A Comparative Analysis of Computational Drug Repurposing Approaches – Proposing a Novel Tensor-Matrix-Tensor Factorization Method: *supplementary material*

by A. Zabihian, J. Asghari, M. Hooshmand, S. Gharaghani

The document contains supplementary information related to the paper of the same title. Section 1 details an exhaustive search through a subset of the hyperparameter space of the proposed ML algorithms. Furthermore, section 2 provides excessive information on the evaluation of several proposed and SOTA methods on a subset of DTINet dataset.

1. Machine learning hyperparameter optimization

Table 1 reports the different values of the parameters tested for each of the machine learning models during the process of hyperparameter optimization.

Table 1: ML parameters examined throughout the grid search

ML Algorithm	Parameters	Parameters' Values
Random Forest	Criterion	gini, entropy, log loss
	No. Estimators	10, 100, 1000
	Max No. Features	$\sqrt{n}, log_2 n$
SVM	Kernel	Linear, poly, RBF
	C	0.1, 1, 10

Tables 2, 3, 4, and 5 detail the performance results of the SVM model regarding different configurations of its hyperparameters on $1:1$, $1:3$, $1:5$, and $1:10$ samples of the DTINet dataset [1], respectively.

Tables 6, 7, 8, and 9 evaluate the performance of the RF model concerning different configurations of its hyperparameters on $1:1$, $1:3$, $1:5$, and $1:10$ samples of the DTINet dataset, respectively.

2. Extra results on the DTINet dataset

This section provides an evaluation on the performance of the proposed machine/deep/graph learning methods along with SOTA ones on three positive:negative samples — $1:1$, $1:3$, and $1:10$ — of the DTINet dataset in Tables 10, 11, 12 respectively.

References

[1] Y. Luo, X. Zhao, J. Zhou, J. Yang, Y. Zhang, W. Kuang, J. Peng, L. Chen, J. Zeng, A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information, Nature communications 8 (1) (2017) 1–13.

[2] F. Wan, L. Hong, A. Xiao, T. Jiang, J. Zeng, Neodti: neural integration of neighbor information from a heterogeneous network for discovering new drug–target interactions, Bioinformatics 35 (1) (2019) 104–111.

Table 2: Performance evaluation of the different configurations of the SVM model on a 1 : 1 sample of the DTINet dataset

Kernel	C	Sp	Re (Se)	Pr	Acc	F1	MCC	AUC-ROC	AUPR
Linear	0.1	0.840±0.021	0.854±0.021	0.842±0.019	0.847±0.018	0.848±0.018	0.694±0.036	0.901±0.010	0.884±0.017
	1	0.787±0.031	0.854±0.032	0.801±0.022	0.821±0.019	0.826±0.019	0.643±0.038	0.881±0.014	0.859±0.018
	10	0.767±0.023	0.835±0.022	0.782±0.017	0.801±0.015	0.807±0.015	0.603±0.030	0.859±0.015	0.835±0.020
Poly	0.1	**0.908±0.016**	0.734±0.028	**0.890±0.016**	0.821±0.013	0.804±0.017	0.653±0.024	0.892±0.011	0.893±0.013
	1	0.854±0.012	0.848±0.019	0.854±0.010	0.851±0.009	0.850±0.010	0.702±0.018	0.917±0.008	0.917±0.009
	10	0.833±0.019	**0.877±0.023**	0.840±0.015	0.855±0.013	0.858±0.014	0.711±0.027	0.915±0.010	0.910±0.014
RBF	0.1	0.858±0.029	0.566±0.038	0.801±0.033	0.712±0.021	0.662±0.029	0.444±0.043	0.786±0.023	0.822±0.021
	1	0.907±0.012	0.735±0.024	0.888±0.013	0.821±0.013	0.804±0.017	0.652±0.026	0.895±0.013	0.906±0.014
	10	0.858±0.013	0.875±0.020	0.860±0.011	**0.866±0.010**	**0.867±0.011**	**0.733±0.020**	**0.931±0.008**	**0.931±0.013**

Table 3: Performance evaluation of the different configurations of the SVM model on a 1 : 3 sample of the DTINet dataset

Kernel	C	Sp	Re (Se)	Pr	Acc	F1	MCC	AUC-ROC	AUPR
Linear	0.1	0.865±0.014	0.858±0.033	0.679±0.022	0.863±0.011	0.758±0.019	0.673±0.027	0.912±0.015	0.733±0.039
	1	0.860±0.018	0.841±0.020	0.668±0.029	0.855±0.014	0.744±0.021	0.654±0.030	0.903±0.014	0.699±0.034
	10	0.861±0.017	0.814±0.025	0.662±0.024	0.849±0.011	0.730±0.015	0.634±0.020	0.883±0.013	0.656±0.034
Poly	0.1	0.908±0.022	0.781±0.025	0.741±0.047	0.876±0.018	0.760±0.028	0.678±0.040	0.911±0.015	0.793±0.042
	1	0.899±0.012	0.848±0.027	0.737±0.025	0.886±0.011	0.788±0.020	0.714±0.027	0.932±0.015	0.831±0.036
	10	**0.917±0.009**	0.833±0.032	0.771±0.023	0.896±0.012	0.800±0.024	0.732±0.032	0.935±0.015	0.858±0.040
RBF	0.1	0.917±0.013	0.595±0.029	0.706±0.037	0.836±0.013	0.645±0.028	0.544±0.036	0.837±0.016	0.738±0.033
	1	0.906±0.022	0.778±0.028	0.737±0.049	0.874±0.019	0.756±0.032	0.673±0.045	0.915±0.018	0.821±0.044
	10	0.917±0.015	**0.862±0.032**	**0.776±0.033**	**0.903±0.014**	**0.816±0.026**	**0.753±0.034**	**0.945±0.014**	**0.865±0.035**

[3] J. Soh, S. Park, H. Lee, Hidti: integration of heterogeneous information to predict drug-target interactions, Scientific reports 12 (1) (2022) 1–12.

[4] J. Li, J. Wang, H. Lv, Z. Zhang, Z. Wang, IMCHGAN: Inductive matrix completion with heterogeneous graph attention networks for Drug-Target interactions prediction, IEEE/ACM Trans Comput Biol Bioinform 19 (2) (2022) 655–665.

Table 4: Performance evaluation of the different configurations of the SVM model on a 1 : 5 sample of the DTINet dataset

Kernel	C	Sp	Re (Se)	Pr	Acc	F1	MCC	AUC-ROC	AUPR
Linear	0.1	0.873±0.013	**0.858±0.025**	0.575±0.026	0.870±0.011	0.688±0.021	0.630±0.026	0.917±0.009	0.648±0.034
	1	0.885±0.011	0.832±0.031	0.591±0.020	0.876±0.008	0.691±0.016	0.630±0.021	0.910±0.007	0.618±0.018
	10	0.893±0.009	0.794±0.032	0.597±0.020	0.876±0.008	0.681±0.020	0.616±0.025	0.898±0.007	0.591±0.019
Poly	0.1	0.910±0.010	0.795±0.029	0.638±0.025	0.890±0.009	0.707±0.022	0.647±0.027	0.919±0.013	0.736±0.035
	1	0.914±0.009	0.849±0.028	0.664±0.022	0.903±0.007	0.745±0.017	0.694±0.021	0.938±0.011	0.789±0.031
	10	**0.947±0.008**	0.818±0.026	**0.757±0.029**	**0.926±0.008**	**0.786±0.022**	**0.742±0.026**	0.944±0.011	**0.839±0.032**
RBF	0.1	0.937±0.009	0.616±0.035	0.663±0.032	0.883±0.009	0.638±0.026	0.569±0.031	0.865±0.016	0.704±0.022
	1	0.911±0.012	0.800±0.029	0.643±0.031	0.892±0.011	0.713±0.024	0.654±0.030	0.926±0.012	0.771±0.035
	10	0.934±0.008	0.855±0.024	0.723±0.024	0.921±0.005	0.783±0.012	0.739±0.015	**0.948±0.012**	0.835±0.037

Table 5: Performance evaluation of the different configurations of the SVM model on a 1 : 10 sample of the DTINet dataset

Kernel	C	Sp	Re (Se)	Pr	Acc	F1	MCC	AUC-ROC	AUPR
Linear	0.1	0.880±0.007	**0.854±0.029**	0.417±0.011	0.878±0.005	0.560±0.011	0.542±0.014	0.922±0.009	0.495±0.025
	1	0.897±0.007	0.820±0.028	0.445±0.014	0.890±0.006	0.576±0.013	0.552±0.015	0.915±0.009	0.461±0.025
	10	0.907±0.007	0.775±0.020	0.457±0.015	0.895±0.005	0.574±0.013	0.543±0.014	0.904±0.009	0.438±0.025
Poly	0.1	0.911±0.008	0.822±0.024	0.480±0.021	0.902±0.007	0.605±0.018	0.580±0.020	0.929±0.010	0.633±0.024
	1	0.931±0.006	0.843±0.028	0.551±0.021	0.923±0.006	0.666±0.020	0.643±0.022	0.944±0.009	0.706±0.034
	10	**0.970±0.004**	0.788±0.025	**0.726±0.023**	**0.954±0.003**	**0.755±0.014**	**0.731±0.016**	0.951±0.008	**0.813±0.027**
RBF	0.1	0.936±0.007	0.673±0.026	0.515±0.026	0.912±0.006	0.583±0.019	0.542±0.021	0.891±0.016	0.631±0.031
	1	0.916±0.007	0.827±0.025	0.497±0.017	0.908±0.005	0.620±0.012	0.596±0.013	0.938±0.011	0.688±0.038
	10	0.953±0.004	0.837±0.025	0.643±0.020	0.943±0.004	0.727±0.015	0.703±0.017	**0.955±0.009**	0.781±0.033

Table 6: Performance evaluation of the different configurations of the RF model on a $1:1$ sample of the DTINet dataset

Criterion	No. Estimators	Max No. Features	Sp	Re (Se)	Pr	Acc	F1	MCC	AUC-ROC	AUPR
gini	10	$log_2 n$	0.881±0.025	0.786±0.027	0.870±0.024	0.834±0.014	0.825±0.016	0.671±0.029	0.903±0.009	0.903±0.010
		$\sqrt{n}$	0.892±0.023	0.792±0.030	0.880±0.024	0.842±0.021	0.833±0.023	0.688±0.041	0.908±0.017	0.903±0.017
	100	$log_2 n$	0.885±0.030	0.816±0.019	0.877±0.030	0.851±0.020	0.846±0.020	0.704±0.041	0.921±0.011	0.930±0.013
		$\sqrt{n}$	0.885±0.018	0.822±0.017	0.877±0.019	0.853±0.015	0.848±0.016	0.708±0.030	0.929±0.008	0.934±0.011
	1000	$log_2 n$	0.888±0.022	0.822±0.018	0.881±0.022	0.855±0.015	0.850±0.015	0.712±0.031	0.925±0.011	0.935±0.012
		$\sqrt{n}$	0.888±0.020	**0.828±0.018**	0.881±0.021	0.858±0.016	0.853±0.017	0.717±0.033	**0.931±0.008**	0.939±0.010
entropy	10	$log_2 n$	0.879±0.021	0.779±0.019	0.866±0.018	0.829±0.009	0.820±0.009	0.662±0.018	0.900±0.008	0.903±0.010
		$\sqrt{n}$	0.889±0.026	0.794±0.026	0.878±0.027	0.842±0.022	0.834±0.024	0.687±0.045	0.904±0.013	0.904±0.016
	100	$log_2 n$	0.890±0.024	0.818±0.019	0.882±0.024	0.854±0.016	0.849±0.016	0.710±0.032	0.923±0.010	0.931±0.012
		$\sqrt{n}$	0.895±0.022	0.822±0.022	0.888±0.023	0.859±0.018	0.853±0.018	0.720±0.035	0.927±0.010	0.935±0.012
	1000	$log_2 n$	0.893±0.026	0.822±0.019	0.885±0.026	0.857±0.018	0.852±0.019	0.717±0.038	0.925±0.011	0.935±0.012
		$\sqrt{n}$	0.897±0.022	0.824±0.017	**0.889±0.022**	**0.860±0.015**	**0.855±0.015**	**0.723±0.031**	0.931±0.009	0.939±0.010
log loss	10	$log_2 n$	0.889±0.022	0.782±0.032	0.876±0.024	0.835±0.022	0.826±0.025	0.675±0.043	0.901±0.016	0.899±0.016
		$\sqrt{n}$	0.897±0.022	0.790±0.022	0.884±0.023	0.843±0.017	0.834±0.019	0.691±0.035	0.905±0.015	0.905±0.015
	100	$log_2 n$	0.890±0.022	0.820±0.020	0.882±0.022	0.855±0.017	0.850±0.017	0.712±0.034	0.922±0.012	0.931±0.012
		$\sqrt{n}$	**0.897±0.021**	0.820±0.020	**0.889±0.022**	0.859±0.017	0.853±0.018	0.719±0.034	0.929±0.010	0.937±0.010
	1000	$log_2 n$	0.889±0.025	0.820±0.018	0.881±0.025	0.854±0.018	0.849±0.018	0.711±0.037	0.925±0.011	0.935±0.012
		$\sqrt{n}$	0.896±0.024	0.824±0.018	0.888±0.025	0.860±0.018	0.855±0.018	0.722±0.037	0.931±0.009	**0.940±0.010**

Table 7: Performance evaluation of the different configurations of the RF model on a 1 : 3 sample of the DTINet dataset

Criterion	No. Estimators	Max No. Features	Sp	Re (Se)	Pr	Acc	F1	MCC	AUC-ROC	AUPR
gini	10	$log_2 n$	0.960±0.011	0.685±0.041	0.852±0.036	0.891±0.012	0.759±0.030	0.697±0.036	0.909±0.014	0.816±0.032
		$\sqrt{n}$	0.963±0.011	0.710±0.032	0.866±0.032	0.900±0.010	0.780±0.023	0.722±0.029	0.919±0.010	0.837±0.021
	100	$log_2 n$	0.964±0.008	0.714±0.036	0.868±0.026	0.901±0.011	0.783±0.027	0.726±0.032	0.933±0.011	0.873±0.023
		$\sqrt{n}$	0.964±0.009	0.735±0.039	0.873±0.029	0.907±0.013	0.797±0.031	0.743±0.037	0.941±0.013	0.889±0.022
	1000	$log_2 n$	0.963±0.010	0.715±0.034	0.866±0.035	0.901±0.014	0.783±0.031	0.726±0.040	0.937±0.011	0.881±0.021
		$\sqrt{n}$	0.967±0.010	**0.742±0.038**	0.884±0.031	**0.911±0.013**	**0.806±0.030**	**0.754±0.037**	**0.945±0.011**	**0.894±0.021**
entropy	10	$log_2 n$	0.966±0.011	0.671±0.038	0.868±0.035	0.892±0.012	0.756±0.030	0.698±0.035	0.913±0.013	0.825±0.030
		$\sqrt{n}$	0.967±0.009	0.686±0.044	0.876±0.030	0.897±0.013	0.769±0.034	0.713±0.039	0.916±0.011	0.834±0.025
	100	$log_2 n$	0.965±0.008	0.714±0.035	0.873±0.027	0.902±0.011	0.785±0.027	0.729±0.033	0.933±0.012	0.874±0.021
		$\sqrt{n}$	**0.969±0.008**	0.718±0.037	**0.885±0.025**	0.906±0.011	0.792±0.027	0.740±0.031	0.940±0.010	0.886±0.020
	1000	$log_2 n$	0.964±0.009	0.713±0.032	0.868±0.031	0.901±0.012	0.782±0.027	0.725±0.034	0.935±0.010	0.880±0.020
		$\sqrt{n}$	0.968±0.009	0.719±0.036	0.884±0.029	0.906±0.012	0.792±0.030	0.739±0.036	0.942±0.011	0.891±0.020
log loss	10	$log_2 n$	0.965±0.010	0.684±0.040	0.868±0.030	0.895±0.012	0.765±0.029	0.707±0.034	0.911±0.013	0.822±0.029
		$\sqrt{n}$	0.966±0.008	0.684±0.036	0.870±0.027	0.896±0.012	0.766±0.030	0.709±0.036	0.916±0.011	0.834±0.029
	100	$log_2 n$	0.965±0.009	0.709±0.032	0.870±0.028	0.901±0.011	0.781±0.026	0.725±0.032	0.932±0.010	0.873±0.022
		$\sqrt{n}$	0.968±0.009	0.723±0.036	0.882±0.029	0.907±0.013	0.795±0.030	0.741±0.036	0.939±0.011	0.885±0.022
	1000	$log_2 n$	0.965±0.010	0.712±0.035	0.873±0.033	0.902±0.013	0.784±0.030	0.728±0.038	0.935±0.011	0.879±0.020
		$\sqrt{n}$	0.969±0.009	0.723±0.032	0.885±0.030	0.907±0.012	0.795±0.027	0.743±0.033	0.942±0.011	0.891±0.020

Table 8: Performance evaluation of the different configurations of the RF model on a 1 : 5 sample of the DTINet dataset

Criterion	No. Estimators	Max No. Features	Sp	Re (Se)	Pr	Acc	F1	MCC	AUC-ROC	AUPR
gini	10	log_2n	0.979±0.005	0.625±0.034	0.858±0.026	0.920±0.005	0.723±0.023	0.689±0.023	0.911±0.011	0.778±0.032
		$\sqrt{n}$	0.980±0.004	0.658±0.033	0.867±0.024	0.926±0.007	0.748±0.026	0.715±0.028	0.920±0.011	0.805±0.023
	100	log_2n	0.979±0.005	0.671±0.033	0.863±0.031	0.927±0.008	0.755±0.028	0.721±0.031	0.935±0.011	0.838±0.024
		$\sqrt{n}$	0.981±0.005	0.694±0.041	0.879±0.029	0.933±0.008	0.775±0.031	0.744±0.034	0.942±0.010	0.859±0.021
	1000	log_2n	0.979±0.006	0.675±0.037	0.867±0.034	0.928±0.009	0.759±0.031	0.725±0.035	0.939±0.011	0.847±0.023
		$\sqrt{n}$	0.982±0.005	**0.704±0.040**	0.887±0.029	**0.936±0.009**	**0.784±0.032**	**0.754±0.035**	**0.947±0.010**	**0.866±0.020**
entropy	10	log_2n	0.978±0.006	0.631±0.037	0.850±0.033	0.920±0.008	0.724±0.030	0.689±0.032	0.910±0.015	0.775±0.031
		$\sqrt{n}$	0.982±0.007	0.652±0.037	0.880±0.043	0.927±0.010	0.748±0.034	0.718±0.039	0.915±0.012	0.796±0.027
	100	log_2n	0.981±0.005	0.668±0.034	0.876±0.032	0.929±0.008	0.758±0.028	0.726±0.031	0.935±0.012	0.840±0.023
		$\sqrt{n}$	0.982±0.005	0.677±0.036	0.885±0.029	0.931±0.008	0.767±0.028	0.737±0.031	0.939±0.010	0.852±0.022
	1000	log_2n	0.981±0.006	0.670±0.035	0.874±0.036	0.929±0.008	0.758±0.030	0.726±0.034	0.938±0.011	0.846±0.023
		$\sqrt{n}$	0.983±0.005	0.682±0.041	0.888±0.031	0.933±0.008	0.771±0.030	0.742±0.033	0.945±0.010	0.862±0.021
log loss	10	log_2n	0.978±0.004	0.613±0.027	0.848±0.026	0.917±0.005	0.711±0.021	0.676±0.022	0.907±0.010	0.777±0.023
		$\sqrt{n}$	0.981±0.005	0.645±0.034	0.874±0.028	0.925±0.005	0.741±0.022	0.710±0.023	0.915±0.013	0.795±0.024
	100	log_2n	0.981±0.007	0.666±0.032	0.876±0.042	0.928±0.009	0.756±0.032	0.724±0.037	0.935±0.012	0.840±0.026
		$\sqrt{n}$	**0.983±0.004**	0.673±0.042	0.891±0.026	0.932±0.008	0.766±0.031	0.737±0.033	0.940±0.009	0.853±0.021
	1000	log_2n	0.981±0.006	0.667±0.035	0.874±0.034	0.928±0.008	0.756±0.030	0.724±0.033	0.938±0.011	0.847±0.023
		$\sqrt{n}$	0.983±0.005	0.679±0.037	**0.892±0.030**	0.933±0.008	0.771±0.029	0.742±0.032	0.944±0.010	0.861±0.022

Table 9: Performance evaluation of the different configurations of the RF model on a 1 : 10 sample of the DTINet dataset

Criterion	No. Estimators	Max No. Features	Sp	Re (Se)	Pr	Acc	F1	MCC	AUC-ROC	AUPR
gini	10	$log_2 n$	0.991±0.002	0.556±0.043	0.856±0.018	0.951±0.003	0.673±0.031	0.666±0.026	0.908±0.009	0.714±0.030
		$\sqrt{n}$	0.992±0.002	0.591±0.037	0.875±0.025	0.955±0.004	0.705±0.031	0.698±0.030	0.919±0.014	0.748±0.027
	100	$log_2 n$	0.991±0.001	0.598±0.040	0.863±0.020	0.955±0.004	0.706±0.033	0.696±0.033	0.938±0.010	0.790±0.024
		$\sqrt{n}$	0.992±0.001	0.631±0.046	0.884±0.016	**0.959±0.004**	0.736±0.034	0.727±0.033	0.949±0.009	0.822±0.021
	1000	$log_2 n$	0.991±0.002	0.605±0.038	0.870±0.024	0.956±0.004	0.713±0.032	0.704±0.032	0.945±0.009	0.800±0.024
		$\sqrt{n}$	0.992±0.001	**0.634±0.045**	0.889±0.015	**0.959±0.004**	**0.739±0.032**	**0.730±0.030**	**0.953±0.010**	**0.830±0.021**
entropy	10	$log_2 n$	0.992±0.001	0.548±0.040	0.866±0.015	0.951±0.003	0.671±0.030	0.666±0.026	0.910±0.012	0.716±0.032
		$\sqrt{n}$	0.992±0.002	0.586±0.038	0.886±0.022	0.956±0.004	0.705±0.031	0.699±0.030	0.917±0.013	0.747±0.026
	100	$log_2 n$	0.991±0.001	0.589±0.036	0.870±0.021	0.955±0.004	0.702±0.030	0.694±0.029	0.936±0.009	0.790±0.025
		$\sqrt{n}$	**0.993±0.002**	0.610±0.038	**0.892±0.026**	0.958±0.004	0.724±0.030	0.717±0.029	0.945±0.009	0.813±0.021
	1000	$log_2 n$	0.991±0.002	0.594±0.035	0.874±0.020	0.955±0.003	0.707±0.028	0.699±0.027	0.943±0.009	0.799±0.023
		$\sqrt{n}$	0.992±0.002	0.611±0.037	0.886±0.025	0.957±0.004	0.723±0.029	0.715±0.028	0.950±0.008	0.821±0.021
log loss	10	$log_2 n$	0.990±0.002	0.557±0.048	0.852±0.025	0.951±0.003	0.672±0.034	0.664±0.028	0.906±0.010	0.714±0.025
		$\sqrt{n}$	0.992±0.001	0.587±0.047	0.876±0.014	0.955±0.003	0.701±0.032	0.695±0.028	0.916±0.015	0.743±0.023
	100	$log_2 n$	0.992±0.002	0.598±0.039	0.878±0.025	0.956±0.004	0.711±0.031	0.703±0.031	0.938±0.011	0.792±0.024
		$\sqrt{n}$	0.992±0.001	0.607±0.034	0.886±0.015	0.957±0.003	0.720±0.026	0.713±0.024	0.946±0.007	0.811±0.020
	1000	$log_2 n$	0.991±0.001	0.601±0.041	0.876±0.021	0.956±0.004	0.712±0.032	0.704±0.031	0.944±0.009	0.800±0.023
		$\sqrt{n}$	0.992±0.002	0.611±0.038	0.884±0.027	0.957±0.004	0.722±0.031	0.715±0.031	0.949±0.009	0.820±0.021

Table 10: Methods performance evaluation on 1 : 1 positive to negative sampling ratio of the DTINet dataset

Methods	Sp	Re (Se)	Pr	Acc	F1	MCC	AUC-ROC	AUPR
NeoDTI[2]	–	0.65	0.57	–	0.76	–	0.83	0.79±0.03
HIDTI[3]	–	0.85	0.85	–	0.85	–	0.92	0.90±0.02
IMCHGAN[4]	0.93±0.03	0.89±0.04	0.93±0.03	0.91±0.01	0.91±0.01	0.82±0.03	0.95±0.01	0.96±0.01
SVM	0.91±0.02	0.88±0.02	0.89±0.02	0.87±0.01	0.87±0.01	0.73±0.02	0.93±0.01	0.93±0.01
RF	0.90±0.02	0.83±0.02	0.89±0.02	0.86±0.02	0.86±0.02	0.72±0.03	0.93±0.01	0.94±0.01
IEDTI	0.82±0.05	0.86±0.03	0.83±0.04	0.84±0.02	0.84±0.02	0.68±0.05	0.91±0.02	0.91±0.02
DEDTI	0.90±0.02	0.89±0.03	0.90±0.02	0.90±0.02	0.89±0.02	0.79±0.03	0.95±0.01	0.95±0.01
MAD-TI	0.98±0.03	0.98±0.02	0.98±0.03	0.98±0.02	0.98±0.02	0.96±0.05	0.99±0.01	0.99±0.02

Table 11: Methods performance evaluation on 1 : 3 positive to negative sampling ratio of the DTINet dataset

Methods	Sp	Re (Se)	Pr	Acc	F1	MCC	AUC-ROC	AUPR
NeoDTI[2]	–	0.65	0.62	–	0.63	–	0.81	0.65±0.11
HIDTI[3]	–	0.79	0.83	–	0.81	–	0.94	0.86±0.04
IMCHGAN[4]	0.94±0.03	0.90±0.03	0.89±0.03	0.93±0.02	0.88±0.03	0.82±0.05	0.96±0.01	0.92±0.02
SVM	0.92±0.01	0.86±0.03	0.78±0.03	0.90±0.01	0.82±0.03	0.75±0.03	0.95±0.01	0.87±0.04
RF	0.97±0.01	0.74±0.04	0.89±0.03	0.91±0.01	0.81±0.03	0.75±0.04	0.95±0.01	0.89±0.02
IEDTI	0.95±0.02	0.74±0.03	0.82±0.04	0.89±0.01	0.78±0.01	0.71±0.02	0.92±0.01	0.84±0.02
DEDTI	0.97±0.01	0.85±0.01	0.90±0.03	0.94±0.01	0.87±0.01	0.83±0.02	0.96±0.01	0.92±0.02
MAD-TI	0.99±0.01	0.99±0.01	0.99±0.01	0.99±0.01	0.99±0.01	0.99±0.01	0.99±0.01	0.99±0.01

Table 12: Methods performance evaluation on 1 : 10 positive to negative sampling ratio of the DTINet dataset

Methods	Sp	Re (Se)	Pr	Acc	F1	MCC	AUC-ROC	AUPR
IMCHGAN[4]	0.94±0.03	0.92±0.03	0.87±0.06	0.93±0.03	0.83±0.03	0.71±0.07	0.96±0.01	0.86±0.03
SVM	0.97±0.01	0.85±0.03	0.73±0.02	0.95±0.01	0.76±0.01	0.73±0.02	0.96±0.01	0.81±0.03
RF	0.99±0.01	0.63±0.05	0.89±0.03	0.96±0.01	0.74±0.03	0.73±0.03	0.95±0.01	0.83±0.02
IEDTI	0.99±0.01	0.62±0.03	0.80±0.04	0.95±0.01	0.70±0.03	0.68±0.03	0.92±0.01	0.73±0.03
DEDTI	0.99±0.01	0.77±0.03	0.87±0.02	0.97±0.01	0.82±0.02	0.80±0.02	0.96±0.01	0.86±0.02
MAD-TI	0.99±0.01	0.99±0.01	0.99±0.01	0.99±0.01	0.99±0.01	0.99±0.02	0.99±0.01	0.99±0.01