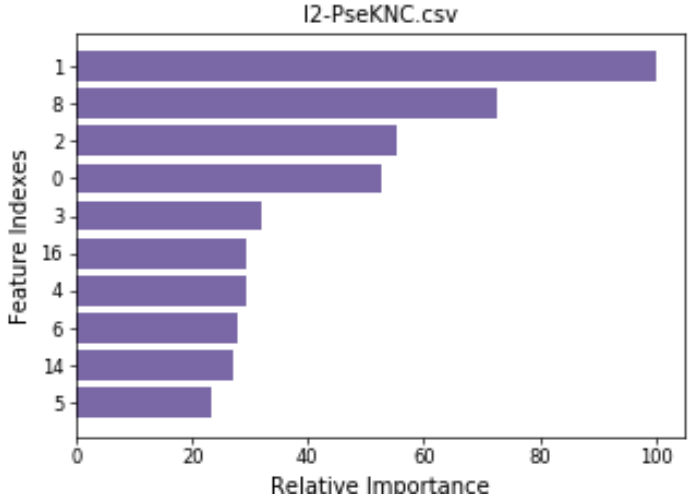
Figure.13. Cascade Level Feature Selection Plot of PseDNC over different threshold values

PseKNC (Pseudo K-nucleotide Composition)

Threshold. 0.03



Threshold. 0.05



Threshold. 0.07

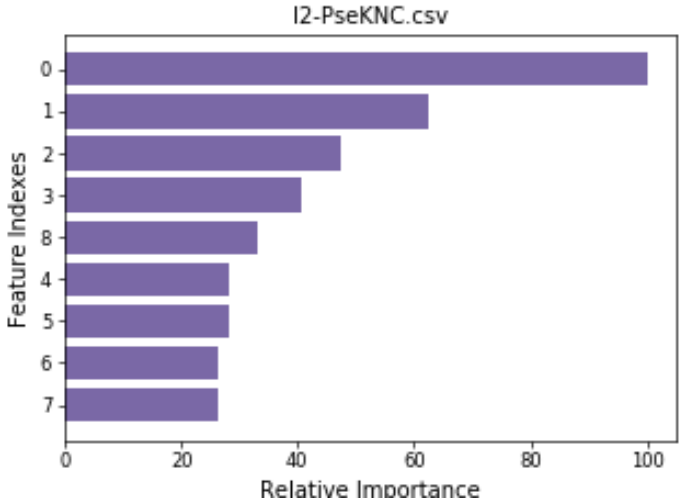


Figure.14. Cascade Level Feature Selection Plot of PseKNC over different threshold values

2. **Primitive Features Classification Results over different classification algorithms**

Detail classification results of the different algorithms over various employed feature extraction are given. As follow. Layer-1

Layer-1 (Individual Feature) ROC Curves

Figure.15. ROC Curve of CKSNAP feature space over Layer-1

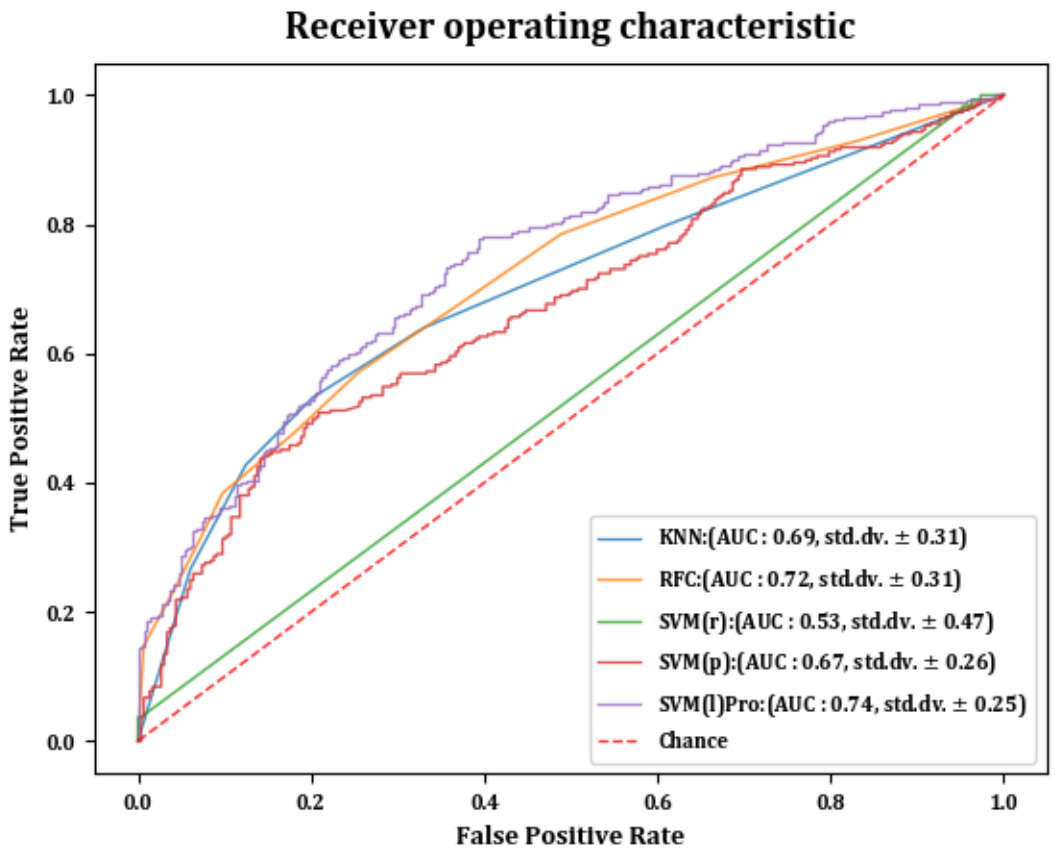


Figure.15. ROC curve of CKSNAP feature space over Layer-1

Figure.16. ROC curve of DCC (Dinucleotide Cross-Correlation) feature space over Layer-1

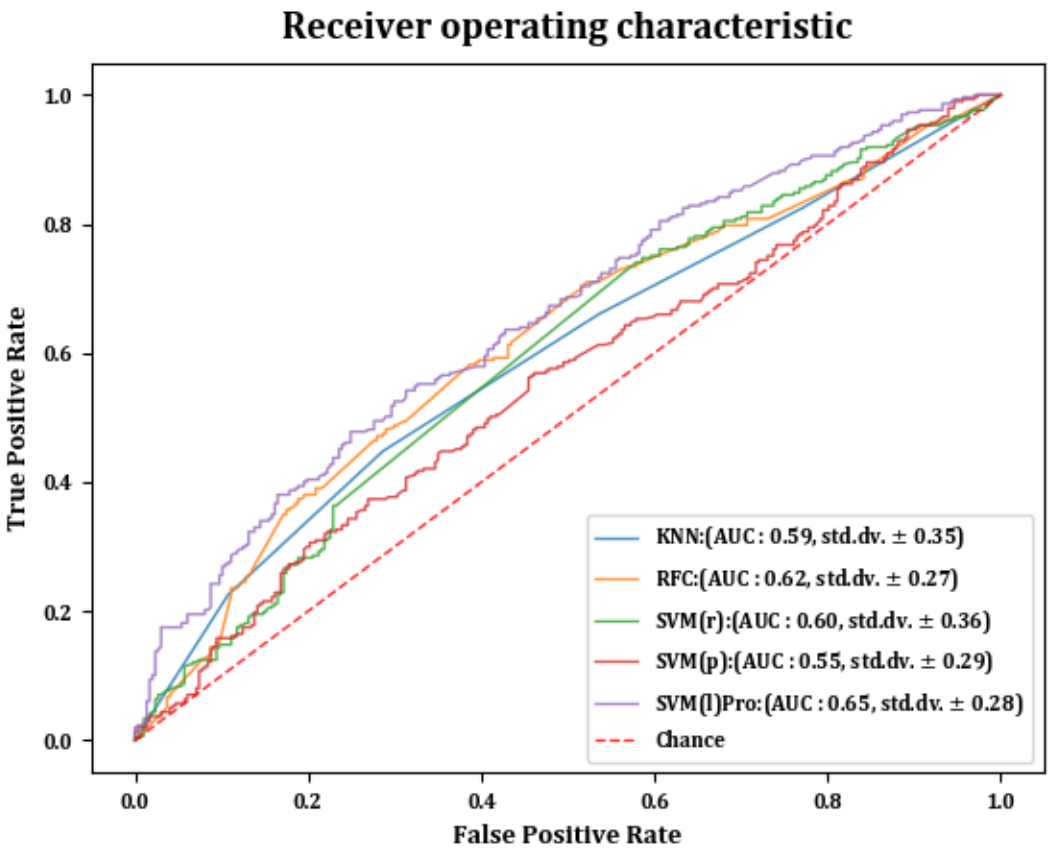


Figure.16. ROC curve of DCC Feature space over Layer-1

Figure.17. ROC curve of FKmer feature space over Layer-1. Fkmer is a fuse feature vector of 3kmer, 4kmer, and 5kmer. Over the fuse feature vector, different classification performances

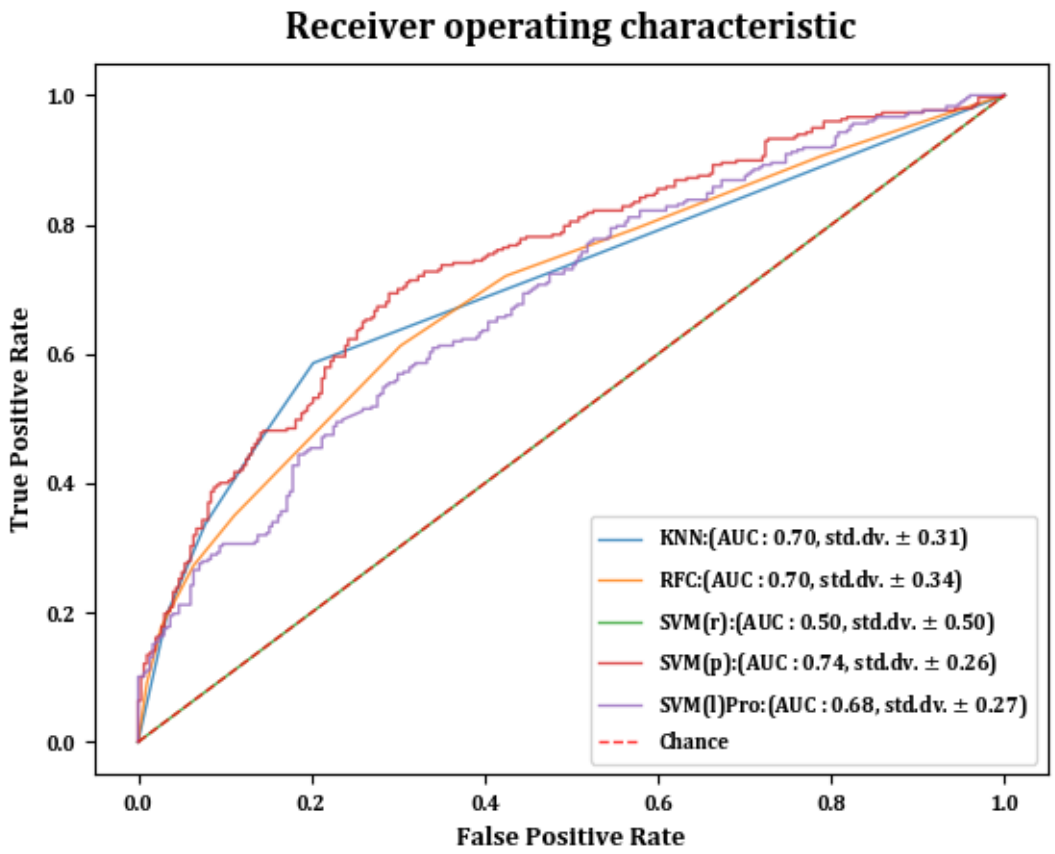


Figure. 17 ROC Curve of Fkmer (Fuse Feature space of 3kmer, 4kmer, and 5kmer) feature space over Layer-1

Figure.18. ROC curve of PseDNC feature space over Layer-1

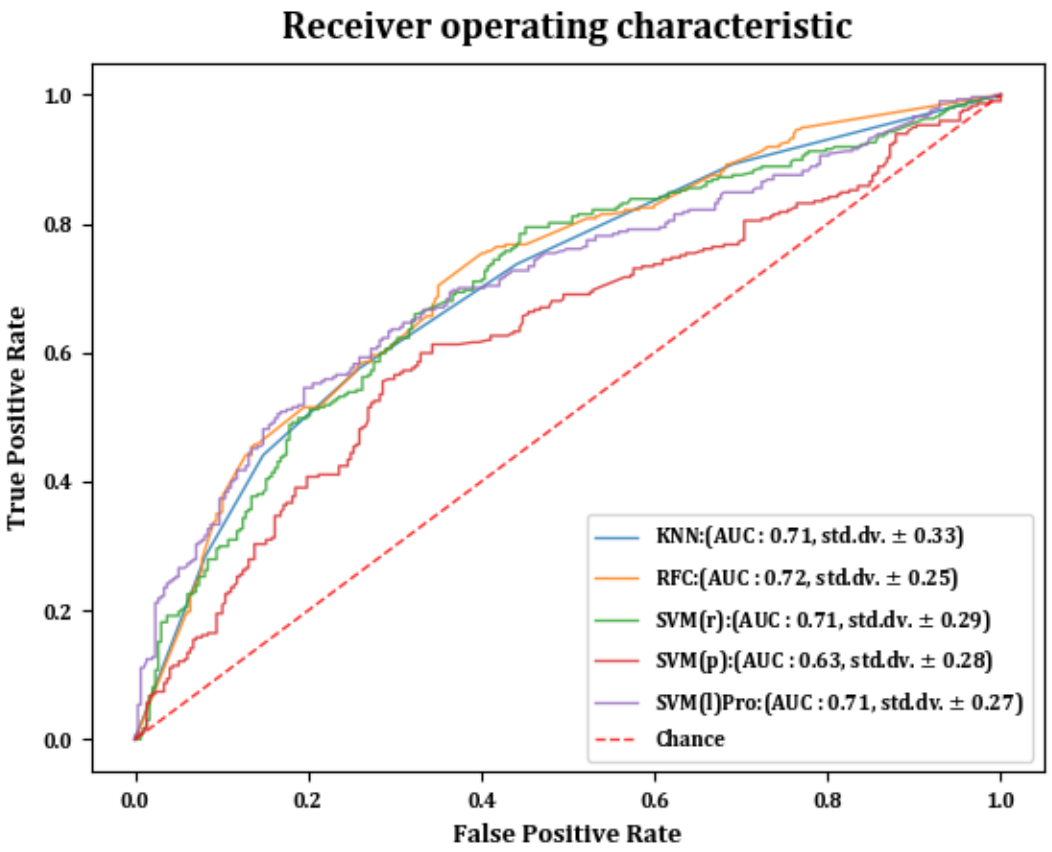


Figure.18. ROC Curve of PseDNC(Pseudo Dinucleotide Composition) feature space

Figure.19. ROC curve of PseTNC feature space over Layer-1

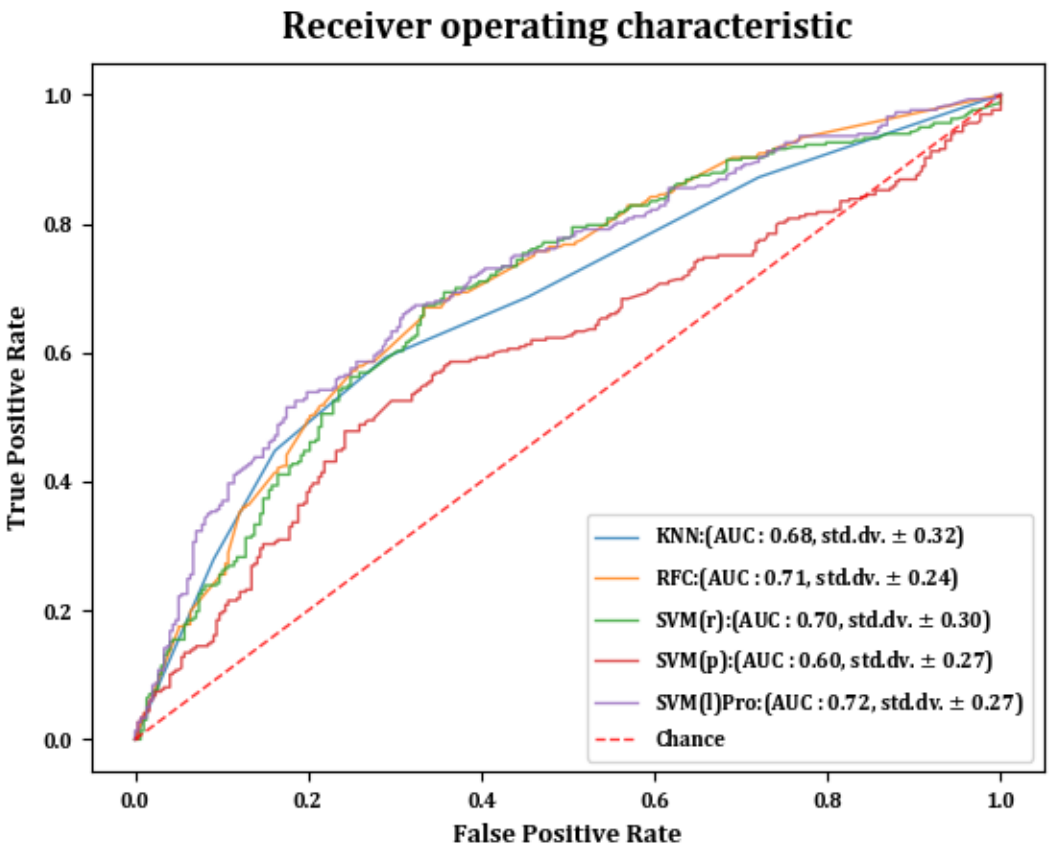


Figure.19. ROC curve of PseTNC feature space over

Layer-2 (Individual Feature) ROC Curves

Figure.20. ROC curve of CKSNAP feature space over Layer-2

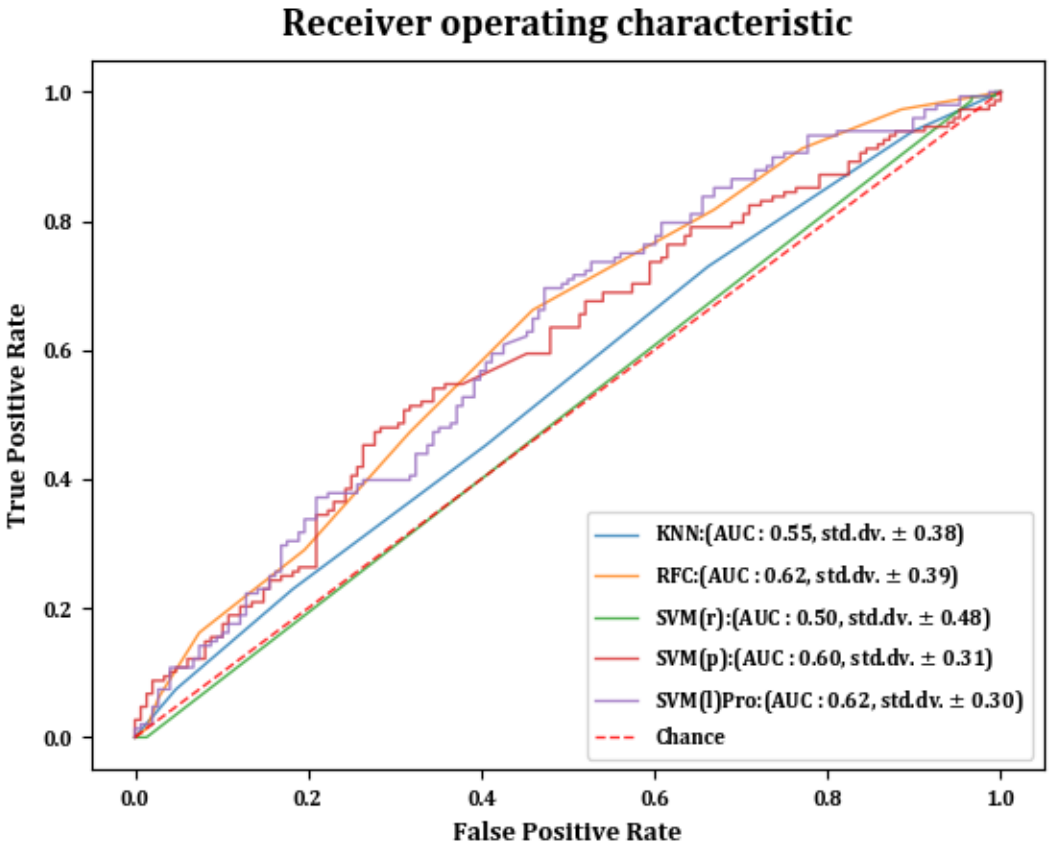


Figure.20. ROC curve of CKSNAP feature space over Layer-2

Figure.21. ROC curve of DCC (Dinucleotide Cross-Correlation) feature space over Layer-2

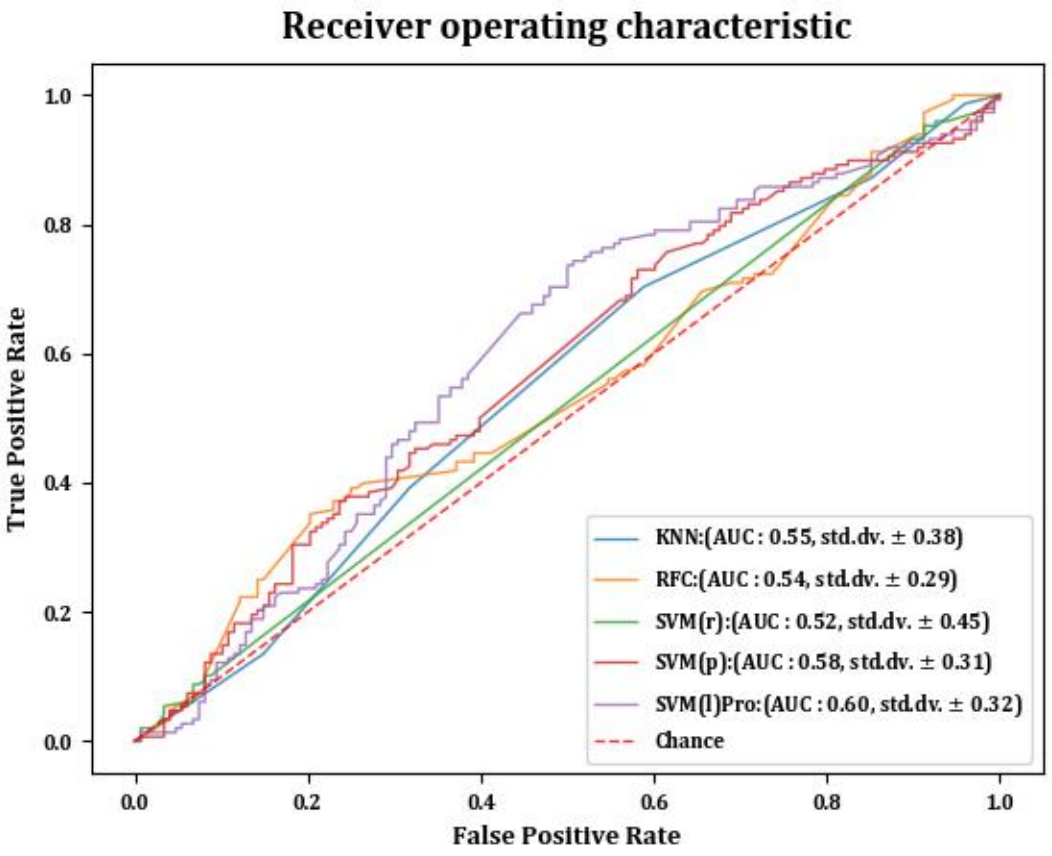


Figure.21. ROC curve of DCC Feature space over Layer-2

Figure.22. ROC curve of Fkmer feature space over Layer-2

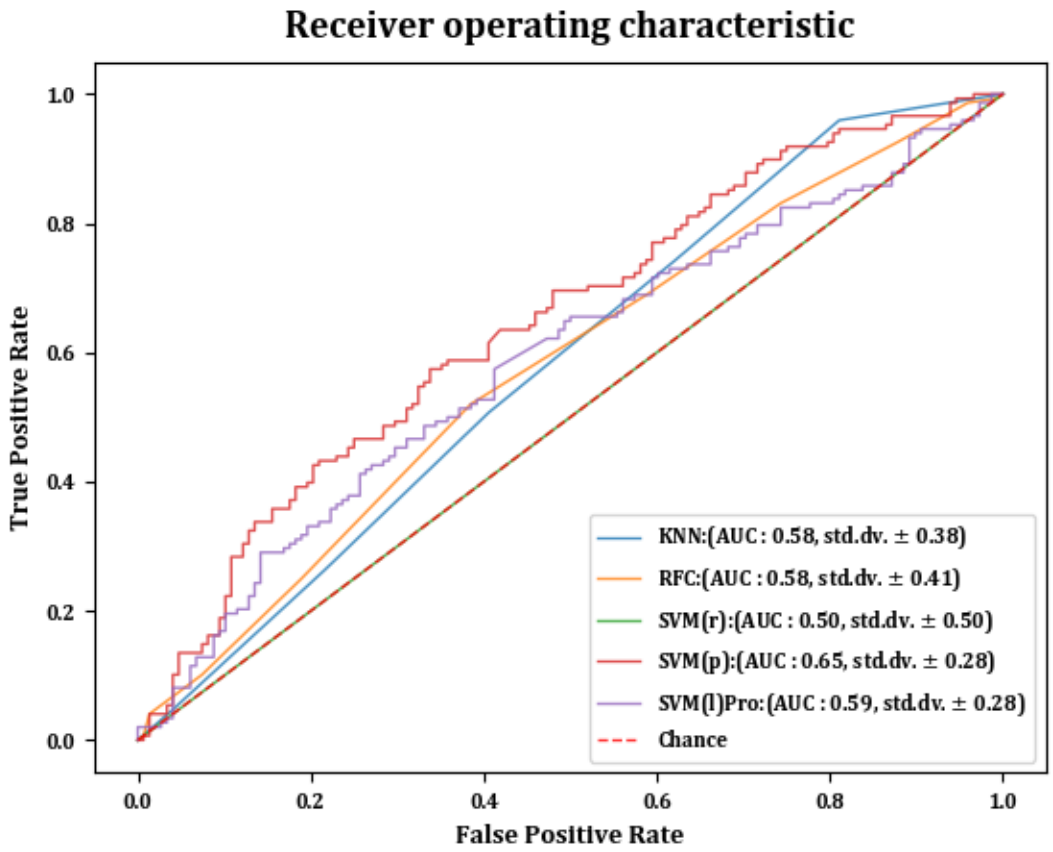


Figure. 22 ROC Curve of Fkmer (Fuse Feature space of 3kmer, 4kmer, and 5kmer) feature space over Layer-2

Figure.29. ROC curve of PseDNC feature space over Layer-2

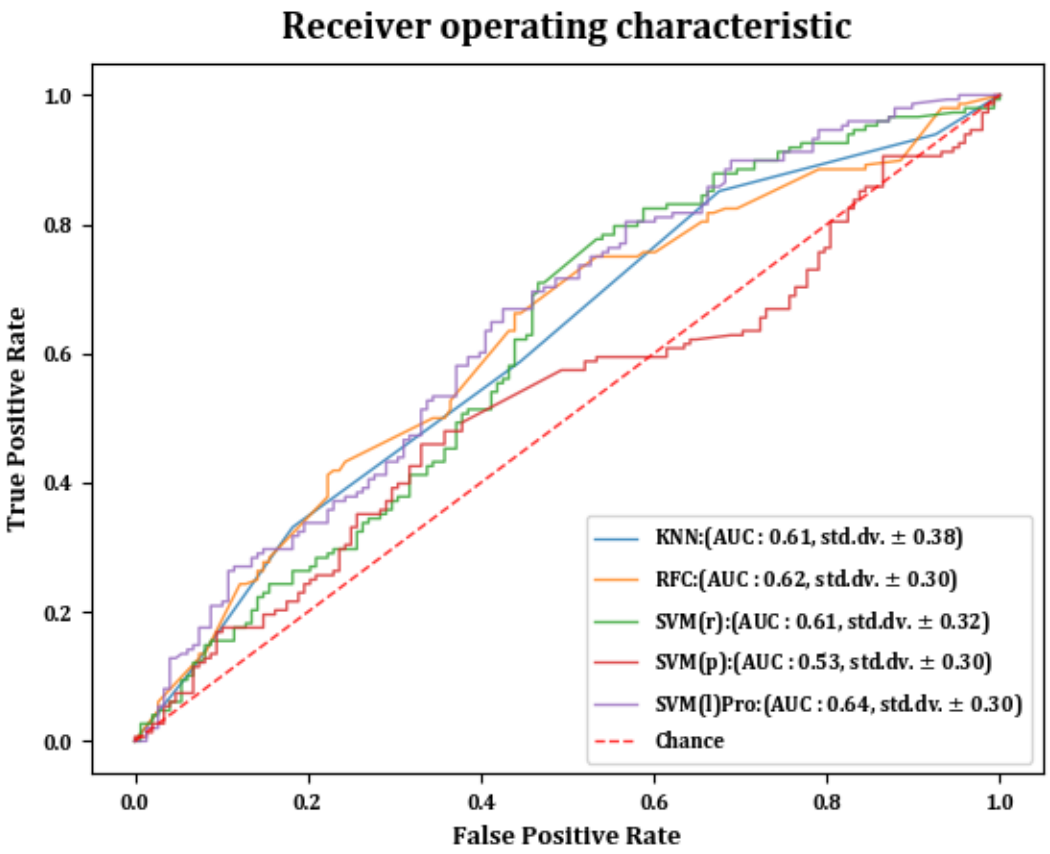


Figure.29. ROC Curve of PseDNC(Pseudo Dinucleotide Composition) feature space over Layer-2

Figure.30. ROC curve of PseTNC feature space over Layer-2

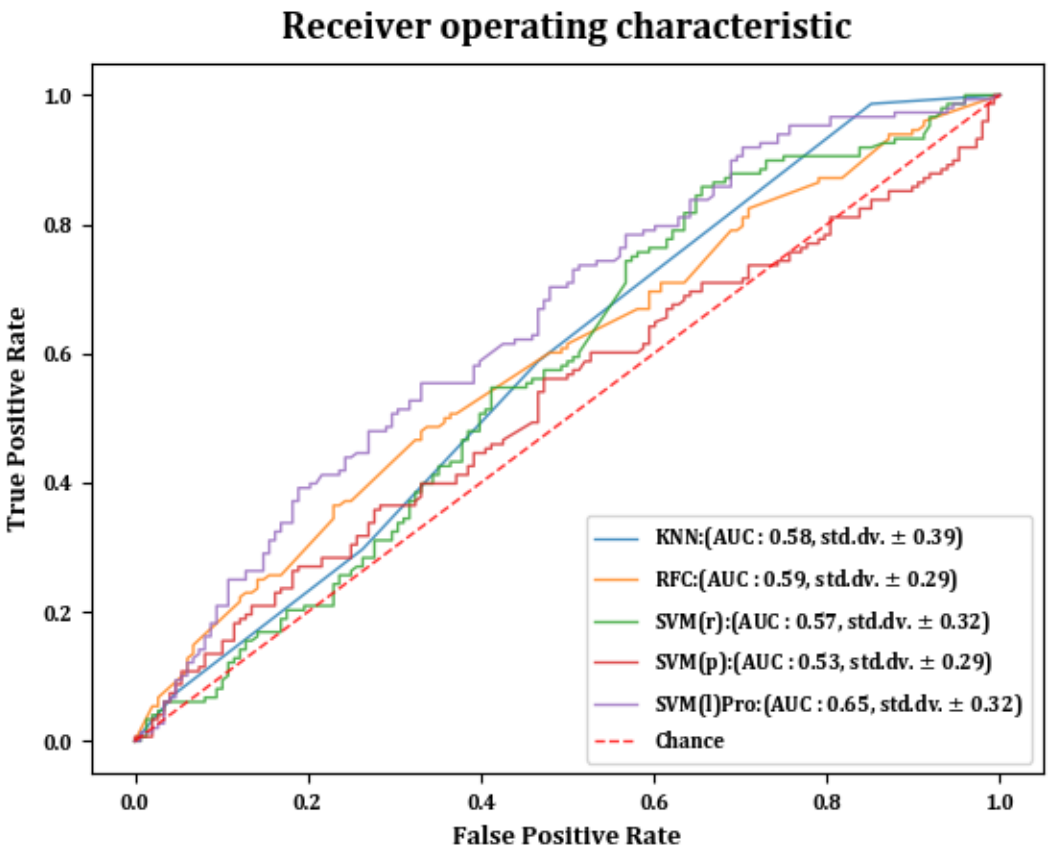


Figure.30. ROC curve of PseTNC feature space over Layer-2

3. Proposed model (baseline SVM(l) linear kernel function) classification over selected feature made a selection via different wrapper classifiers

This section discusses the performance outcomes of the bassline SVM(l) model over Cascade Multi-Level Subset Feature space. These features were drawn and selected via employing different wrapper classification algorithm. These wrappers classifier-based feature simulation result is given as follow.

Result over Cascade Multi-level Feature selection (Wrapper Algorithm: Adaboost)

Layer-1 Classification outcomes

Table.1. Resulting outcomes of Proposed baseline classifier, via Feature selection Cascade Multi-level Subset selection via Adaboost Classifier on Layer-1

Feature:	Clf	AAC	Sn	Sp	Recall	F1	MCC	Kappa	APR	ROC
layer-1-PriCFS.csv	KNN	86.19	91.58	80.80	0.9158	0.869	0.7281	0.7239	0.8921	0.9196
	RFC	84.51	88.55	80.47	0.8855	0.8511	0.6925	0.6902	0.8994	0.9172
	SVM(r)	52.69	5.38	100	0.0539	0.1022	0.1664	0.0539	0.5293	0.5317
	SVM(p)	81.31	88.21	74.41	0.8822	0.8252	0.6323	0.6263	0.8892	0.8923
	SVM(l)Pro	87.87	95.28	80.47	0.9529	0.8871	0.766	0.7576	0.9636	0.9603

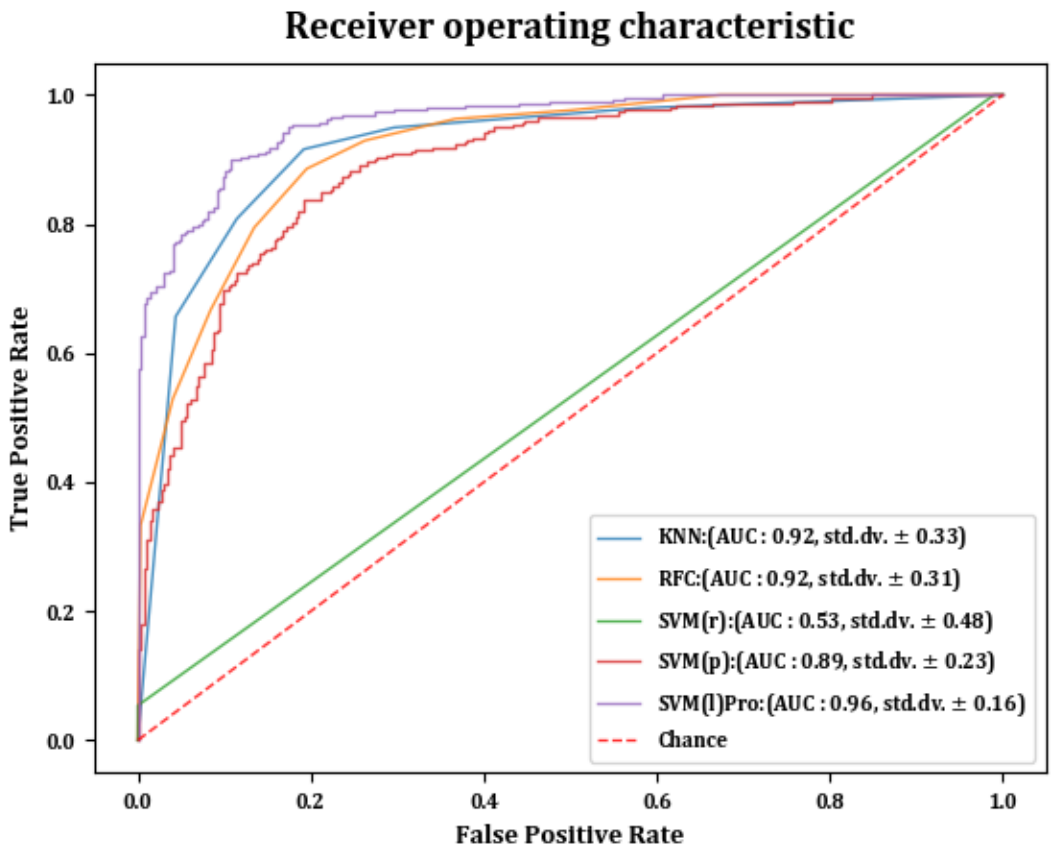


Figure. 31. ROC curve of Propose baseline, method, via Cascade Multi-level Feature selection via Adaboost Classifier on Layer-1

Layer-2 Classification outcomes

Table.2. Resulting outcomes of Proposed baseline classifier, via Feature selection Cascade Multi-level Subset selection via Adaboost Classifier on Layer-2

Feature:	Clf	AAC	Sn	Sp	Recall	F1	MCC	Kappa	ROC
layer-2-PriCFS.csv	KNN	69.257	66.892	71.622	0.6689	0.6851	0.3856	0.3851	0.7025
	RFC	64.865	53.378	76.351	0.5338	0.6031	0.3055	0.2973	0.6591
	SVM(r)	50	0	100	0	0	0	0	0.5
	SVM(p)	66.216	66.216	66.216	0.6622	0.6622	0.3243	0.3243	0.7105
	SVM(l)Pro	68.243	65.541	70.946	0.6554	0.6736	0.3654	0.3649	0.7339

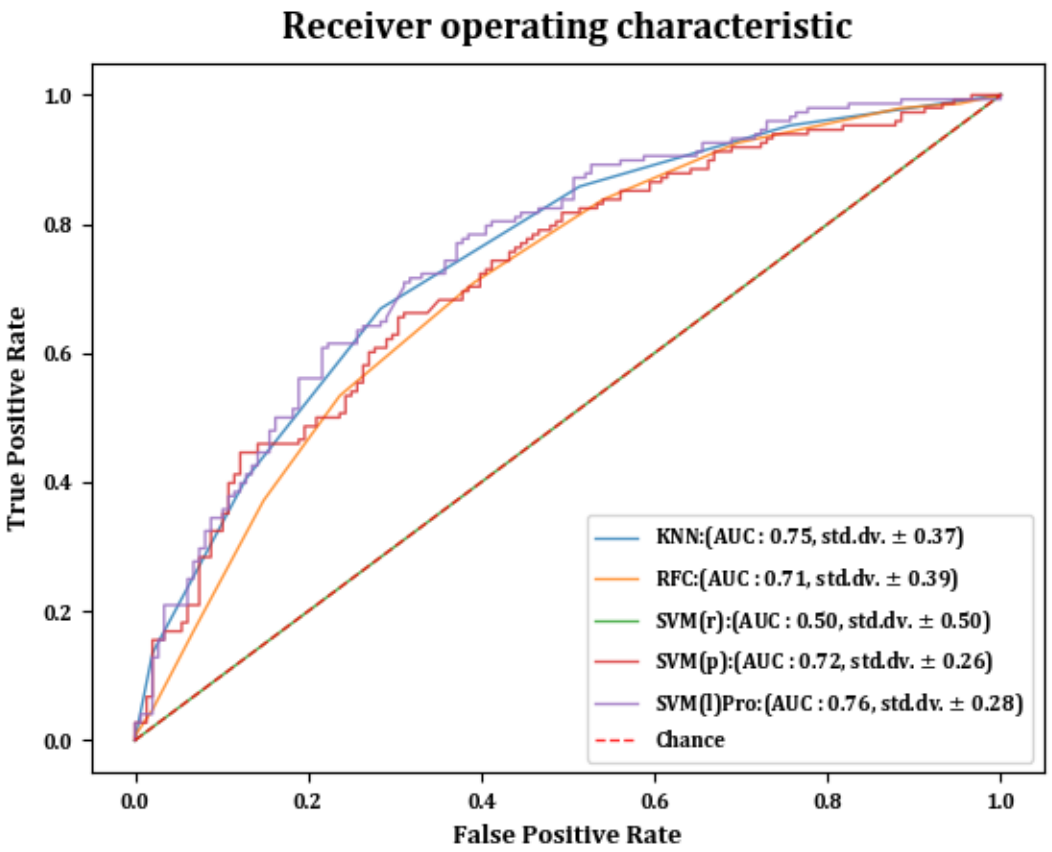


Figure. 32. ROC curve of Propose baseline, method, via Cascade Multi-level Feature selection via Adaboost Classifier on Layer-2

Result over Cascade Multi-level Feature selection (Wrapper Algorithm : Decision Tree)

As we have mentioned in the manuscript, that for implementing feature selection, we have used various wrapper classification technique. These classifiers were used to yield the optimum feature selection and those features were made subject to different values of the threshold for cascade multi-level selection to further bubble out the optimum features.

Classification of the baseline propose model via selected features, whose selection was made via Decision Tree wrapper classifier

Layer-1 Classification outcomes

Table.3. Resulting outcomes of Proposed baseline classifier, via Feature selection Cascade Multi-level Subset selection via Decision Tree Classifier on Layer-1

Feature:	CLF	F	AAC	Sn	Sp	Recall	F1	MCC	Kappa	APR	ROC
layer-1-PriCFS.csv	KNN	1.0	88.2	86.5	89.9	0.865	0.880	0.765	0.764	0.915	0.930
	RFC	1.0	86.4	84.8	87.9	0.849	0.862	0.728	0.727	0.920	0.934
	SVM(r)	1.0	50.0	0.0	100.0	0.000	0.000	0.000	0.000	0.500	0.500
	SVM(p)	1.0	87.2	89.6	84.8	0.896	0.875	0.745	0.744	0.932	0.934
	SVM(l)Pro	1.0	86.5	89.9	83.2	0.899	0.870	0.732	0.731	0.929	0.935

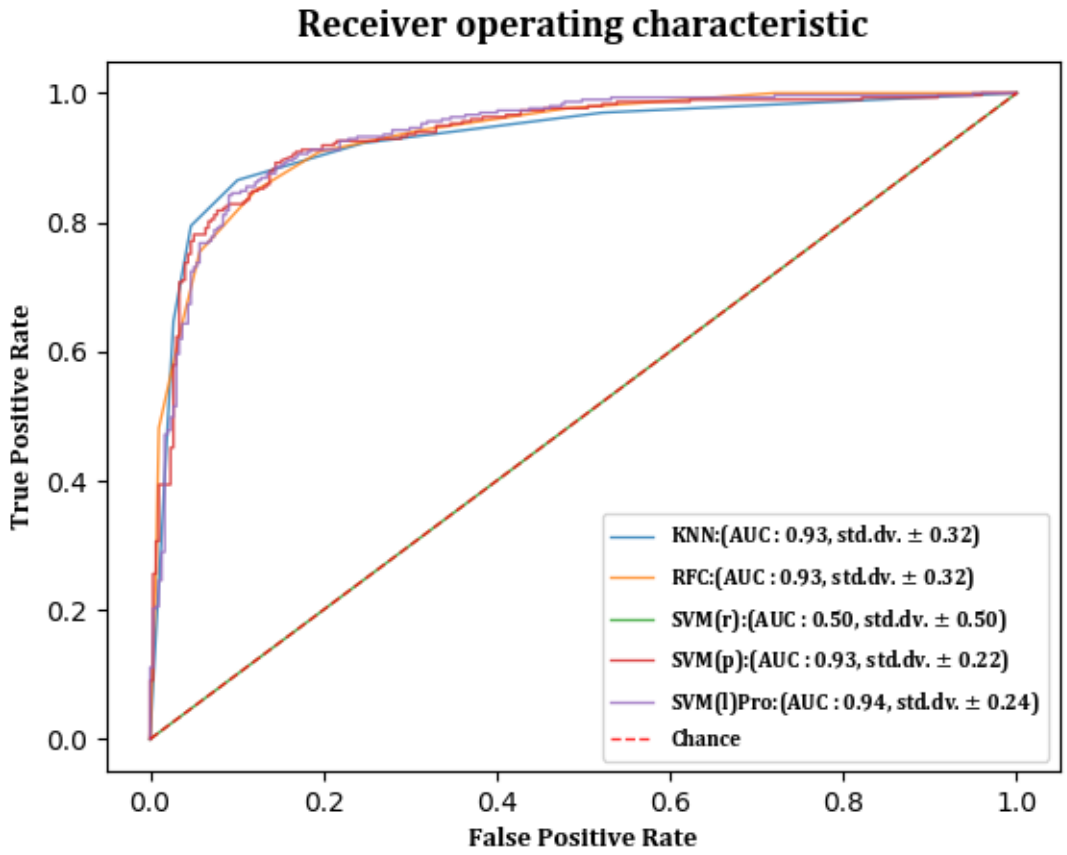


Figure. 33. ROC curve of Propose baseline, method, via Cascade Multi-level Feature selection via Decision Tree Classifier on Layer-1

Layer-2 Classification outcomes

Table.4. Resulting outcomes of Proposed baseline classifier, via Feature selection Cascade Multi-level Subset selection via Decision Tree Classifier on Layer-2

Feature:		Clf	AAC	Sn	Sp	Recall	F1	MCC	Kappa	APR	ROC
layer-2-PriCFS.csv	KNN	3.0	67.6	60.1	75.0	0.601	0.650	0.355	0.351	0.667	0.717
	RFC	3.0	64.2	54.1	74.3	0.541	0.602	0.290	0.284	0.634	0.654
	SVM(r)	3.0	50.0	0.0	100.0	0.000	0.000	0.000	0.000	0.500	0.500
	SVM(p)	3.0	73.0	67.6	78.4	0.676	0.714	0.462	0.460	0.777	0.773
	SVM(l)Pro3	3.0	67.6	61.5	73.6	0.615	0.655	0.354	0.351	0.751	0.740

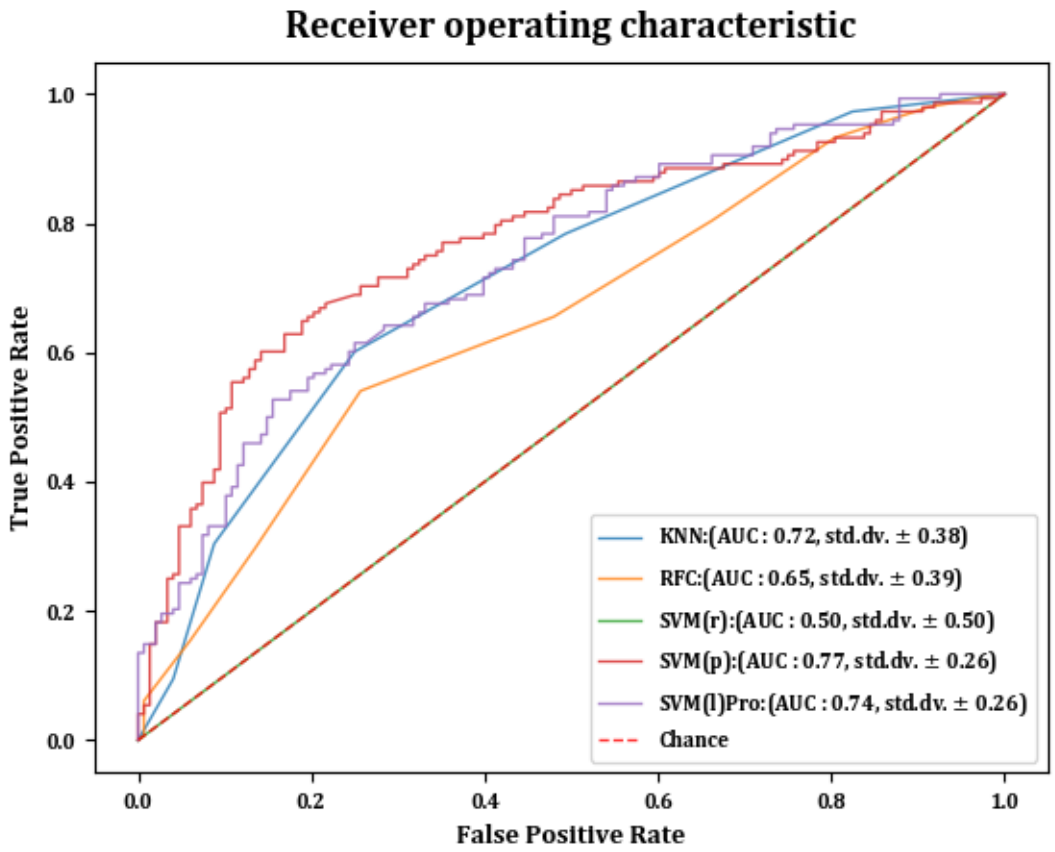


Figure. 34. ROC curve of Propose baseline, method, via Cascade Multi-level Feature selection via Decision Tree Classifier on Layer-2

Result over Cascade Multi-level Feature selection (Wrapper Classifer: Random Forest)

Classification of the baseline propose model via selected features, whose selection was made via Random Forest wrapper classifier

Layer-1 Classification outcomes

Table.5. Resulting outcomes of Proposed baseline classifier, via Feature selection Cascade Multi-level Subset selection via Random Forest Classifier on Layer-1

Feature:		Clf	AAC	Sn	Sp	Recall	F1	MCC	Kappa	APR	ROC
layer-1-PriCFS.csv	KNN	1.0	87.7	87.9	87.5	0.879	0.877	0.754	0.754	0.907	0.930
	RFC	1.0	86.5	85.5	87.5	0.855	0.864	0.731	0.731	0.907	0.924
	SVM(r)	1.0	50.0	0.0	100.0	0.000	0.000	0.000	0.000	0.500	0.500
	SVM(p)	1.0	86.4	89.9	82.8	0.899	0.868	0.729	0.727	0.931	0.938
	SVM(l)Pro1	1.0	86.4	89.2	83.5	0.9	0.867	0.729	0.727	0.926	0.939

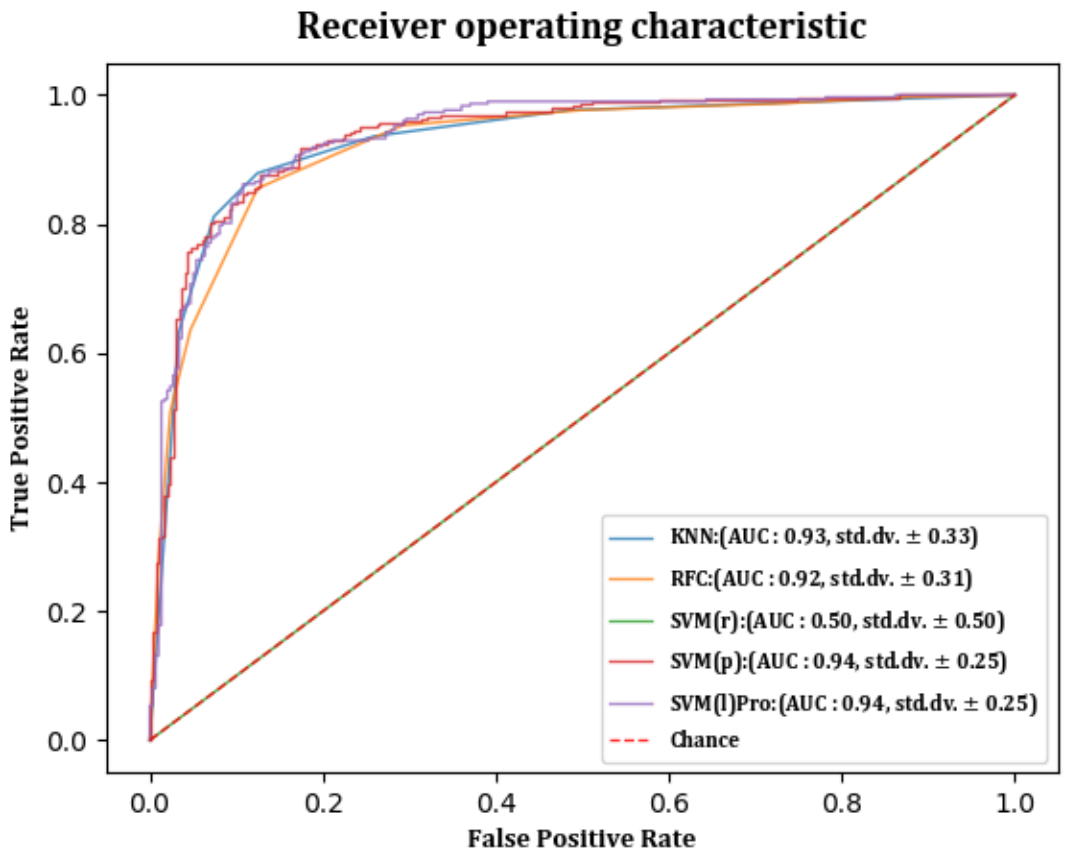


Figure. 35. ROC curve of Propose baseline, method, via Cascade Multi-level Feature selection via Random Forest Classifier on Layer-1

Layer-2 Classification outcomes

Table.6. Resulting outcomes of Proposed baseline classifier, via Feature selection Cascade Multi-level Subset selection via Random Forest Classifier on Layer-6

Feature:	Clf	F	AAC	Sn	Sp	Recall	F1	MCC	Kappa	APR	ROC
layer-2-PriCFS.csv	KNN	1.0	63.4	57.7	69.1	0.577	0.612	0.270	0.269	0.653	0.690
	RFC	1.0	60.4	53.7	67.1	0.537	0.576	0.210	0.208	0.618	0.667
	SVM(r)	1.0	50.0	0.0	100.0	0.000	0.000	0.000	0.000	0.500	0.500
	SVM(p)	1.0	67.1	64.4	69.8	0.644	0.662	0.343	0.342	0.744	0.727
	SVM(l)Pro	1.0	62.8	62.4	63.1	0.6	0.626	0.255	0.255	0.668	0.677

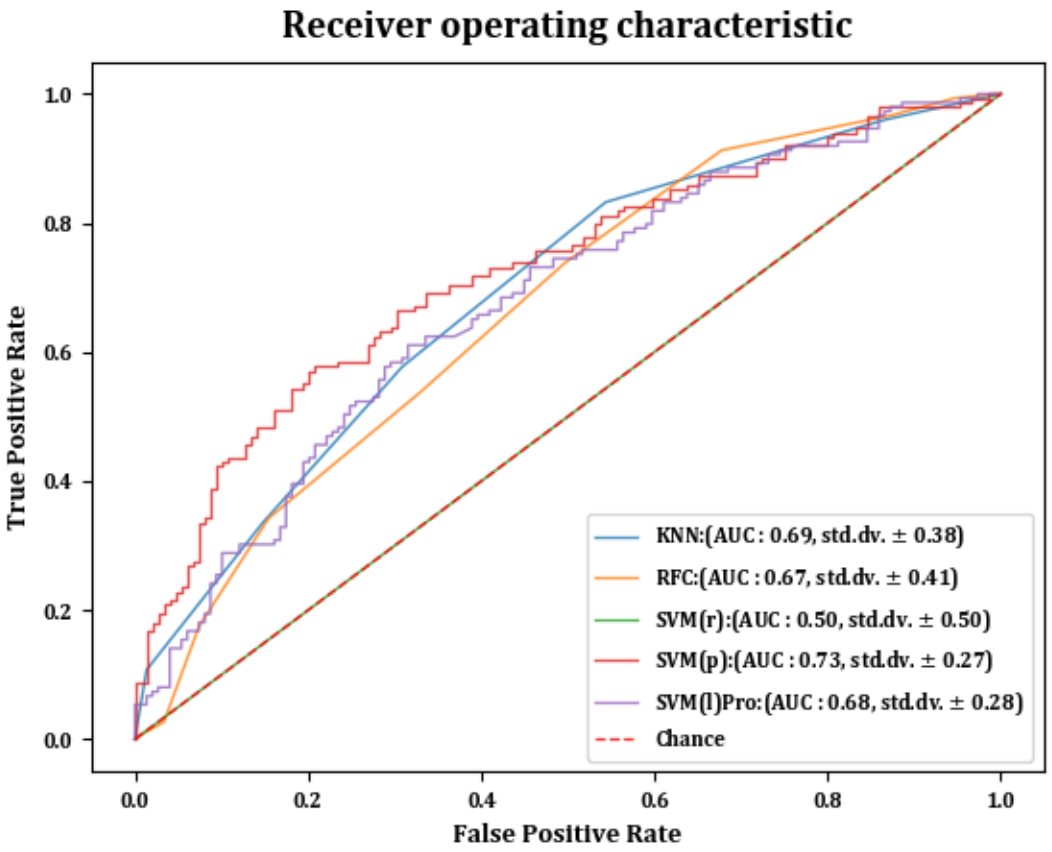


Figure. 36. ROC curve of Propose baseline, method, via Cascade Multi-level Feature selection via Random Forest Classifier on Layer-2

Result over Cascade Multi-level Feature selection (Wrapper Classifier: RFRegressor)
Classification of the baseline propose model via selected features, whose selection was made via RFRegressor wrapper classifier

Layer-1 Classification outcomes
Table.7. Resulting outcomes of Proposed baseline classifier, via Feature selection Cascade Multi-level Subset selection via Random Forest Regressor Classifier on Layer-1

Feature:	S	Clf	Clf	AAC	Sn	Sp	Recall	F1	MCC	Kappa	APR	ROC
layer-1-PriCFS.csv	179	KNN	1	88.9	90.2	87.5	0.9024	0.8904	0.7781	0.7778	0.9086	0.9373
	179	RFC	1	88.6	86.2	90.9	0.8628	0.8828	0.7719	0.7710	0.9317	0.9484
	179	SVM(r)	1	50.0	0.0	100.0	0	0	0	0	0.5	0.5
	179	SVM(p)	1	87.5	90.2	84.8	0.9024	0.8787	0.7519	0.7508	0.9333	0.9407
	179	SVM(l)Pro	1	87.5	91.6	83.5	0.9158	0.8803	0.7533	0.7508	0.9356	0.9448

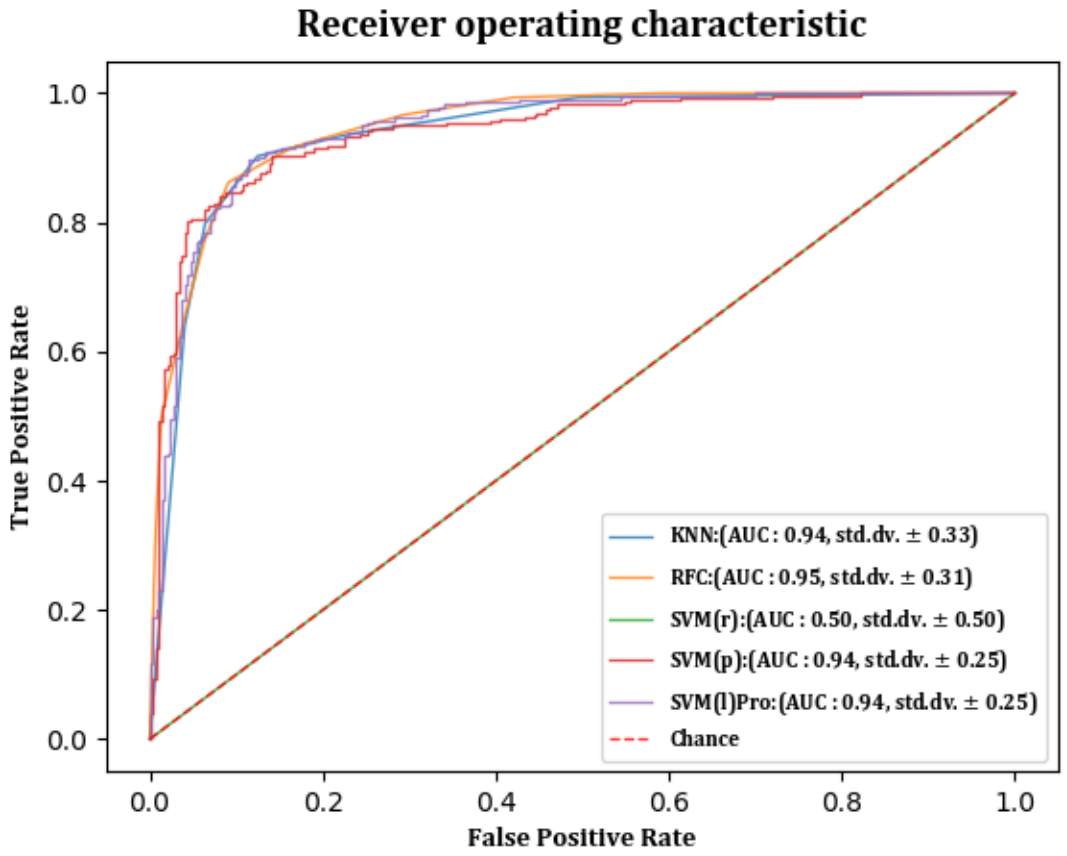


Figure. 37. ROC curve of Propose baseline, method, via Cascade Multi-level Feature selection via Random Forest Regressor Classifier on Layer-1

Layer-2 Classification outcomes

Table.8. Resulting outcomes of Proposed baseline classifier, via Feature selection Cascade Multi-level Subset selection via Random Forest Regressor Classifier on Layer-2

F:	S	CLF	F	AAC	Sn	Sp	Recall	F1	MCC	Kappa	APR	ROC
layer-2-PriCFS.csv	296	KNN	3	67.6	58.8	76.4	0.5878	0.6444	0.3569	0.3514	0.6731	0.7189
	296	RFC	3	65.9	53.4	78.4	0.5338	0.61	0.328	0.3176	0.698	0.723
	296	SVM(r)	3	50.0	0.0	100.0	0	0	0	0	0.5	0.5
	296	SVM(p)	3	73.6	68.9	78.4	0.6892	0.7234	0.4751	0.473	0.7593	0.7853
	296	SVM(l)Pro	3	67.9	61.5	74.3	0.6149	0.657	0.3611	0.3581	0.739	0.7439

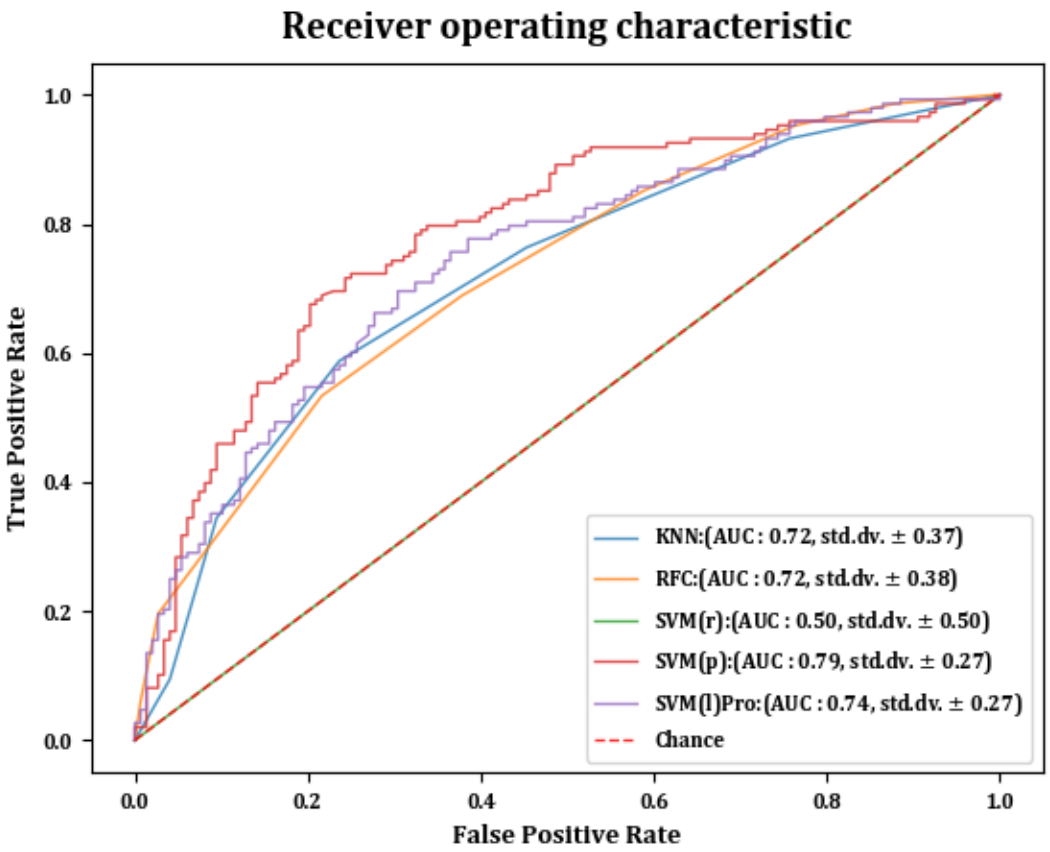


Figure. 38. ROC curve of Propose baseline, method, via Cascade Multi-level Feature selection via Random Forest Regressor Classifier on Layer-2

Result over Cascade Multi-level Feature selection (Wrapper Classifier: XGboost)

Classification of the baseline propose model via selected features, who selection was made via XGboost wrapper classifier

Layer-1 Classification outcomes

Table.9. Resulting outcomes of Proposed baseline classifier, via Feature selection Cascade Multi-level Subset selection via Xgboost Classifier on Layer-1

F	S	CLF	F	AAC	Sn	Sp	Recall	F1	MCC	Kappa	APR	ROC
layer-1-PriCFS.csv	256	KNN	1	87.4	87.5	89.2	0.8754	0.8829	0.7678	0.7677	0.9222	0.9353
	256	RFC	1	85.7	84.2	87.2	0.8418	0.8547	0.7141	0.7138	0.9029	0.9223
	256	SVM(r)	1	50.0	0.0	100.0	0	0	0	0	0.5	0.5
	256	SVM(p)	1	87.0	90.2	83.8	0.9024	0.8744	0.7423	0.7407	0.9517	0.9503
	256	SVM(l)Pro	1	87.7	92.6	82.8	0.9259	0.8828	0.7578	0.7542	0.9552	0.9548

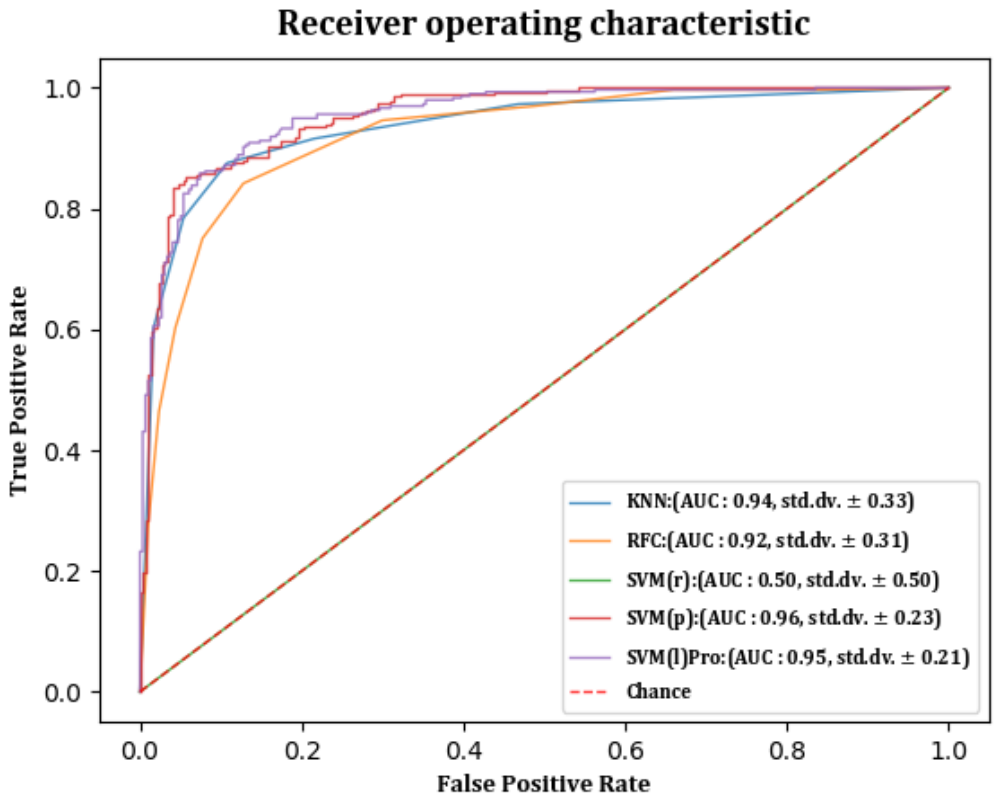


Figure. 39. ROC curve of Propose baseline, method, via Cascade Multi-level Feature selection via XGBoost Classifier on Layer-1

Layer-2 Classification outcomes

Table.10. Resulting outcomes of Proposed baseline classifier, via Feature selection Cascade Multi-level Subset selection via Xgboost Classifier on Layer-2

F	S	CLF	F	AAC	Sn	Sp	Recall	F1	MCC	Kappa	APR	ROC
PriCFS.csv	279	KNN	4	64.9	58.1	71.6	0.5811	0.6232	0.3	0.2973	0.6511	0.6798
	279	RFC	4	62.5	56.8	68.2	0.5676	0.6022	0.2517	0.25	0.6556	0.6957
	279	SVM(r)	4	50.0	0.0	100.0	0	0	0	0	0.5	0.5
	279	SVM(p)	4	65.5	66.2	64.9	0.6622	0.6577	0.3108	0.3108	0.7146	0.7258
	279	SVM(l)Pro	4	68.6	68.9	68.2	0.6892	0.6869	0.3716	0.3716	0.6973	0.7301

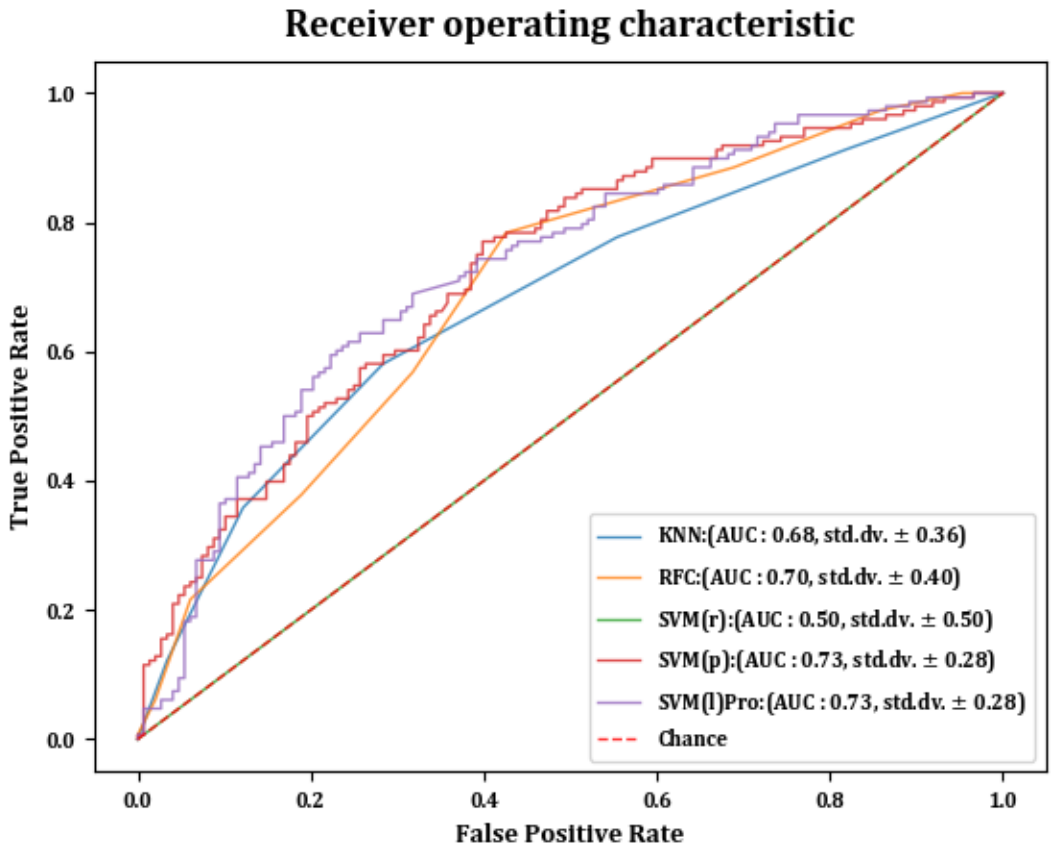


Figure. 40. ROC curve of Propose baseline, method, via Cascade Multi-level Feature selection via XGBoost Classifier on Layer-2

4. Webserver Guide

Many literature studies in the field of computational biology and bioinformatics, indicates the importance and development of a user-friendly publicly accessible web server. Further, a web server simulates intuitions and signifies the importance and future direction for both academians and experimental scientists through carrying various kinds of biological(medical) computational analysis and reporting.

For the ease of end-user and experimental biologist, publicly accessible web server, from where can the end-user can obtain required results without going through technical and mathematical details and accessing a web server. A simple stepwise brief guideline is given as follows.

Step.1. Open Browser by providing the link address http://bienhancer.pythonanywhere.com as shown in Figure.41

Step.2. Type or copy past the Fasta sequence in the Empty area textbox provided that the sequence should be in Fasta format, e.g. see Figure.41

Step.3. By clicking the submit button, the page will be redirected to the prediction result page, as shown in Fig.42.

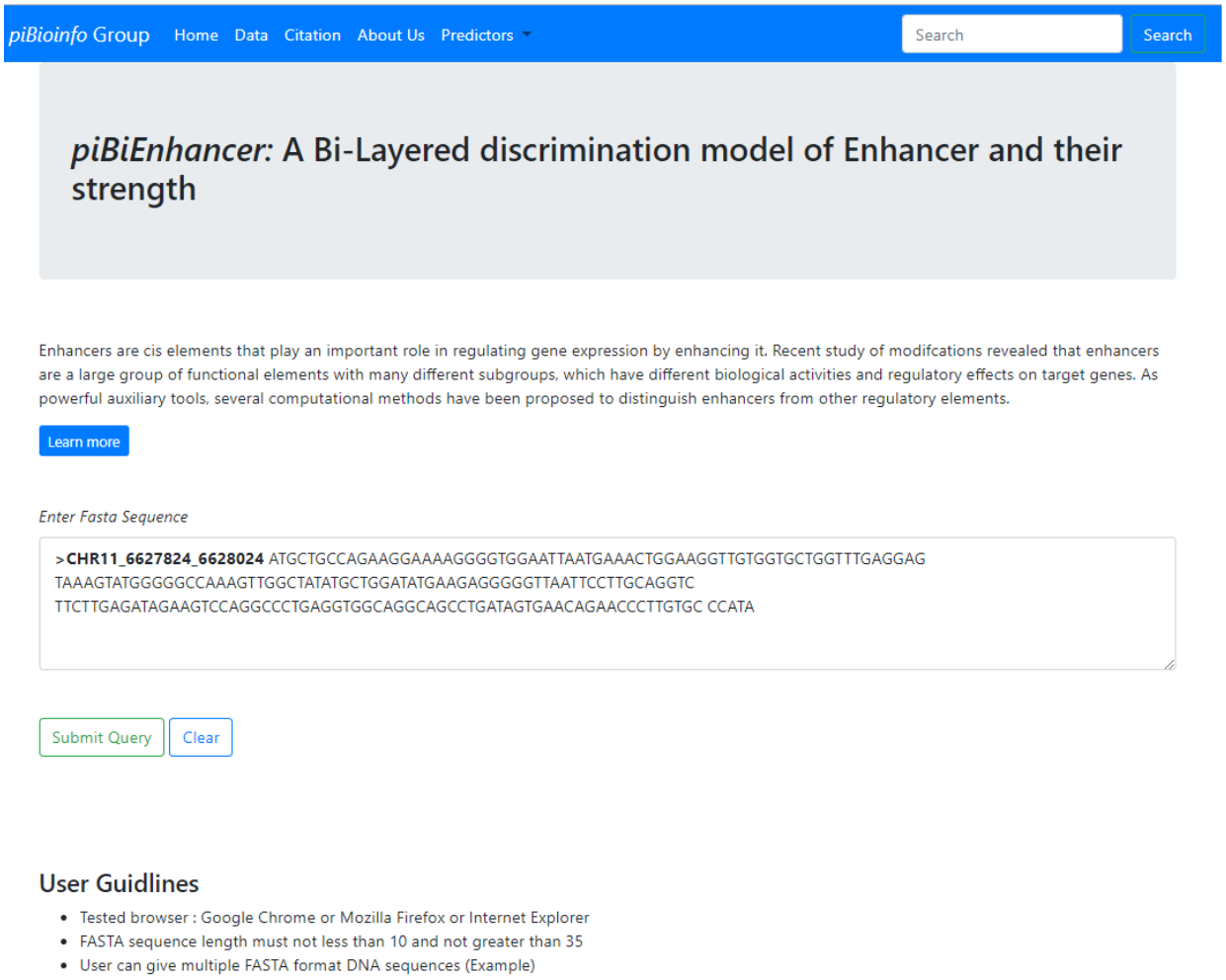


Figure.41. Index Page of webserver Fasta File Entry page

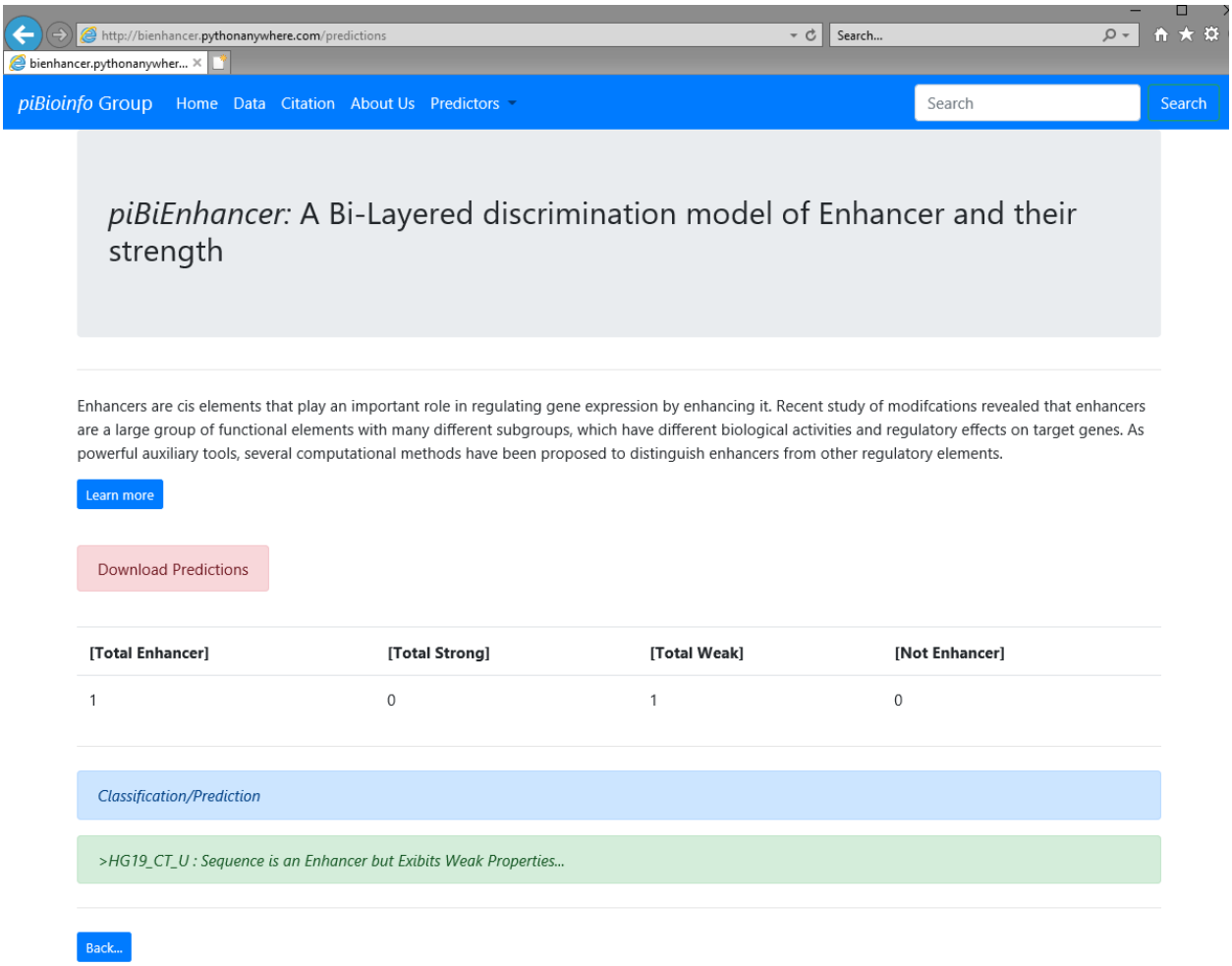


Figure.42. Model prediction over test Fasta Sequence