

Histone-Net: Supplementary Data

Principal Component Analysis:

Principal Component Analysis (PCA) [1] is a most widely used approach in diverse forms of data analysis ranging from computer graphics to neuroscience [2]. It is a very simple and non-parametric approach of acquiring relevant information from highly complex datasets. More specifically, it transforms higher dimensional (n-dimensional) dataset into lower dimensional (k-dimensional) dataset while retaining most of the information [1]. It exploits the correlation among dimensions to come up with least number of variables which can maximally approximate the actual distribution of data [1]. Generally, value of least number of principal components k is determined through empirical evaluation. To illustrate better, consider a hypothetical example in which we want to reduce 100-dimensional data into 50-dimensional data, then PCA performs following steps for dimensionality reduction. First, it computes mean of every data column and centers the value in every column by subtracting the average column value. Then, it computes co-variance matrix of the centered matrix and estimates eigen decomposition of co-variance matrix where eigen vectors represent components for reduced space.

T-distributed Stochastic Neighbor Embedding:

Another extensively used dimensionality reduction approach is T-Distributed Stochastic Neighbor Embedding (t-SNE)[3]. T-SNE is based on a probabilistic paradigm which mainly alleviates the divergence among 2 distributions, one out of which measures pairwise relatedness of input components and other one computes pairwise similarities of respective low-dimensional components of embedding. Unlike PC, T-SNE preserves merely local similarities. It is a non-linear approach which adapts to underlying data to find set of points which can represent higher dimensional data.

We have used three machine learning classifiers (Random Forest, AdaBoost, and Support Vector Machine) to perform extrinsic evaluation of DNA2Vec and SuperDNA2Vec.

Random forest: Random forest [4] classifier is composed of large number of decision trees which operate like an ensemble. For a given instance, every standalone decision tree in random forest infers a corpus class and the corpus class which gets the most votes eventually becomes final prediction of the model. As ensemble usually performs better when underlay decision trees do not mimic highly similar behavior. Therefore, to achieve diverse behaviour of decision trees, random forest makes use of bagging where random forest allows every

standalone decision tree to take random subset of data from the corpus. While in bagging, random forests are constructed through diverse bootstrap samples of data, in boosting, classification trees are developed through splitting every node and using the best predictor among all the predictors randomly chosen at particular node.

AdaBoost: Another classifier used to evaluate the integrity of both distributed representation learning schemes is known as “AdaBoost” [5]. It comes under the hood of meta-classifier which builds a strong classifier through combining the yields of several weak classifiers to a weighted sum. In this manner, AdaBoost manages to correctly classify instances that may be misclassified by standalone weak classifiers.

Support Vector Machine: Discriminative classifier SVM [6] maps each input into a multi-dimensional feature space using a kernel function. SVM makes use of one-against-all or one-against-one approach to perform multi-class classification. In one-against-one approach, $k(k-1)/2$ binary classifiers are trained where each classifier learns to discriminate distinct pair of k classes. For inference, the class with majority vote is chosen as final prediction. Whereas in one-against-all paradigm, multi-class classification problem is transformed into binary classification problem where actual class label is treated as positive and all other class labels are treated as negative to train K binary classifiers. For inference, among all binary classification models, model having highest confidence is used.

4 kinds of Histone proteins

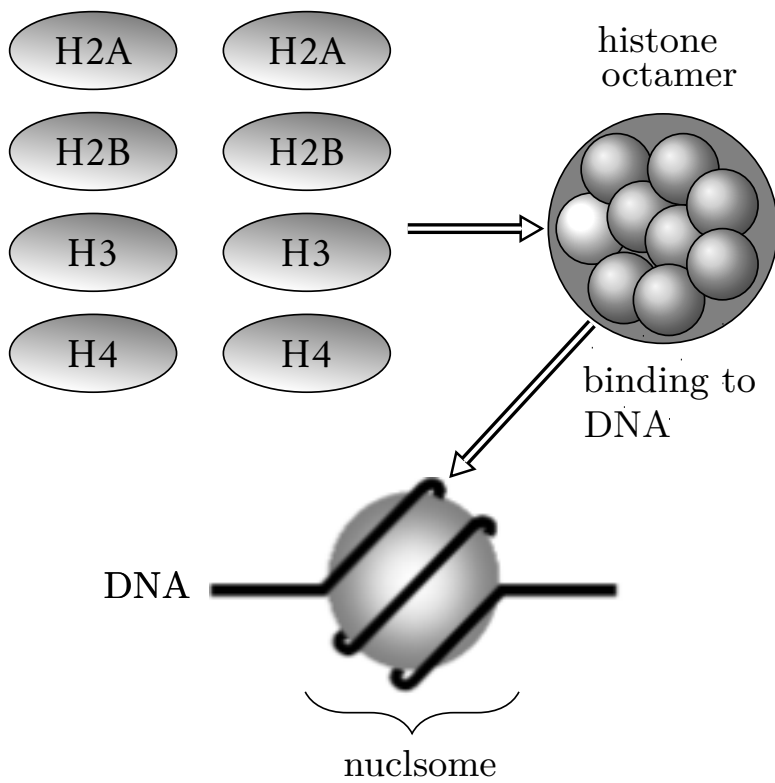


Figure 1: Nucleosome Construction with Histone Octamer Generation and DNA Binding

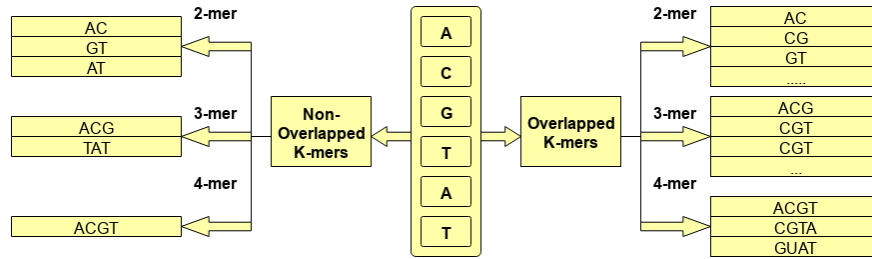


Figure 2: Process of Generating Different Overlapping and Non-Overlapping Higher Order Residues

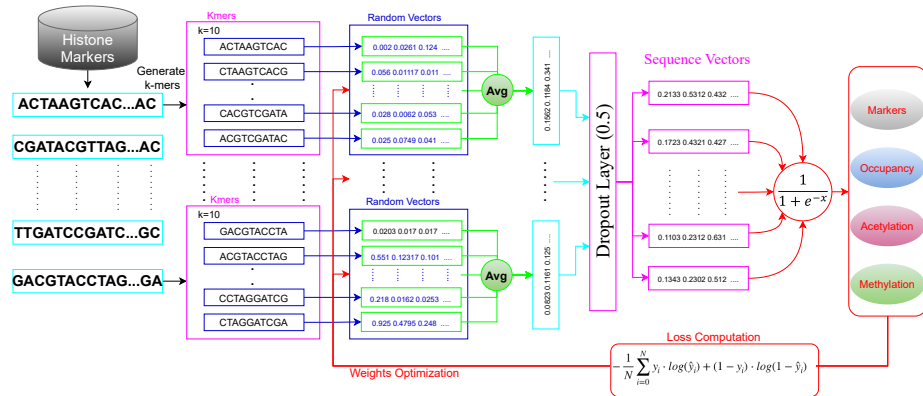


Figure 3: Workflow of Histone-Net Multi-Label Classification Paradigm for Histone Marker, Occupancy, Acetylation, and Methylation Prediction

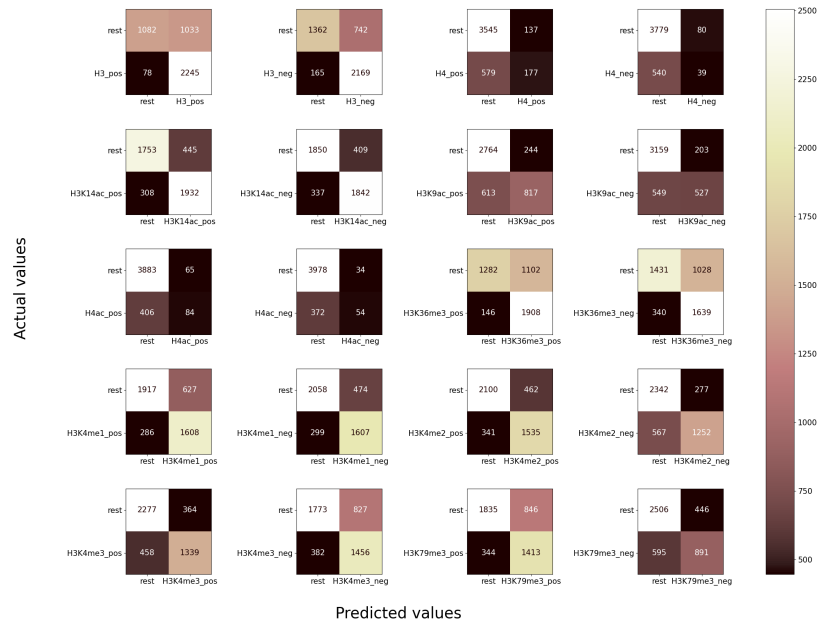


Figure 6: Performance of Histone-Net Approach on balanced dataset where 20 different classes has correct predictions and wrong predictions

DataSets	Kmer	SVM				RFC				Adaboost			
		Accu- racy	Preci- sion	Recall	F1 score	Accu- racy	Preci- sion	Recall	F1 score	Accu- racy	Preci- sion	Recall	F1 score
H3	7	0.6299	0.6299	0.6287	0.6293	0.5671	0.5792	0.5667	0.541	0.6166	0.6162	0.6178	0.6164
	8	0.5778	0.578	0.578	0.5746	0.5404	0.5519	0.5410	0.4939	0.5498	0.549	0.5486	0.5471
	9	0.5391	0.5381	0.539	0.5364	0.5337	0.5476	0.5323	0.4713	0.5431	0.5422	0.5427	0.5406
	10	0.5311	0.5299	0.53	0.5284	0.5043	0.4919	0.5102	0.4362	0.513	0.5116	0.5101	0.511
H4	7	0.6179	0.6191	0.6181	0.6153	0.5337	0.5463	0.5340	0.474	0.5471	0.5467	0.5469	0.547
	8	0.6468	0.6413	0.6471	0.6434	0.5551	0.5168	0.556	0.4201	0.6099	0.6024	0.6112	0.6062
	9	0.5948	0.5855	0.595	0.5863	0.5613	0.6132	0.5621	0.4167	0.5791	0.5689	0.5865	0.5726
	10	0.5674	0.5532	0.5666	0.5495	0.5565	0.5285	0.5570	0.4063	0.5243	0.5062	0.527	0.512
H3K4me1	7	0.5667	0.5521	0.5566	0.5136	0.5551	0.4781	0.556	0.4021	0.5346	0.5067	0.5434	0.5003
	8	0.5921	0.586	0.5942	0.5915	0.5599	0.559	0.5601	0.4269	0.5551	0.5406	0.5561	0.543
	9	0.5568	0.546	0.5571	0.5377	0.5451	0.5229	0.5448	0.3962	0.548	0.5353	0.551	0.5282
	10	0.5559	0.5446	0.5560	0.5319	0.5426	0.4221	0.5431	0.3868	0.5398	0.5223	0.5401	0.5064
H3K4me2	7	0.547	0.5333	0.5468	0.5217	0.5445	0.5017	0.5447	0.3905	0.5432	0.5273	0.5382	0.5116
	8	0.5448	0.5292	0.5451	0.5118	0.5414	0.4281	0.5412	0.3884	0.5322	0.51	0.5319	0.4918
	9	0.5347	0.5161	0.5323	0.5035	0.5464	0.5564	0.547	0.3946	0.53	0.5027	0.5304	0.478
	10	0.594	0.5581	0.5938	0.5095	0.5914	0.2957	0.5923	0.4396	0.5816	0.5286	0.5821	0.5081
H3K4me3	7	0.5839	0.5101	0.584	0.4709	0.5911	0.2956	0.5901	0.4394	0.5917	0.5505	0.6011	0.4944
	8	0.5865	0.5081	0.5856	0.4599	0.5911	0.2956	0.5875	0.4394	0.5833	0.5039	0.5763	0.4678
	9	0.5852	0.5115	0.5832	0.4675	0.5911	0.2956	0.5901	0.4394	0.5888	0.5329	0.5847	0.4728
	10	0.5557	0.5504	0.5561	0.5444	0.5345	0.5618	0.535	0.3819	0.5277	0.5176	0.5267	0.5087
H3K79me3	7	0.5236	0.5142	0.5231	0.5098	0.5326	0.5165	0.5331	0.3801	0.5416	0.5337	0.5425	0.5152
	8	0.5323	0.524	0.5312	0.5186	0.5315	0.4899	0.5321	0.3777	0.5228	0.51	0.5235	0.4966
	9	0.516	0.5061	0.514	0.5034	0.5318	0.5041	0.5306	0.3833	0.5288	0.5172	0.5291	0.5008
	10	0.531	0.522	0.523	0.5151	0.531	0.482	0.5301	0.3784	0.5277	0.517	0.5267	0.506
H3K36me3	7	0.5922	0.594	0.5865	0.5444	0.5949	0.5444	0.412	0.5832	0.5802	0.5832	0.5771	0.5771
	8	0.5655	0.5614	0.5645	0.5567	0.5281	0.4702	0.5311	0.3823	0.5572	0.5524	0.5582	0.5451
	9	0.5565	0.5516	0.556	0.5478	0.5326	0.5245	0.5336	0.3885	0.5454	0.539	0.5463	0.5305
	10	0.5485	0.5427	0.5475	0.5366	0.5302	0.496	0.532	0.3856	0.5277	0.5172	0.5283	0.5048
H4ac	7	0.5402	0.5335	0.5387	0.5287	0.5326	0.5233	0.5336	0.3918	0.5218	0.5102	0.5225	0.4995
	8	0.5837	0.5785	0.5841	0.5728	0.5413	0.5133	0.5401	0.3884	0.5516	0.5422	0.5545	0.5378
	9	0.5456	0.5348	0.5460	0.5291	0.5401	0.4662	0.5387	0.3844	0.545	0.5331	0.5441	0.5219
	10	0.5467	0.5351	0.5472	0.5225	0.5399	0.458	0.5345	0.3843	0.529	0.5098	0.5301	0.4939
H3K14ac	7	0.5378	0.5236	0.5412	0.5117	0.5413	0.5128	0.5421	0.3879	0.5344	0.5151	0.534	0.4892
	8	0.5376	0.5234	0.538	0.5122	0.541	0.4958	0.5412	0.3848	0.5381	0.5215	0.5367	0.498
	9	0.5437	0.5337	0.5441	0.5287	0.5393	0.5007	0.5401	0.3845	0.539	0.525	0.541	0.5056
	10	0.5387	0.5268	0.5371	0.5182	0.5408	0.5407	0.5414	0.3892	0.5229	0.5004	0.5231	0.4809
H3K9ac	7	0.5302	0.5156	0.531	0.5062	0.5367	0.4609	0.5381	0.3862	0.5264	0.5052	0.5282	0.4835
	8	0.5217	0.5068	0.5221	0.502	0.5416	0.5706	0.5432	0.3886	0.534	0.5153	0.5332	0.4847
	9	0.534	0.5207	0.532	0.5118	0.5378	0.4362	0.5391	0.3812	0.5331	0.5186	0.5327	0.5067
	10	0.5691	0.5454	0.5701	0.5336	0.5679	0.5341	0.5681	0.4141	0.5607	0.5334	0.5624	0.5253
H3K9ac	7	0.5722	0.5486	0.5732	0.5228	0.5679	0.534	0.5681	0.4125	0.5637	0.5332	0.5645	0.5087
	8	0.5707	0.5456	0.571	0.5157	0.5682	0.6174	0.5687	0.4127	0.5564	0.5122	0.5549	0.4806
	9	0.5622	0.5279	0.561	0.4966	0.5676	0.2839	0.568	0.4113	0.5592	0.521	0.5586	0.4909
	10	0.5601	0.5236	0.5604	0.494	0.5676	0.2839	0.5663	0.4113	0.5558	0.5105	0.5533	0.4793
H3K9ac	7	0.561	0.5446	0.559	0.5311	0.5524	0.3923	0.5516	0.3968	0.5491	0.5293	0.5501	0.5212
	8	0.5502	0.5297	0.5499	0.518	0.5542	0.5047	0.5532	0.402	0.5419	0.5166	0.5426	0.5038
	9	0.5452	0.5204	0.5467	0.5044	0.5556	0.5694	0.5561	0.4008	0.5282	0.4912	0.5352	0.4787
	10	0.5376	0.5069	0.5329	0.4906	0.5542	0.4997	0.5552	0.4008	0.5437	0.5139	0.5461	0.4888
11	0.5747	0.5631	0.5751	0.5449	0.557	0.6219	0.5569	0.404	0.5506	0.5285	0.5542	0.5106	

Table 1: Performance figures of three machine learning classifiers using DNA2Vec unsupervised DNA sequence vectors

Datasets	K-mers	SVM				RFC				Adabost				SoftMax			
		Accuracy	Precision	Recall	F1 score	Accuracy	Precision	Recall	F1 score	Accuracy	Precision	Recall	F1 score	Accuracy	Precision	Recall	F1 score
H3	7	0.865	0.8649	0.8637	0.865	0.8616	0.8615	0.8626	0.8616	0.8596	0.8596	0.8576	0.8596	0.863	0.8629	0.863	0.863
	8	0.8864	0.8869	0.8854	0.8863	0.891	0.8912	0.8901	0.891	0.8824	0.8827	0.8814	0.8823	0.885	0.8857	0.885	0.8849
	9	0.8917	0.8926	0.8867	0.8916	0.8937	0.8937	0.8927	0.8937	0.8683	0.8682	0.8673	0.8683	0.8951	0.8952	0.8951	0.895
	10	0.8971	0.8973	0.8981	0.897	0.8884	0.8884	0.8798	0.8884	0.871	0.8709	0.869	0.871	0.8991	0.899	0.8991	0.8991
	11	0.8817	0.8819	0.8717	0.8817	0.875	0.8749	0.873	0.875	0.8556	0.8556	0.8506	0.8556	0.8777	0.8787	0.8777	0.8776
H4	7	0.8712	0.8691	0.8732	0.8714	0.8726	0.8706	0.8716	0.8727	0.8603	0.8582	0.8601	0.8604	0.8726	0.8706	0.8726	0.8729
	8	0.8815	0.8799	0.8825	0.8815	0.8815	0.8797	0.8805	0.8816	0.8788	0.8767	0.8708	0.8789	0.8836	0.8818	0.8836	0.8836
	9	0.8945	0.8928	0.8925	0.8946	0.8918	0.8902	0.8921	0.8918	0.887	0.8865	0.8871	0.8868	0.8911	0.8893	0.8911	0.8912
	10	0.8877	0.8856	0.8817	0.8878	0.8815	0.8795	0.8785	0.8816	0.8664	0.8643	0.8662	0.8666	0.8863	0.8843	0.8863	0.8865
	11	0.8781	0.8763	0.8771	0.8782	0.8753	0.8732	0.8749	0.8755	0.8493	0.8482	0.8491	0.8491	0.8719	0.8725	0.8719	0.8715
H3K4me1	7	0.6932	0.6905	0.6992	0.6925	0.6866	0.6838	0.6966	0.6861	0.6926	0.6899	0.6956	0.6924	0.6847	0.6832	0.6847	0.6819
	8	0.7342	0.7322	0.7322	0.7338	0.7311	0.729	0.7211	0.7306	0.7263	0.7243	0.7258	0.7257	0.7304	0.7283	0.7304	0.73
	9	0.7863	0.786	0.7843	0.7855	0.785	0.7833	0.7795	0.785	0.7661	0.7643	0.7601	0.7663	0.7885	0.7871	0.7885	0.7882
	10	0.816	0.815	0.815	0.8157	0.8144	0.8128	0.8104	0.8145	0.7841	0.7824	0.7851	0.784	0.8198	0.8193	0.8198	0.8194
	11	0.8207	0.8198	0.8205	0.8204	0.822	0.8204	0.852	0.8221	0.7951	0.7939	0.7953	0.7955	0.8251	0.8252	0.8251	0.8246
H3K4me2	7	0.689	0.6781	0.685	0.6832	0.6851	0.6733	0.6841	0.6807	0.6812	0.6689	0.6832	0.6776	0.6861	0.6756	0.6861	0.6788
	8	0.7451	0.7372	0.7441	0.743	0.736	0.7275	0.735	0.7338	0.7347	0.7281	0.7357	0.7304	0.7405	0.7316	0.7405	0.7396
	9	0.7947	0.7947	0.7927	0.7912	0.7891	0.7847	0.7879	0.7872	0.7761	0.7716	0.7757	0.7736	0.7966	0.7916	0.7966	0.7953
	10	0.8256	0.8229	0.8216	0.8243	0.822	0.8182	0.825	0.821	0.8025	0.801	0.8031	0.8	0.8276	0.8267	0.8276	0.8258
	11	0.8331	0.8297	0.8321	0.8322	0.8256	0.8209	0.8251	0.825	0.8116	0.8077	0.8106	0.8103	0.8312	0.8283	0.8312	0.83
H3K4me3	7	0.6821	0.6807	0.6723	0.6812	0.6712	0.6696	0.6722	0.6705	0.6647	0.663	0.6707	0.6644	0.6731	0.6751	0.6731	0.6688
	8	0.7549	0.7542	0.7489	0.7545	0.7484	0.7476	0.7464	0.748	0.7351	0.735	0.7348	0.734	0.7527	0.7518	0.7527	0.7525
	9	0.8321	0.8318	0.8291	0.8319	0.8171	0.8165	0.8161	0.817	0.8111	0.8104	0.8121	0.8111	0.8329	0.8322	0.8329	0.8329
	10	0.8731	0.8724	0.8721	0.8731	0.8693	0.8686	0.8703	0.8693	0.8484	0.8476	0.8474	0.8484	0.8731	0.8727	0.8731	0.8731
	11	0.8717	0.871	0.8712	0.8718	0.8726	0.8719	0.8723	0.8727	0.8568	0.8561	0.8558	0.8569	0.8726	0.8723	0.8726	0.8725
H3K79me3	7	0.8124	0.8121	0.8124	0.8122	0.8051	0.8047	0.8052	0.8049	0.8037	0.8031	0.8029	0.8037	0.8089	0.8085	0.8089	0.8088
	8	0.837	0.8367	0.835	0.8369	0.8386	0.8385	0.8376	0.8384	0.8317	0.831	0.8327	0.8317	0.8384	0.8379	0.8384	0.8384
	9	0.8748	0.8742	0.8688	0.8749	0.8707	0.8704	0.8705	0.8706	0.861	0.8603	0.8601	0.861	0.8748	0.8755	0.8748	0.8746
	10	0.8984	0.8979	0.8972	0.8984	0.8963	0.8959	0.8953	0.8963	0.8894	0.8893	0.8893	0.8893	0.8991	0.8999	0.8991	0.899
	11	0.9026	0.9023	0.9016	0.9026	0.897	0.8965	0.8971	0.897	0.8752	0.875	0.8762	0.8751	0.9022	0.9027	0.9022	0.9021
H3K36me3	7	0.7477	0.7464	0.7527	0.7471	0.7411	0.7396	0.7421	0.7406	0.7408	0.7392	0.7418	0.7405	0.7471	0.7479	0.7471	0.7454
	8	0.7953	0.794	0.7893	0.7952	0.7864	0.7857	0.7864	0.7859	0.7867	0.7853	0.7857	0.7866	0.7921	0.7914	0.7921	0.7917
	9	0.8435	0.8427	0.8445	0.8433	0.8412	0.8404	0.8432	0.8411	0.8265	0.8278	0.8259	0.8258	0.8446	0.8436	0.8446	0.8446
	10	0.8759	0.8754	0.8769	0.8758	0.8724	0.8718	0.8704	0.8724	0.8501	0.8499	0.8511	0.8498	0.8744	0.8744	0.8744	0.8743
	11	0.8724	0.8723	0.8714	0.8723	0.8727	0.8718	0.8717	0.8727	0.8564	0.8552	0.8554	0.8565	0.8756	0.8765	0.8756	0.8753
H4ac	7	0.7125	0.7106	0.7135	0.7123	0.7096	0.7077	0.7106	0.7093	0.7061	0.7041	0.7063	0.706	0.7064	0.7045	0.7064	0.7058
	8	0.7776	0.7762	0.7759	0.7776	0.7738	0.7724	0.7731	0.7738	0.7697	0.7683	0.7689	0.7697	0.7759	0.7748	0.7759	0.7755
	9	0.8416	0.8405	0.8396	0.8416	0.8398	0.8388	0.8391	0.8398	0.8181	0.8169	0.8171	0.8183	0.8486	0.8479	0.8486	0.8486
	10	0.8744	0.8735	0.8684	0.8745	0.8648	0.8638	0.8638	0.8648	0.856	0.8549	0.8516	0.8561	0.8759	0.8757	0.8759	0.8758
	11	0.875	0.874	0.872	0.8751	0.873	0.8719	0.874	0.8731	0.8548	0.8539	0.8538	0.855	0.8747	0.8739	0.8747	0.8747
H3K14ac	7	0.7256	0.7209	0.7376	0.724	0.7183	0.7129	0.7203	0.7175	0.7228	0.7176	0.7218	0.7221	0.7207	0.7155	0.7207	0.7208
	8	0.7909	0.7883	0.7879	0.79	0.7882	0.7852	0.7812	0.7874	0.7776	0.7748	0.7786	0.7765	0.7873	0.7833	0.7873	0.7873
	9	0.8563	0.8545	0.8613	0.856	0.8475	0.8464	0.8465	0.847	0.8275	0.8254	0.8205	0.8271	0.8545	0.8531	0.8545	0.8541
	10	0.8856	0.8843	0.8806	0.8855	0.8829	0.8809	0.8819	0.8829	0.8663	0.8637	0.8643	0.8663	0.8865	0.8885	0.8865	0.8859
	11	0.895	0.8946	0.891	0.8948	0.8923	0.8911	0.8932	0.8921	0.8805	0.8788	0.8815	0.8804	0.8962	0.8985	0.8962	0.8956
H3K9ac	7	0.7469	0.7442	0.7389	0.7463	0.7473	0.7444	0.7503	0.7468	0.7358	0.7326	0.7308	0.7355	0.7455	0.7425	0.7455	0.7451
	8	0.7793	0.7773	0.7783	0.7788	0.7693	0.7671	0.7683	0.7686	0.7711	0.7705	0.7751	0.7696	0.7775	0.7751	0.7775	0.7772
	9	0.8182	0.8163	0.8122	0.818	0.8125	0.8111	0.8124	0.812	0.7995	0.7971	0.7975	0.7995	0.8182	0.8182	0.8182	0.8174
	10	0.8359	0.8353	0.8319	0.8354	0.8312	0.829	0.8332	0.8312	0.829	0.8268	0.8219	0.8294	0.8384	0.8365	0.8384	0.8384
	11	0.8402	0.8396	0.8392	0.8397	0.8445	0.8426	0.8435	0.8445	0.8182	0.8159	0.8187	0.8185	0.8359	0.8359	0.8359	0.8352

Table 2: Performance figures of three machine learning classifiers and SoftMax using SuperDNA2Vec Supervised DNA sequence vectors

Histone Occupancy and Modification Predictor	Histone Markers									
	H3	H4	H3K4me1	H3K4me2	H3K4me3	H3K79me3	H3K36me3	H4ac	H3K14ac	H3K9ac
	Precision									
HCNN [7]	0.8733	0.8700	0.7367	0.7400	0.7467	0.8267	0.7767	0.7833	0.7500	0.7900
DeepHistone [8]	0.8689	0.8952	0.6927	0.6334	0.6523	0.8104	0.7604	0.7141	0.7296	0.7424
Histone-Net	0.8952	0.8893	0.8252	0.8283	0.8723	0.9027	0.8765	0.8739	0.8985	0.8365
	Recall									
HCNN [7]	0.8733	0.8700	0.7300	0.7233	0.7467	0.8233	0.7667	0.7767	0.7367	0.7900
DeepHistone [8]	0.8697	0.8979	0.6944	0.6496	0.6533	0.8111	0.7609	0.7152	0.7334	0.7433
Histone-Net	0.8951	0.8911	0.8251	0.8312	0.8726	0.9022	0.8756	0.8747	0.8962	0.8384

Table 3: Performance comparison of proposed Histone-Net with state-of-the-art histone occupancy and modification predictor HCNN [7] in terms of precision and recall. Precision and recall values of HCNN are taken from Table 3 of Yin et al. [7]

Multi-Label DNA Sequence Analysis Datasets	K-mers	Accuracy	Precision	Recall	F1	Average Precision	Coverage	One Error	Ranking Loss	Hamming Loss
Imbalance Dataset	7	0.475	0.5275	0.7656	0.6146	0.529	15.0033	0.7701	0.495	0.3269
	8	0.5129	0.5802	0.7503	0.6451	0.529	15.223	0.7981	0.4404	0.2772
	9	0.5535	0.6406	0.7384	0.6787	0.529	15.683	0.8371	0.3966	0.2291
	10	0.5714	0.6882	0.7257	0.693	0.529	15.8996	0.8556	0.3812	0.2049
	11	0.57	0.6752	0.7363	0.6921	0.529	15.6819	0.8517	0.3811	0.2109
Balanced Dataset	7	0.5486	0.5979	0.8086	0.676	0.5712	14.8986	0.7929	0.4305	0.2818
	8	0.5783	0.6413	0.7838	0.6973	0.5712	15.2021	0.7902	0.3843	0.2443
	9	0.6099	0.6963	0.7644	0.7217	0.5712	15.6519	0.8215	0.3489	0.2068
	10	0.6171	0.7113	0.7652	0.7299	0.5712	15.7452	0.8324	0.3412	0.1977
	11	0.6078	0.7112	0.7432	0.7187	0.5712	15.4712	0.8175	0.3529	0.2005

Table 4: Performance Produced by Histone-Net using Different Higher Order Residues based SuperDNA2Vec Sequence Representation over Imbalanced and Balanced Version of Multi-Label Dataset

References

- [1] Ian Jolliffe. Principal component analysis. *Encyclopedia of statistics in behavioral science*, 2005.
- [2] Basna Mohammed Salih Hasan and Adnan Mohsin Abdulazeez. A review of principal component analysis algorithm for dimensionality reduction. *Journal of Soft Computing and Data Mining*, 2(1):20–30, 2021.
- [3] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008.
- [4] Yanjun Qi. Random forest for bioinformatics. In *Ensemble machine learning*, pages 307–323. Springer, 2012.
- [5] Elham Pashaei, Mustafa Ozen, and Nizamettin Aydin. Splice sites prediction of human genome using adaboost. In *2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, pages 300–303. IEEE, 2016.
- [6] Chenghai Xue, Fei Li, Tao He, Guo-Ping Liu, Yanda Li, and Xuegong Zhang. Classification of real and pseudo microRNA precursors using local structure-sequence features and support vector machine. *BMC bioinformatics*, 6(1):310, 2005.
- [7] Bojian Yin, Marleen Balvert, Davide Zambrano, Alexander Schönhuth, and Sander Bohte. An image representation based convolutional network for dna classification. *arXiv preprint arXiv:1806.04931*, 2018.
- [8] Qijin Yin, Mengmeng Wu, Qiao Liu, Hairong Lv, and Rui Jiang. Deephi-stone: a deep learning approach to predicting histone modifications. *BMC genomics*, 20(2):193, 2019.