JINGUI CHEN used a deep learning frame work with are current neural network (RNN) and a convolutional neural network (CNN) to execute a new glycation site predictor named DeepGly to identify protein glycation sites from imbalanced data. They have been used three types of dataset. One is dataset A (used CPLM database), then dataset B (used lysine glycation site dataset) and dataset C (employed to build glycation site predictor ,it used before dataset A and dataset B) .They used LSTM RNNs model to solve imbalance dataset problem using oversample peptide chains with glycation sites and for data redundancy they used CD-HIT with a threshold of 50% . In this study ,they worked on dataset A and dataset B using 10-fold cross-validation with CNN and SVM method. In dataset A, the model obtained performance using CNN, Sp 96.4%,Sn 78.9%,Acc 90.5% and AUC 0.766 and using SVM Sp 76.2%, Sn 100% ,Acc 76.6% and AUC 0.233 and in dataset B, the model gained performance using CNN, Sp 97.7%,Sn 84.4%,Acc 92.1% and AUC 0.838 and using SVM Sp 60.3%, Sn 100%, Acc 65.2% and AUC 0.405. DeepGly has been compared with existing methods such as Gly-PseAAC, iProtGly-SS and PreGly. DeepGly gave better performance than all of methods.

@article{chen2019deepgly,

title={DeepGly: A Deep Learning Framework With Recurrent and Convolutional Neural Networks to Identify Protein Glycation Sites From Imbalanced Data},

author={Chen, Jingui and Yang, Runtao and Zhang, Chengjin and Zhang, Lina and Zhang, Qian},

journal={IEEE Access},

volume={7},

pages={142368--142378},

year={2019},

publisher={IEEE}

}