597 Sequence and Network features description:

40 features from Liu et al:

'gene_size','prot_size','T3s','C3s','A3s','G3s','CAI','Nc','Gravy','Ala','Arg','Asn','Asp','Cys','Gln','Gly','His','Ile','Leu','Lys','Met','Phe','Pro','Ser','Thr','Trp','Tyr','Val','Isoelectric_Point','Tiny','Aromatic','Nonpolar','Polar','Basic','Acidic','First60','Extracellular_Score','Cytoplasmic_Score','Final_Score','Hurst'

A detailed description of all these features is available in the original paper[1].

283 network features:

14 network measures

'Closeness_Centrality','Information_Centrality','Subgraph_Centrality','Random_Walk_Betweenness_Centrality','Harmonic_Centrality','Betweenness_Centrality','Load_Centrality','Page_Rank','Reaching_Centrality','Eigenvector_Centrality','Edge_Clustering_Coefficient','Clustering_Coefficient','Degree_Centrality','Clique_Number'

267 ReFeX features

'xe1-wnm1-wn1-wem1','wnm1-wnm1-wn1-wnm1-wnm1-xem1','wn1-xe1-wnm1-wem1','xe1-wet1','wnm1-wnm1-xet0','xe1-wnm1-wn1-wnm1-xem1','wn1-wnm1-wnm1-xet1','xe1-wn1-wnm1-xet1','xe1-wnm1-wnm1-xem0','xe1-wnm1-xe1-wnm1-wn1-xem0','wn1-wnm1-wnm1-xe1-wem1','wnm1-wn1-wn1-xem0','xe1-wnm1-wn1-xet1','wn1-wnm1-wnm1-wn1','wnm1-wnm1-wn1-wn1-wem1','xe1-wnm1-wn1','wn1-wem1','wnm1-xe1-xe1-xem1','xe1-wn1-wnm1-xem1','wnm1-wnm1-xet1','wn1-xet0','xe1-xe1-wn1','wnm1-weu1','wnm1-wnm1-wn1-xem0','xe1-wnm1-wnm1-xe1-wem1','wn1-wnm1-wn1-wn1-wnm1-wem1','wnm1-xet0','wnm1-xet1','xe1-wnm1-xe1-wnm1-wet1','xe1-wnm1-xe1-wnm1-xem0','xe1-wnm1-xe1-wnm1-xem1','wnm1-wnm1-wnm1-wem1','wn1-wnm1-xe1-wnm1-xe1-xem0','wnm1-wnm1-wnm1-xe1-wnm1-xem1','wn1-wn1-wem1','wnm1-wn1-wnm1-wem1','wnm1-wnm1-wnm1-wn1-xet1','wnm1-wnm1-wnm1-xe1-xem0','xe1-wnm1-xe1-wn1-xem0','wnm1-wnm1-wnm1-wnm1-xem1','xe1-wnm1-xeu1','wnm1-wnm1-xe1-wnm1-xem0','xe1-weu1','wnm1-wnm1-xe1-wnm1-xem1','wnm1-wnm1-xe1-wnm1-xe1-wem1','wnm1-wnm1-wet1','wnm1-wnm1-wnm1-wn1','wnm1-wem1','xe1-wnm1-wnm1-xem1','wn1-wnm1-wn1','xe1-xe1-wnm1-wem1','wn1-wnm1-wem1','xe1-xe1-xet0','xe1-wn1-xem1','wnm1-wnm1-wnm1-xe1-wn1','wnm1-xe1-wn1','wnm1-wn1-wnm1-xe1-wnm1-wn1-xem0','wnm1-xe1-wnm1-xe1-xem1','wnm1-wnm1-xe1-xe1-xem1','wn1-wnm1-wnm1-xe1-wnm1-xem1','xeu1','wnm1-wn1-wn1-wnm1-xem0','wnm1-wn1-wn1-wnm1-xem1','xe1-wnm1-wnm1-wem1','wn1-wet1','wn1-wnm1-wn1-xem0','wnm1-xe1-wn1-wnm1-xem1','xe1-wnm1-xe1-wet1','wnm1-xe1-xe1-wem1','wnm1-xe1-wn1-xem1','xe1-wnm1-wnm1-xet1','wn1-wnm1-xe1-wnm1-xem0','wn1-wnm1-xe1-wnm1-xem1','xe1-xe1-xem0','xe1-xe1-xem1','xe1-xe1-xet1','wnm1-xe1-xe1-wnm1-xem0','xe1-wnm1-wn1-wn1-xem0','wn1-wnm1-xe1-xem0','wnm1-wnm1-xe1-wnm1-wnm1-wem1','xe1-wnm1-wnm1-xe1-wet1','wnm1-wnm1-wn1-wn1-wnm1-xem1','wnm1-wnm1-wn1-wn1-wnm1-xem0','wnm1-wnm1-weu1','xe1-wn1-xem0','wnm1-wnm1-xe1-wnm1-wem1','xe1-wnm1-wnm1-wn1-wem1','wn1-wnm1-wnm1-xe1-wnm1-wem1','xe1-wnm1-xe1-xet0','xem0','xem1','wnm1-xe1-wn1-xem0','xe1-wnm1-xe1-xem1','xe1-wnm1-xe1-xem0','xe1-xem1','xe1-xem0','xe1-wnm1-wnm1-xeu1','wnm1-wnm1-xe1-wn1-xem1','xe1-xet0','xe1-xet1','wnm1-wnm1-wem1','wnm1-xe1-wnm1-wnm1-xem0','wnm1-xe1-wnm1-wnm1-xem1','wnm1-wn1','wn1-wnm1-wn1-wnm1-wem1','wnm1-xe1-wnm1-wnm1-wem1','wnm1-xe1-wn1-wem1','xet1','xet0','wnm1-wnm1-wnm1-wnm1-xet1','wn1-wn1-wnm1-xem1','wn1-

wn1-wnm1-xem0','wnm1-xe1-wnm1-wnm1-xe1-wem1','wnm1-xe1-xet1','wnm1-wnm1-wn1','wn1-wnm1-wnm1-wnm1-xem1','wn1-wnm1-wet1','wn1-wnm1-xet1','wn1-wnm1-xet0','wnm1-xe1-wnm1-xe1-wnm1-xem1','wnm1-wnm1-xe1-xem1','wnm1-wnm1-xe1-xem0','wnm1-wnm1-xe1-xet0','xe1-xe1-wet1','wnm1-xe1-wnm1-wn1','wnm1-xe1-wnm1-xet0','wn1-wn1-xem0','wn1-wn1-xem1','wn1-wn1-wnm1-wem1','wn1-xe1-wnm1-xem0','xe1-wn1-wem1','wnm1-xe1-wnm1-wem1','xe1-wnm1-xem1','xe1-wn1-wnm1-wem1','wn1-wnm1-xe1-wem1','wnm1-wn1-wnm1-wnm1-wn1-xem1','wnm1-wnm1-xe1-wem1','wnm1-wnm1-xe1-wnm1-xet0','xe1-wnm1-xe1-wnm1-wem1','wn1-wn1-wnm1-wnm1-xem0','wn1-wn1-wnm1-wnm1-xem1','wn1-wnm1-xem0','wn1-wnm1-xem1','xe1-xe1-wnm1-xem1','xe1-xe1-wnm1-xem0','wn1','xe1-wnm1-xe1-wem1','wnm1-xe1-wnm1-wet1','wnm1-xe1-xeu1','wnm1-xe1-wem1','wnm1-wnm1-xe1-wn1-wem1','wnm1-wn1-wnm1-xem1','wnm1-wn1-wnm1-xem0','wnm1-wn1-wnm1-xet1','wnm1-wn1-wn1','xe1-xe1-wnm1-wnm1-wem1','xe1-wnm1-xe1-wn1-wem1','wnm1-wnm1-xe1-wnm1-xeu1','wnm1-xe1-wn1-wnm1-wn1','wnm1-xe1-wn1-wnm1-wem1','wnm1-wn1-wem1','wn1-wnm1-wnm1-wem1','wnm1-wn1-wnm1-wnm1-xe1-xem0','wn1-wnm1-wnm1-xe1-wnm1-wn1','wnm1-wn1-xe1-wnm1-xem0','xe1-xe1-wnm1-wnm1-xe1-wet1','xe1-xe1-wnm1-wn1','xe1-wnm1-wnm1-xe1-xem0','xe1-wn1-wnm1-wn1','wnm1-xe1-xet0','wnm1-wnm1-xe1-wnm1-wn1-wem1','wn1-wnm1-wnm1-xem1','wn1-wnm1-wnm1-xem0','wnm1-xe1-weu1','wnm1-wn1-wnm1-wnm1-xem1','xe1-xe1-wnm1-xeu1','wn1-wnm1-wnm1-xe1-xem0','wnm1-wnm1-wnm1-xe1-wet1','wnm1-wnm1-wn1-wnm1-xe1-wn1','wnm1-xe1-xem0','wnm1-xe1-xem1','wn1-wn1-wnm1-wn1-wem1','wnm1-wnm1-wnm1-wn1-wnm1-xem1','xe1-wnm1-wnm1-wn1-xet1','wnm1-wnm1-xe1-wet1','wn1-xet1','wn1-wnm1-wn1-wnm1-xem0','wnm1-wnm1-wn1-xet1','xe1-wnm1-xe1-wnm1-xe1-wem1','wnm1-wnm1-wnm1-xem1','wnm1-wnm1-wnm1-xem0','xe1-wnm1-wnm1-wn1-xem1','xe1-wnm1-wnm1-wnm1-xe1-xem0','wnm1-wn1-wnm1-wn1-wem1','wnm1-wnm1-wnm1-xe1-wem1','xe1-xe1-wem1','xe1-xeu1','xe1-wn1','wnm1-wnm1-wnm1-xet1','xe1-wnm1-wem1','wnm1-xe1-wnm1-wn1-wem1','xe1-xe1-wnm1-xet1','wnm1-wet1','wn1-wn1-wnm1-wnm1-wnm1-xem1','wnm1-xe1-wnm1-xet1','wn1-xem1','wn1-xem0','wnm1-xe1-wet1','xe1-wnm1-wn1-xem1','xe1-wnm1-wn1-xem0','xe1-wnm1-wnm1-wnm1-xem1','wn1-wnm1-xe1-xem1','wnm1-wn1-xet1','wnm1-wn1-xem0','wnm1-wn1-xem1','wn1-wnm1-wnm1-wn1-wem1','wnm1-xe1-xe1-xet1','wn1-wnm1-wnm1-wn1-xem0','wn1-wnm1-wnm1-wn1-xem1','wn1-wnm1-xe1-wn1','wnm1-wnm1-xe1-wn1','xe1-wnm1-wet1','wnm1-xe1-wnm1-xem1','wnm1-xe1-wnm1-xem0','wnm1-xe1-wnm1-xe1-wem1','wet1','wnm1-wnm1-wnm1-xeu1','wnm1-wnm1-xe1-wnm1-wn1','wn1-wnm1-xe1-wnm1-wn1-xem0','wnm1-wnm1-wnm1-wet1','wnm1-wnm1-xeu1','wnm1-wn1-wnm1-xe1-wn1','xe1-xe1-wnm1-wnm1-xem1','wnm1-xem1','wnm1-xem0','wnm1-wnm1-wnm1-wnm1-xeu1','wnm1-xe1-wnm1-xeu1','wnm1-wn1-wnm1-wn1-xem0','wnm1-wn1-wn1-wem1','wnm1-wnm1-wn1-wem1','wnm1-xe1-wnm1-xe1-wnm1-wem1','wn1-wnm1-wn1-wem1','wnm1-wn1-wn1-xem1','xe1-wnm1-xet1','xe1-wnm1-xet0','xe1-wnm1-xem0','wnm1-wnm1-wn1-wnm1-xem1','wnm1-xe1-xe1-wn1','wnm1-wnm1-xe1-wnm1-wnm1-xem1','wn1-wn1','xe1-wem1','wnm1-wnm1-xem0','wnm1-wnm1-xem1','wnm1-wnm1-wnm1-wn1-wem1','wnm1-wn1-wn1-wnm1-wem1','wnm1-wnm1-wnm1-xet0','xe1-wnm1-wnm1-xe1-wnm1-wn1','weu1','wnm1-xe1-wnm1-xe1-xem0','wnm1-xe1-xe1-wnm1-xet1','wnm1-xeu1','wem1','xe1-wnm1-xe1-wnm1-wn1','wnm1-xe1-wnm1-wn1-xem0','wn1-xe1-wn1-wnm1-wn1','wnm1-wnm1-wn1-xem1','wnm1-xe1-wnm1-wnm1-wnm1-xem1'

2 RIDER Properties:
'biconnected_components','weighted_degree'
The following are the 274 ZUPLS features that we used. For more details, please refer ZUPLS article[2]

93 ZCURVE features:
Phase independent then phase dependent
Phase-independent parameters of mononucleotide -- 3 'X','Y','Z'
Phase-independent parameters of di-nucleotide -- 9 'Xa','Ya','Za','Xt','Yt','Zt','Xc','Yc','Yg'
Phase-independent parameters of tri-nucleotide -- 22
'Xat','Zat','Yag','Xta','Zta','Xtt','Ytt','Ztt','Xca','Xct','Zct','Xcc','Zcc','Xcg','Ycg','Zga','Xgt','Xgc','Ygc','Zgc','Ygg','Zgg'
Phase-specific parameters of mononucleotide -- 9 'Xf1','Yf1','Zf1','Xs2','Ys2','Zs2','X3','Y3','Z3'
Phase-specific parameters of di-nucleotides --14
'Xa1','Za1','Xt1','Yt1','Zt1','Yc1','Zc1','Xa2','Za2','Xt2','Zt2','Zc2','Xt3','Zt3'
Phase-specific parameters of tri-nucleotides --36
'Xat1','Zat1','Zac1','Yag1','Xta1','Xtt1','Ytt1','Xtg1','Ytg1','Ztg1','Xca1','Zcc1','Zgg1','Zaa2','Zat2','Zag2','Xta2','Xtt2','Ztt2','Xtg2','Yca2','Xct2','Xcc2','Zcc2','Ygg2','Zgg2','Xtt3','Xtc3','Ytc3','Ytg3','Ztg3','Xct3','Yct3','Xcc3','Ycc3','Xgt3'

Paralogs:
'paralogs'

180 Orthologs: (KEGG 3-letter codes)
'aae','aci','afu','ama','ana','ape','atu','bab','ban','bba','bbr','bbu','bce','bcl','bfl','bfr','bfs','bga','bha','bhe','bja','bli','blo','bma','bme','bms','bpa','bpe','bps','bqu','bsu','bth','btk','buc','cac','cbu','cca','ccr','cdi','cef','cgl','cje','cmu','cpe','cpn','ctc','cte','ctr','cvi','det','dps','dra','dvu','eba','eca','eco','efa','erg','eru','erw','fnu','ftu','gka','gox','gsu','gvi','hal','hdu','hhe','hin','hma','hpy','ilo','lac','lil','lin','ljo','lla','lmo','lpl','lpn','lxx','mac','mbo','mca','mfl','mga','mge','mhy','mja','mka','mle','mlo','mma','mmo','mmp','mmy','mpa','mpe','mpn','mpu','msu','mth','mtu','neq','neu','nfa','ngo','nme','oih','pab','pac','pae','pai','pcu','pfu','pgi','pho','plu','pma','pmm','pmu','poy','ppr','ppu','pst','pto','rba','rco','rpa','rpr','rso','rty','sac','sag','sco','sep','ser','sfl','sil','sma','sme','smu','son','spn','spt','spy','sso','sth','stm','sto','sty','syc','syn','tac','tde','tel','tko','tma','tpa','tte','tth','tvo','twh','uur','vch','vfi','vpa','vvu','wbm','wbr','wol','wsu','xac','xcc','xfa','xoo','ype','yps','zmo'

Undersampling strategy Description[1]:
Step 1: The entire dataset was split into 2/3 rd training set and 1/3 rd testing set. This was repeated 5 times with different random splitting of train and test datasets.
Step 2: Then, under-sampled training data was obtained from 2/3 rd training set and this was repeated 10 times (Under sampling of the dataset of non-essential genes was repeated 10 times whereas essential genes is fixed for all 10 times)
Step 3: Now, new train set (under sampled non-essential and essential) was constructed and 5 fold cross validation was done to find the best parameters using SVM.
Step 4: The best classifier was found and tested on 1/3rd testing set.
Step 5: Then,different metrics are reported - Sensitivity, Specificity, AUC, Average of sensitivity and specificity, Precision and Accuracy.

Step 6: Thus, 5 times random splitting of the entire dataset combined with 10 times under sampling of training data gives 50 values. The average of this is taken and reported.

References

[1] X. Liu, B. Wang, L. Xu, H. Tang, and G. Xu, “Selection of key sequence-based features for prediction of essential genes in 31 diverse bacterial species,” pp. 1–13, 2017.

[2] K. Song, T. Tong, and F. Wu, “Predicting essential genes in prokaryotic genomes using a linear method: ZUPLS,” *Integr. Biol.*, vol. 6, no. 4, pp. 460–469, 2014.