4. BINDING AFFINITY PREDICTION MODEL USING PROTEIN-LIGAND DOCKING AND REGRESSION TECHNIQUES

Binding affinity prediction for a hereditary disorder is significant for drug detection in therapeutic field. It is intricate to predict binding affinity when there is a structural change in protein due to binding with ligands and also with mutations that occur in protein structure. Hence, it is required to predict binding affinity for SCA to monitor the structural changes in protein and also the changes in their physio-chemical properties that enables in accurate prediction of binding affinity. The predicted binding affinity further reveals specific path for drug designing.

Binding affinity prediction methods use general approaches and the machine learning algorithms aids in accurate prediction. Machine learning algorithms are used in genomics to recognize the genetic causes, treatments through genes and proteins and also the affinity prediction of proteins-ligands, protein-protein etc. Affinity prediction of various disorders and drug-target identification has been effectively done using various regression techniques. Some of the works performed using general approaches and machine learning are macromolecule-ligand interaction [81], protein-protein affinity prediction [58], drug-target interaction prediction [82], protein-RNA interactions [83].

This chapter illustrates the development of binding affinity predictive models built through protein-ligand docking and supervised regression techniques.

4.1 PROTEIN-LIGAND DOCKING BASED BINDING AFFINITY PREDICTIVE MODELS USING SUPERVISED LEARNING

This work explains binding affinity prediction models built through protein-ligand docking using supervised regression techniques. The binding affinity prediction problem is formulated as regression task and the regression models are built through the intelligence attained from the training data. The performance of the models is evaluated using performance metrics such as explained variance score, mean squared error, root mean squared error, median absolute error, mean absolute error.

Methodology

Binding affinity predictive model is constructed by accumulating the protein structures from PDB corpus and gathering ligands from gene cards. Protein structures associated with six types of SCA are considered and the ligands relating to SCA are treated from various literatures and docked using autodock. Docked complexes are employed for feature extraction and PLD dataset is created. This dataset is trained using various regression techniques such as linear regression, random forest, support vector regression and artificial neural network to build predictive models. The proposed framework of binding affinity prediction model based on protein-ligand docking is shown in Fig. 4.1. The model includes four components namely corpus development, feature extraction and dataset creation, model building and evaluation of the binding affinity predictive models.

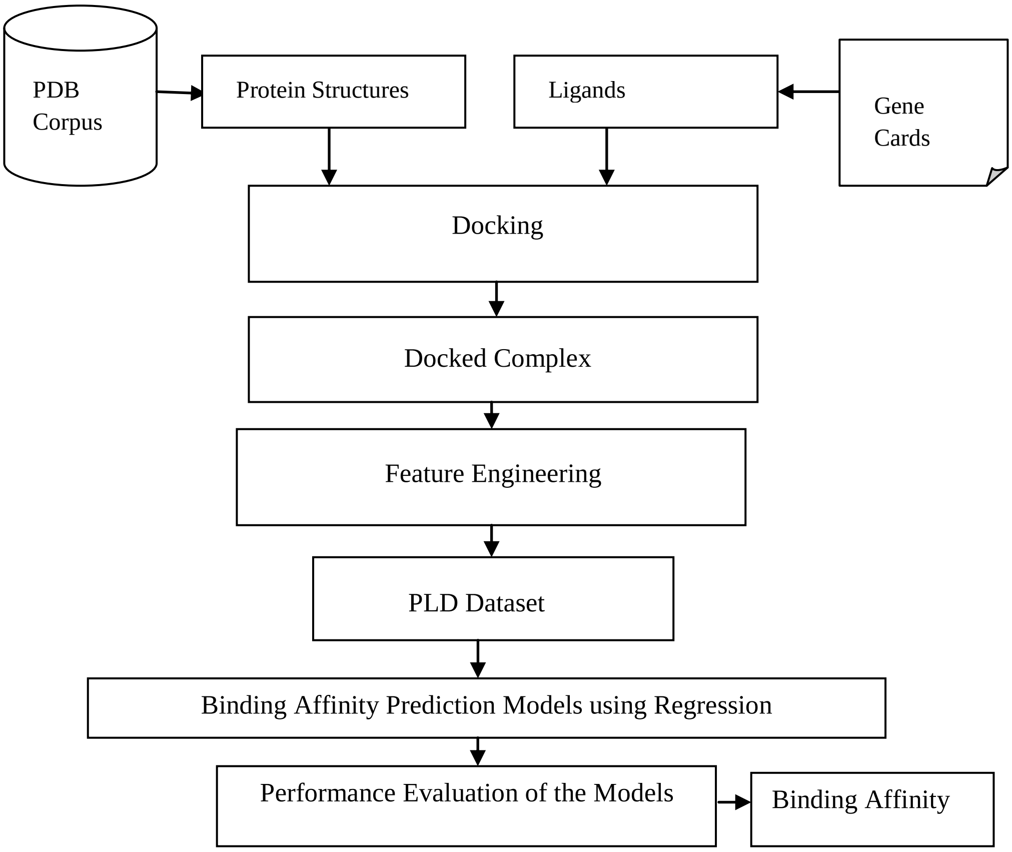


Fig. 4.1 Proposed Framework of Binding Affinity Prediction Model Based on Protein-Ligand Docking

Corpus Development

Protein structures associated with six types of SCA as given in Table III are taken from PDB and ligand from genecards. Structures of ligands shown in Fig. 1.10 are chosen from genecards. The ligands such as amantadine, benztropine, biperiden, bromocriptine, carbidopa, donepezil, entacapone, galantamine, levodopa, pergolide, pramipexole, procyclidine, rivastigmine, ropinirole, selegiline and tacrine are used to dock with proteins. Flexible docking is preferred where ligand rotates the protein and optimum pose is selected to get docked complex. Protein-ligand docking is performed in autodock and totally seventeen protein structures and eighteen ligands are considered for docking. The protein structures are prepared for docking by adding hydrogen bonds and computing charges. The protein structures are then added with grid and autogrid file is computed. Ligand is prepared by finding its root to dock with the protein. The autodock file is computed for ligand. Protein and ligand is docked where the cluster of conformation energies are produced. The minimum energy is chosen for binding energy as it indicates the strong binding. Each protein is docked with ligand in order to produce the docked complex. Totally 307 complexes are created and the corpus is developed. The detailed description of PL corpus development has been given in section 3.2 of chapter 3.

Feature Extraction and Dataset Creation

Efficient features are derived from the docked complexes obtained through protein-ligand docking. Dimensions like energy calculations and physical properties of protein and ligand are extracted using autodock, autodock vina and pymol. Energy calculations like vanderwaals energy, desolvation energy, torsional energy, electrostatic energy, inhibition constant, ligand efficiency etc., and physical properties like molecular weight of both protein, ligand, complex; surface area of solvent for protein, ligand and complex are extracted from the docked complexes of protein-ligand docking.

Energy Calculations: Energy calculations are important in predicting binding affinity where the vanderwaals energy plays role in attraction and repulsion of atoms, molecules and surfaces. Vanderwaals is the weakest attraction, along with this bond covalent and ionic bonds helps in proper function of protein. Solvation energy is important because the complex is fed to a solvent where the chemical reaction occurs between the bonds and the energy is measured. Torsional energy is measured as the dihedral angles in the protein structure where the energy changes due to rotation and the best value of energy is calculated. Electrostatic energy is the energy that occurs due to reaction of charged atoms and it is measured to calculate the changes in the energy. Inhibition constant measures the potency of inhibitor. These energy values are very significant to calculate binding affinity. Binding affinity is the measure, how well the structure binds with the ligand. Binding energy is the energy of docked complex that is bound together with the ligand. Physical properties are measured for ligand, complex and protein structure where the changes occur in molecular weight, atom count, charge and surface area of all the solvents are changed due to docking where they are collected from pymol. The derivation of each feature is given in detail below:

Binding Energy Range: The binding energy range describes at which cluster the binding energy falls. The binding energy range is the difference between highest and lowest energies among the protein-ligand complexes. For example, binding energy range obtained for the complex 1oa8 with amantadine shown in Fig. 3.6 is 0.26.

Binding Energy: Binding energy is released when a drug molecule associates with a target, that leads to lower the overall energy of the complex. The release in binding energy transforms the ligand from its minimum energy to its bound conformation with the protein. Lower the binding energy more stable the complex. The binding energy is calculated using the equation given in 4.1.

$$\Delta G = (V_{\mathrm{bound}}^{\mathrm{L-L}} - V_{\mathrm{unbound}}^{\mathrm{L-L}}) + (V_{\mathrm{bound}}^{\mathrm{P-P}} - V_{\mathrm{unbound}}^{\mathrm{P-P}}) + (V_{\mathrm{bound}}^{\mathrm{P-L}} - V_{\mathrm{unbound}}^{\mathrm{P-L}} + \Delta S_{\mathrm{conf}}) \quad \textbf{(4.1)}$$

Where P refers to the protein, L refers to the ligand, V refers to the pair-wise evaluations, and ΔS~conf~ denotes the loss of conformational entropy upon binding. Intermolecular energy and torsional energy both are significant to calculate binding energy. The energy of ligand and protein in the unbound state is calculated and then the energy of the protein-ligand complex is calculated. The binding energy is calculated as the difference between energy in unbound state and energy of protein-ligand complex. For example, binding energy obtained for the complex 1oa8 with amantadine shown in Fig. 3.6 is -6.47.

Ligand Efficiency: Ligand efficiency is a measurement of the binding energy per atom of a ligand to its binding partner, such as a receptor or enzyme. Mathematically, ligand efficiency (LE) can be defined as the ratio of Gibbs free energy (ΔG) to the number of non-hydrogen atoms of the compound. Ligand efficiency is calculated using the equation given in 4.3.

$$LE = (\Delta G)/N \quad \textbf{(4.2)}$$

where $\Delta G = -RT\ln K_i$ and N is the number of non-hydrogen atoms. It is transformed to the equation:

$$LE = 1.4(-\mathit{log}IC_{50})/N \quad \textbf{(4.3)}$$

For example, ligand efficiency obtained for the complex 1oa8 with amantadine shown in Fig. 3.6 is -0.59.

Inhibition Constant (pIC50): The inhibitor constant, K_i, is an indication of how potent an inhibitor is. It is the concentration required to produce half maximum inhibition. The IC50 value is determined at only one concentration of substrate over a range of inhibitor concentrations. While K_i is a constant value for a given compound with an enzyme, an IC50 is a relative value, whose magnitude depends upon the concentration of sub- strate. According to the FDA, IC50 represents the concentration of a drug that is required for 50% inhibition in vitro. The inhibition constant is calculated using the equation given in 4.4.

K_i = dissociation constant of the enzyme-inhibitor complex = K_d

$K_i = [E][I]/[EI]$

$\ln K_b = -\ln K_i$

deltaG(binding) = -R*T*ln K_b

deltaG(inhibition) = R*T*ln K_i

Binding and Inhibition occur in opposite directions, so the minus-sign is lost:

deltaG = R*T*lnK_i,

deltaG/(R*T) = lnK_i

K_i = exp(deltaG/(R*T)) **(4.4)**

For example, inhibition constant obtained for the complex 1oa8 with amantadine shown in Fig. 3.6 is 18.24.

Intermolecular Energy: Intermolecular energy is the energy between non-bonded atoms that is the energy between atoms separated by 3-4 bonds or between atoms in different molecules. For example, intermolecular energy obtained for the complex 1oa8 with amantadine shown in Fig. 3.6 is -6.76.

Vanderwaal's Desolvation Energy: Desolvation energy is the static van der waals energy. It is the lose of the interaction between substance or organic compound and solvent upon binding describes the energy. For example electro-statically bound particles, dissociate by releasing water in an aqueous solution. The desolvation energy is calculated using the equation given in 4.5.

$$\Delta G_{desolv} = W_{desolv} \sum_{i\,(C),\,j} (S_i * V_j * \exp(-r_{ij}^2 / (2 * \sigma^2))) \quad \textbf{(4.5)}$$

For example, the desolvation energy obtained for the complex 1oa8 with amantadine shown in Fig. 3.6 is -5.35.

Electrostatic Energy: Electrostatic energy is the long term interaction between charged atoms. The example of electrostatic energy is, to hold balloon against ceiling. The electrostatic energy is calculated using the equation given in 4.6.

$$\Delta G_{elec} = W_{elec} \sum_{i,\,j} (q_i * q_j) / (\varepsilon(r_{ij}) * r_{ij}) \quad \textbf{(4.6)}$$

For example, the electrostatic energy obtained for the complex 1oa8 with amantadine shown in Fig. 3.6 is -1.42.

Total Internal Energy: Total energy is that the total of changes of all energetic terms enclosed in rating operates of matter or supermolecule upon binding, and the changes upon binding of the entropic terms. For example, the total internal energy obtained for the complex 1oa8 with amantadine shown in Fig. 3.6 is 0.06.

Torsional Energy: Torsion energy is related to dihedral term of internal energy. Torsional energy is calculated using the equation given in 4.7.

$$\Delta G_{tor} = W_{tor}\, N_{tor} \qquad \textbf{(4.7)}$$

where N_{tor} is the number of all rotatable bonds, excluding guanidinium and amide bonds *etc.* For example, torsional energy obtained by the complex 1oa8 with amantadine shown in Fig. 3.6 is 0.3.

clRMS: It is the root mean squared error of difference between current conformation and the lowest energy conformation in its cluster. For example, the clRMS obtained for the complex 1oa8 with amantadine shown in Fig. 3.6 is 0.

refRMS: It is the root mean squared error of distinction between current conformation coordinates and current reference structure. By default the input substance is utilized as a result of the reference. For example, the refRMS obtained by the complex 1oa8 with amantadine shown in Fig. 3.6 is 30.29.

Binding Affinity: Affinity is a measure of the strength of attraction between a molecule and legend. High affinity binding has strong intermolecular force, whereas low affinity binding has weak intermolecular force. Affinity is calculated using the equation given in 4.11.

$$[R]\,[R]\,K_1 = [DR]\,K\text{-}1 \qquad \textbf{(4.8)}$$

$$K_1/K-1 = [RR]/[R][R] \qquad \textbf{(4.9)}$$

$$\text{Binding Affinity} = K_1/K-1 \qquad \textbf{(4.10)}$$

$$K_d = K-1/K_1 \qquad \textbf{(4.11)}$$

Here, K_d is called as binding affinity constant, K_1 is termed as association constant and k-1 is rate constant. For example, the binding affinity obtained by the complex 1oa8 with amantadine shown in Fig. 3.6 is -4.7.

RMSD: The root mean squared deviation is used to validate the docking with respect to biological configuration. RMSD is the measure of the average distance between the atoms. The value of RMSD is obtained using the equation given in 4.12. The rmsd l.b denotes lower bound of root mean squared error whereas rmsd u.b denotes upper bound of root mean squared error.

$$RMSD = \sqrt{1/\mathrm{N} \sum_{\mathrm{i}=1}^{\mathrm{N}} (\mathrm{x_{ci}} - \mathrm{x_{di}})^2 + (\mathrm{y_{ci}} - \mathrm{y_{di}})^2 + (\mathrm{z_{ci}} - \mathrm{z_{di}})^2} \qquad \textbf{(4.12)}$$

For example, the RMSD obtained for the complex 1oa8 with amantadine shown in Fig. 3.6 is 7.888, 9.231 as rmsd l.b and rmsd u.b respectively.

Physical Properties: Physical properties such as molecular weight of ligand, molecular weight of complex, atom count in protein, atom count in ligand, atom count for complex, surface area of protein, surface area-solvent access of protein, surface area of ligand, surface area-solvent access of ligand, surface area solvent of complex and charges of protein are derived from pymol. For example, the physical properties obtained for for the complex 1oa8 with amantadine are given below.

molecular weight of ligand = 151.2487

molecular weight of complex = 13041.16

atom count in protein = 975

atom count in ligand = 28

atom count for complex = 1190

surface area of protein = 12734.07

surface area-solvent access of protein = 7834.376

surface area of ligand = 700.997

surface area-solvent access of ligand = 2760.954

surface area solvent of complex = 13291.47

charges of protein = -7

Feature Importance using Correlation Matrix: Correlation matrix shows the correlation coefficients between two variables. In this work, pearson correlation matrix is used as the dataset constitutes continuous variables. The pearson correlation values lies between -1 to 1 wherein the value 1 refers positive correlation, -1 demotes negative correlation and 0 refers to there is no correlation. The value below -0.5 or above 0.5 is referred to as notable correlation and values below these values are suggested as less notable correlation. The correlation matrix of feature vectors is shown in Fig. 4.2. In this correlation matrix the feature molecular weight of complex has the value of 1, root mean squared deviation upper bound have the value of 0.99, vanderwaals desolvation energy posses the high value of 0.9, atom count in protein holds the value of 0.82, torsional energy has the value of 0.7, binding energy posses the value of -0.59. The features, binding energy range and ligand efficiency have the low correlation values of 0.2 and 0.02 respectively. Electrostatic energy posses the value of -0.4 and the binding energy range has the value of -0.2. This matrix determines the relation of independent variables (X) with dependent variable (Y).

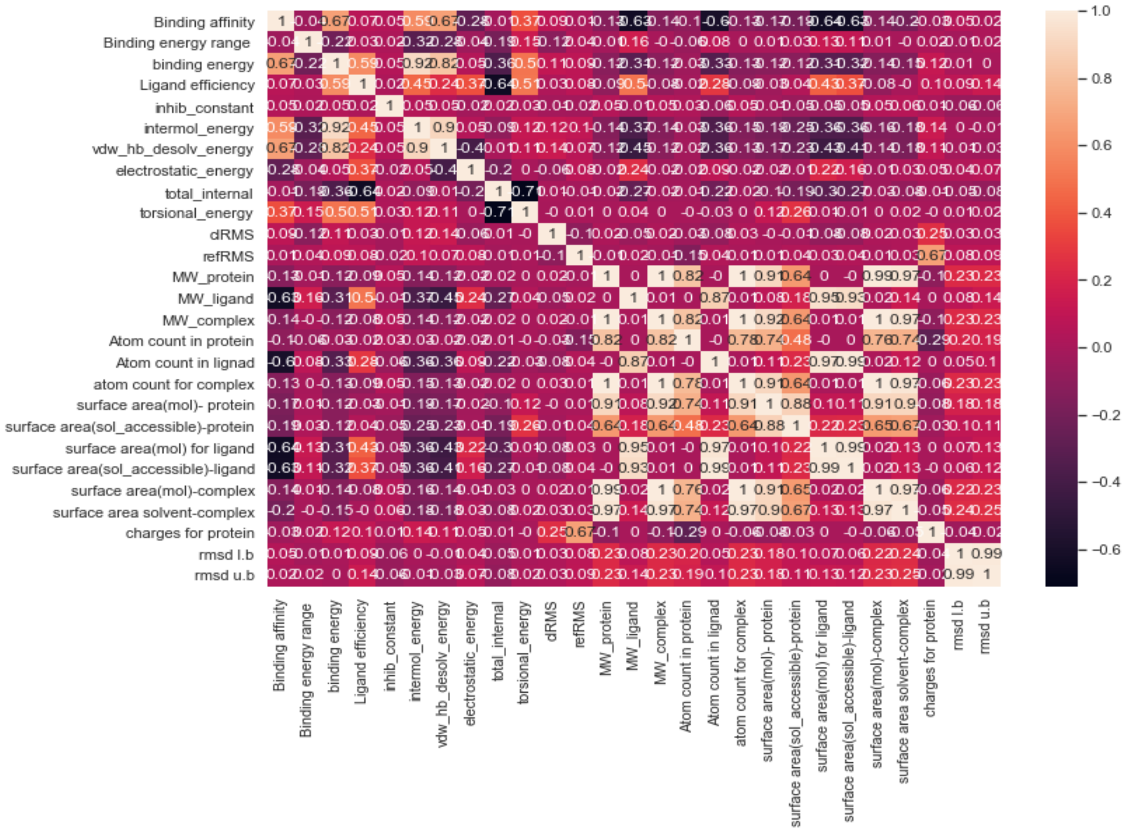


Fig. 4.2 Correlation Matrix of Feature Vectors

Feature importance refers to assigning scores for each input feature that indicates the relative importance of each feature in predictions. The higher the value the most contributive is the feature. Permutation Feature Importance (PFI) calculates relative importance score that is independent of the model used. It works by randomly changing the values of each feature column, one column at a time, and then evaluating the model.

In this work, PFI is used as the feature importance measure by re-estimating the model after permuting one variable. The feature importance based on correlation matrix enables in identifying the contributive feature set with respect to binding affinity. By this way the feature values are validated and the P-value is determined for the contributive feature to discover its relationship with binding affinity. Permutation feature importance of features and the scores for each feature value are shown in Fig. 4.3 and Fig. 4.4 respectively.

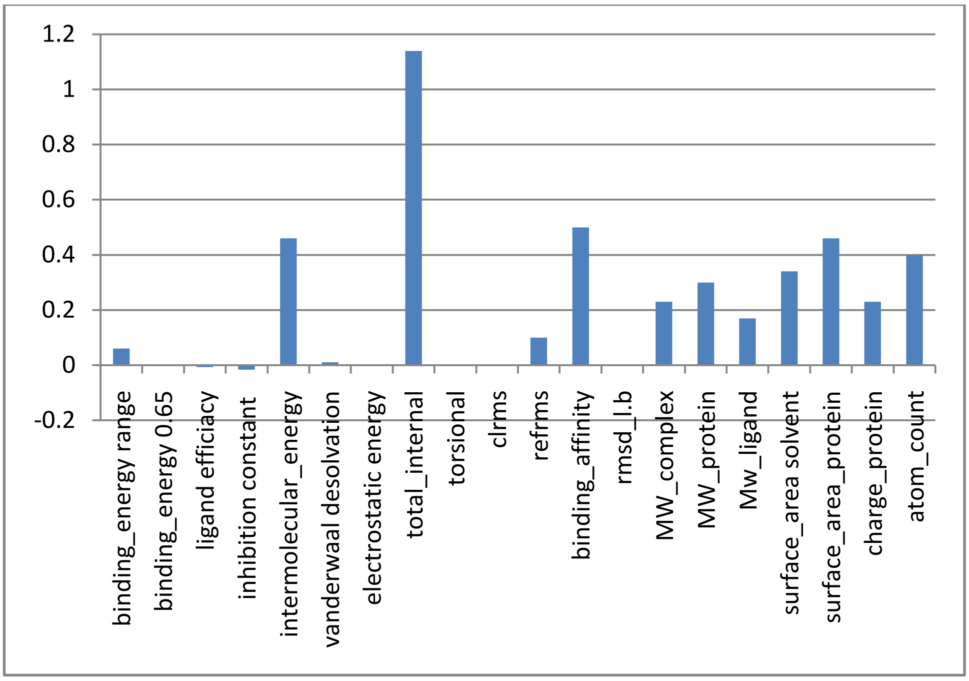
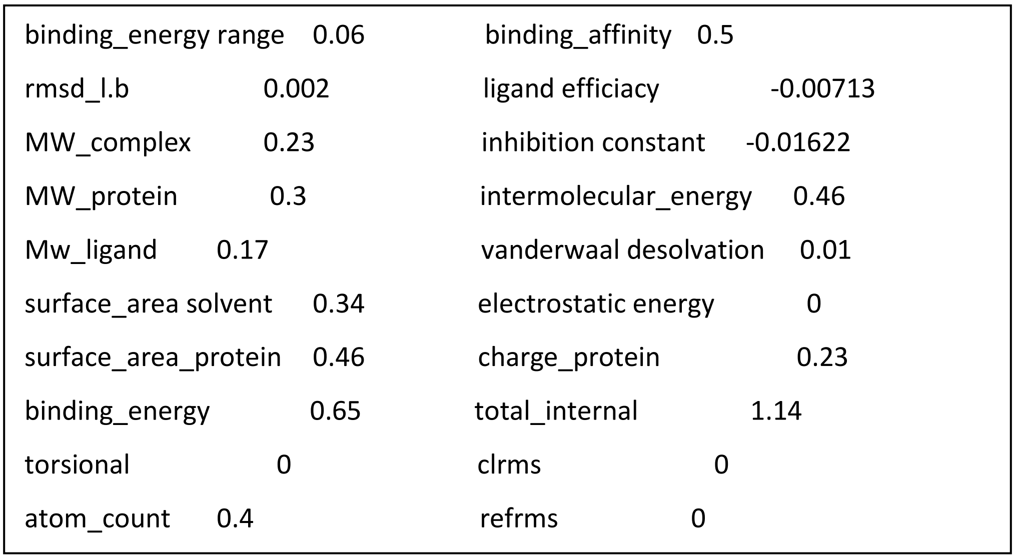


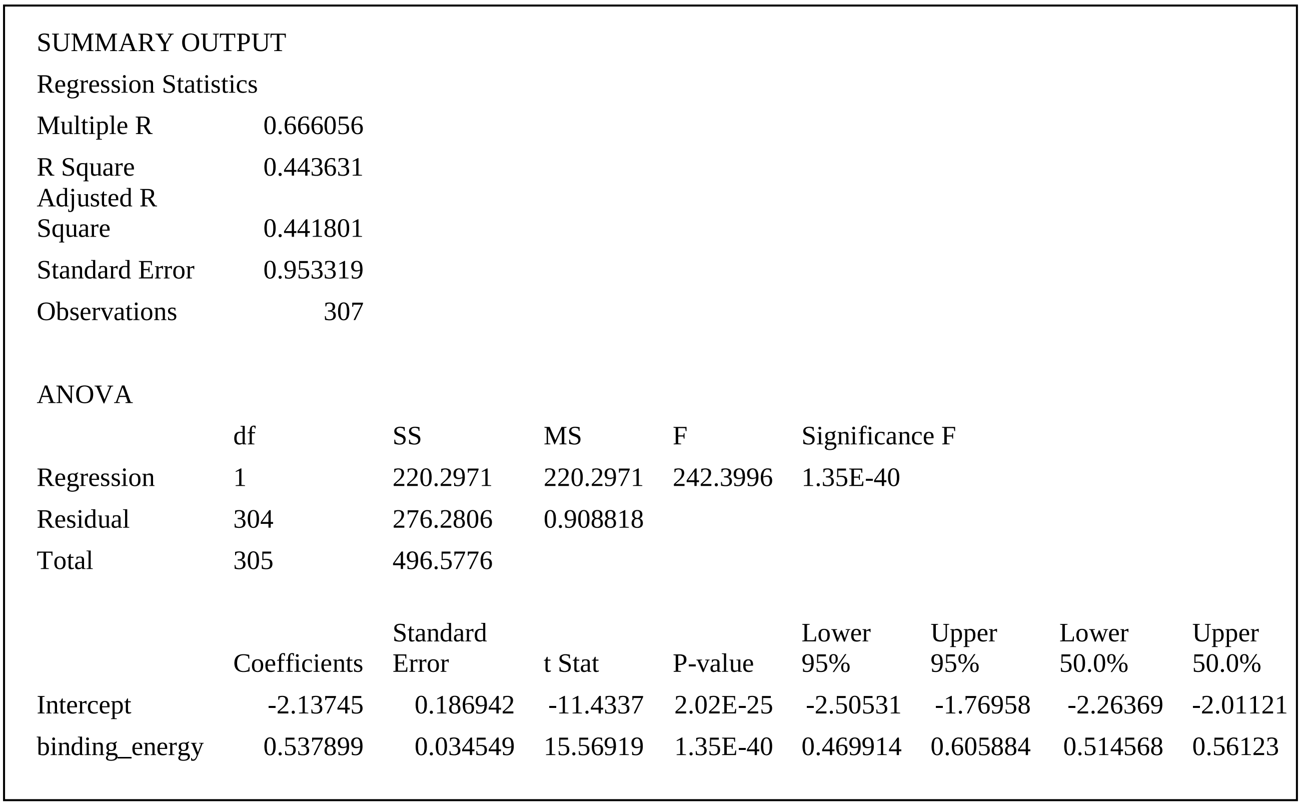
Fig. 4.3 Permutation Feature Importance of Features



binding_energy range	0.06	binding_affinity	0.5
rmsd_l.b	0.002	ligand efficiacy	-0.00713
MW_complex	0.23	inhibition constant	-0.01622
MW_protein	0.3	intermolecular_energy	0.46
Mw_ligand	0.17	vanderwaal desolvation	0.01
surface_area solvent	0.34	electrostatic energy	0
surface_area_protein	0.46	charge_protein	0.23
binding_energy	0.65	total_internal	1.14
torsional	0	clrms	0
atom_count	0.4	refrms	0

Fig. 4.4 Scores of Features

P-value is calculated to reveal how strong the relationship is between dependant variable and independent variable. The feature importance based on correlation matrix shows that the most contributive feature is binding energy. Binding energy is computed based on torsional energy, desolvation energy, ligand efficacy, intermolecular energy, RMSD, electrostatic energy, total internal energy and physical properties. Thus the energy values are important for binding affinity prediction. The P-value obtained is less than 0.05 and it reveals that the relationship between binding affinity and binding energy is strong. The P-value for binding energy and binding affinity is shown in Fig. 4.5.



```
SUMMARY OUTPUT
Regression Statistics
Multiple R          0.666056
R Square            0.443631
Adjusted R
Square              0.441801
Standard Error      0.953319
Observations             307

ANOVA
                 df            SS          MS          F          Significance F
Regression       1             220.2971    220.2971    242.3996   1.35E-40
Residual         304           276.2806    0.908818
Total            305           496.5776

                               Standard                           Lower       Upper       Lower       Upper
                 Coefficients  Error       t Stat      P-value    95%         95%         50.0%       50.0%
Intercept            -2.13745    0.186942  -11.4337    2.02E-25   -2.50531    -1.76958    -2.26369    -2.01121
binding_energy       0.537899    0.034549  15.56919    1.35E-40   0.469914    0.605884    0.514568    0.56123
```

Fig. 4.5 P-value of Binding Affinity with Binding Energy

Features from 307 docked complexes are extracted and feature values are normalized using min-max normalization to scale the values from 0 to 1. Binding affinity values are derived from autodock and augmented with feature vectors. The summary of the above features are portrayed below.

Features	**Count**	**Features**	**Count**
Binding energy range	1	Ligand efficiency	1
Binding energy	1	inhib_constant	1
intermol_energy	1	electrostatic_energy	1
vdw_hb_desolv_energy	1	total_internal	1
torsional_energy	1	clRMS	1
refRMS	1	MW_protein	1
MW_ligand	1	MW_complex	1
Atom count in protein	1	Atom count in ligand	1
Atom count in complex	1	Surface area(mol)- protein	1

Surface area(sol_accessible)protein	1	Surface area(mol) for ligand	1
Surface area(sol_accessible)ligand	1	Surface area(mol)-complex	1
Surface area solvent-complex	1	Charges for protein	1
rmsd u.b	1	rmsd l.b	1
Binding affinity	1	Total	27

Totally 27 features are extracted from each docked complex and PLD dataset with 307 feature vectors of dimension 27 is developed. The feature vector corresponding to the docked complex of 1oa8 with amantadine is given below. The sample dataset is shown in Appendix A.

```
0.671428571     0.101960784 0.41125  0.109090909 0.018654086 0.526315789  0.638586957
0.597894737     1    0.14354067 0 0.06389217    0.213711243     0 0.210970899  0.077147016
0.451612903     0.213884786     0.229693173     0.382196914     0.214054128     0.313410678
0.215396263     0.179535747     0.378378378     0.148810443     0.155097236
```

Model Building

Binding affinity predictive models are built by training the PLD dataset using various regression algorithms such as linear regression, random forest, support vector regression and artificial neural network. Various hyper parameters like number of iterations, learning rate, number of estimators etc., are used here to build the predictive models. The hyper parameters are used for random forest and artificial neural network. The number of estimators is used for random forest as it denotes the number of trees. The parameter number of iterations implies that the number of times the model executes while training. Learning rate parameter is used to control the rate or speed at which the model learns. Tuning of hyper parameter aids in achieving the better prediction rate. The dataset of 307 instances is split into training and testing set where 275 instances for training, 31 instances for testing. The performances of the models are assessed by means of various metrics such as explained variance score, mean squared error, root mean squared error, R2 score, median absolute error and mean absolute error.

Performance metrics like explained variance score and mean squared error are considered as significant metrics in regression task where explained variance score should be higher and the error rate should be low. The other error metrics like root mean squared error, median absolute error and mean absolute error should be minimal. R2 score value should be higher and P-value should be less than 0.05 to determine the relationship stronger. The experimental results and performance analysis of predictive models based on protein-ligand docking dataset is given in section below.

4.2 EXPERIMENT AND RESULTS

Experiments have been carried out by implementing standard regression techniques namely linear regression, support vector regression, artificial neural network and random forest with PLD dataset using the scikit learn tool. Scikit-learn is an open source machine learning library in python. In scikit learn classification, regression and clustering algorithms is built on top of numpy, scipy and matplotlib libraries and also it contains the tools for statistical modelling. The standard 10-fold cross validation technique is used to estimate the collision on the predictive performance. The results attained from the regression models are investigated through performance measures namely explained variance score, mean squared error, root mean squared error, median absolute error and mean absolute error. The results are tabulated in Table 4.1.

Table 4.1 Performance Results of Binding Affinity Predictive Models Based on Protein-Ligand Docking

Regression Algorithms	Explained Variance Score	R2 score	Mean Squared error	Root Mean Squared Error	Median Absolute Error	Mean Absolute error
LR	0.70	0.70	0.32	0.57	0.35	0.23
SVR	0.76	0.76	0.30	0.52	0.30	0.22
RF	**0.85**	**0.85**	**0.20**	**0.44**	**0.25**	**0.15**
ANN	0.82	0.82	0.20	0.44	0.22	0.15

Table 4.1 shows that the results of linear regression predictive model based on protein-ligand docking obtained the explained variance score of 0.70 and mean squared error of 0.32. The results of root mean squared error, median absolute error, mean absolute error and R2 score obtained the values as 0.57, 0.35, 0.23 and 0.70 respectively. The support vector regression predictive model based on protein-ligand docking acquired the explained variance score of 0.76 and mean squared error of 0.30. The results of root mean squared error, median absolute error, mean absolute error and R2 score obtained the values as 0.52, 0.30, 0.22 and 0.76 respectively. The random forest predictive model based on protein-ligand docking yields the explained variance score of 0.85 and mean squared error of 0.20. The results of root mean

squared error, median absolute error, mean absolute error and R2 score obtained the values as 0.44, 0.25, 0.15 and 0.85 respectively. The artificial neural network predictive model based on protein-ligand docking produces the explained variance score of 0.82 and mean squared error of 0.20. The results of root mean squared error, median absolute error, mean absolute error and R2 score obtained the values as 0.44, 0.22, 0.15 and 0.82 respectively. Among all the predictive models based on protein-ligand docking random forest achieves the highest prediction rate and low error rate. Random forest produces efficient results as it acts an estimator algorithm which aggregates the result of many decision trees and then outputs the most optimal result. The other predictive models with linear regression, support vector regression and artificial neural network obtained the lower prediction rate and higher error rate. The performance results of the binding affinity predictive models with PLD dataset to various metrics are portrayed in Fig. 4.6 to Fig. 4.11.

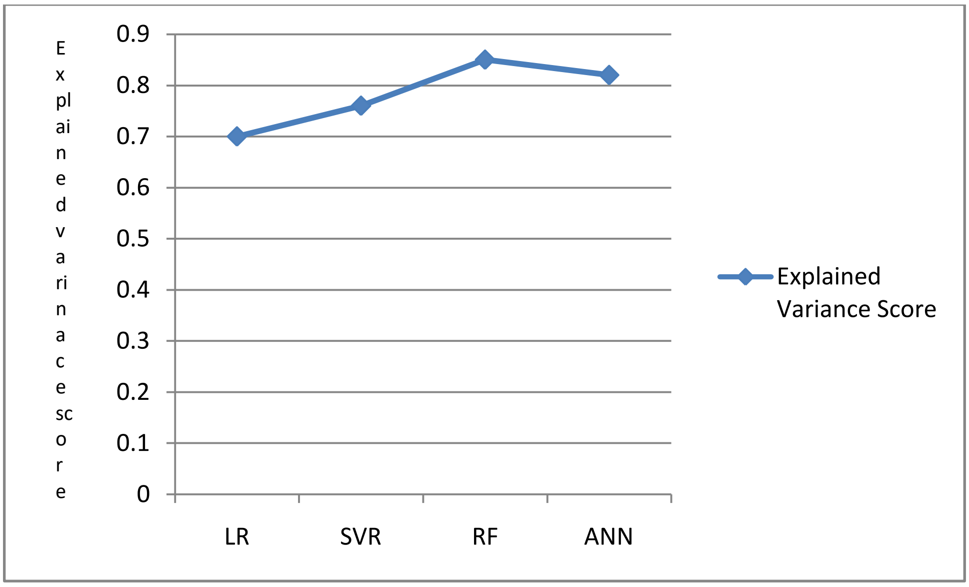


Fig. 4.6 Explained Variance Score of Binding Affinity Predictive Models Based on Protein-Ligand Docking

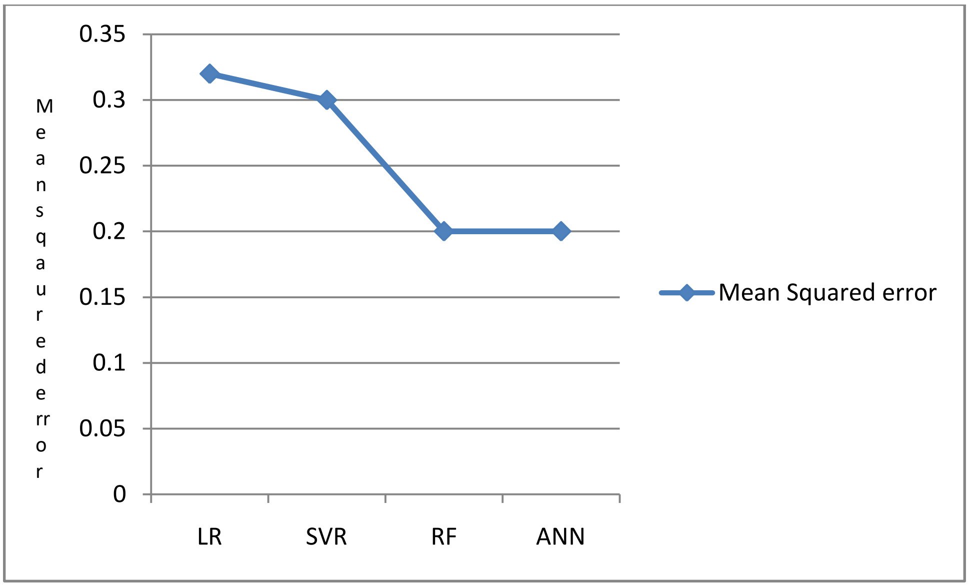


Fig. 4.7 Mean Squared Error of Binding Affinity Predictive Models Based on Protein-Ligand Docking

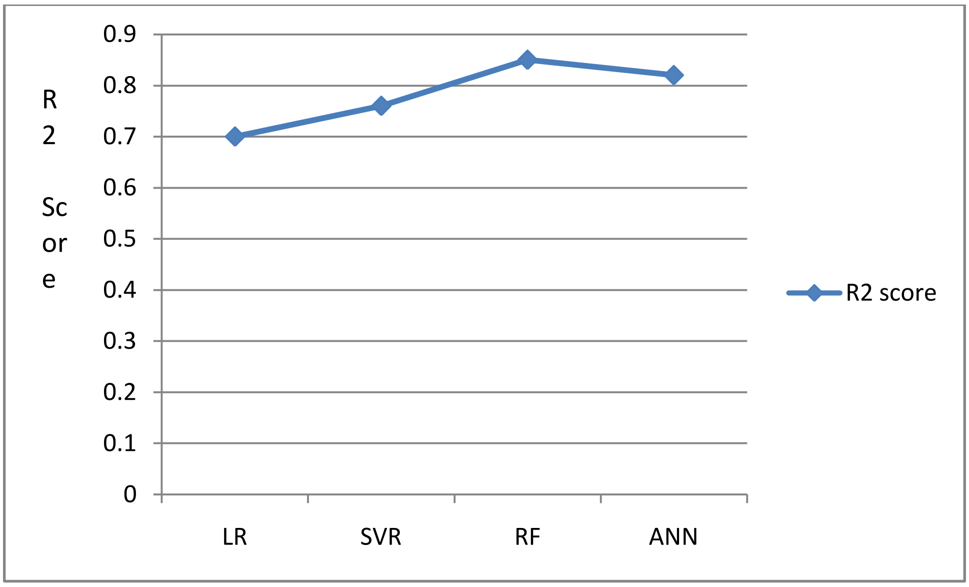


Fig. 4.8 R2 Score of Binding Affinity Predictive Models Based on Protein-Ligand Docking

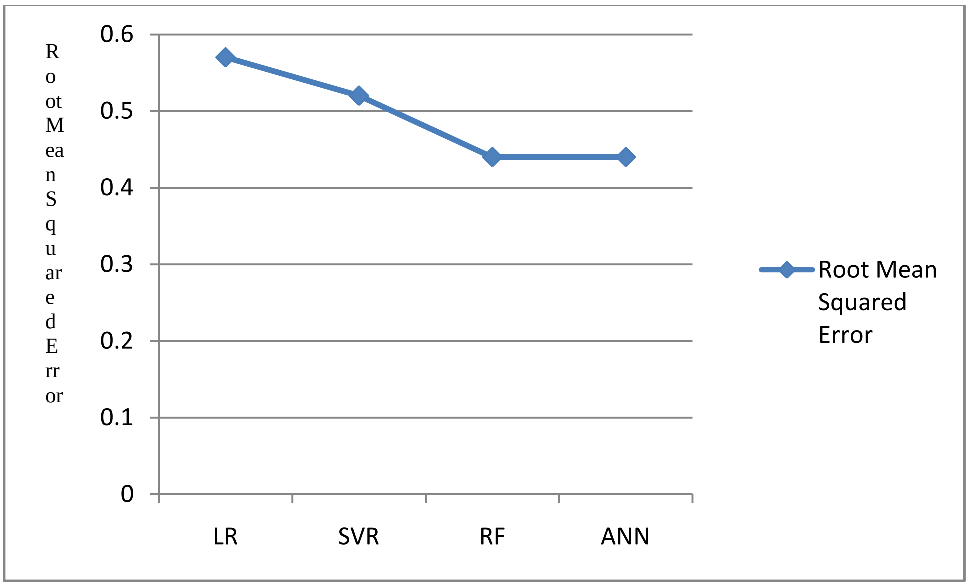


Fig. 4.9 Root Mean Squared Error of Binding Affinity Predictive Models Based on Protein-Ligand Docking

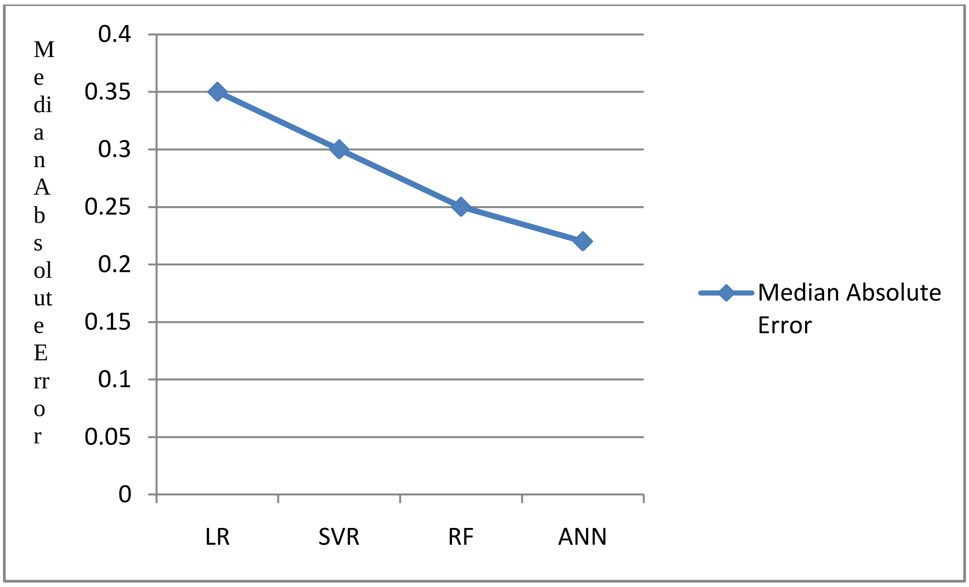


Fig. 4.10 Median Absolute Error of Binding Affinity Predictive Models Based on Protein-Ligand Docking

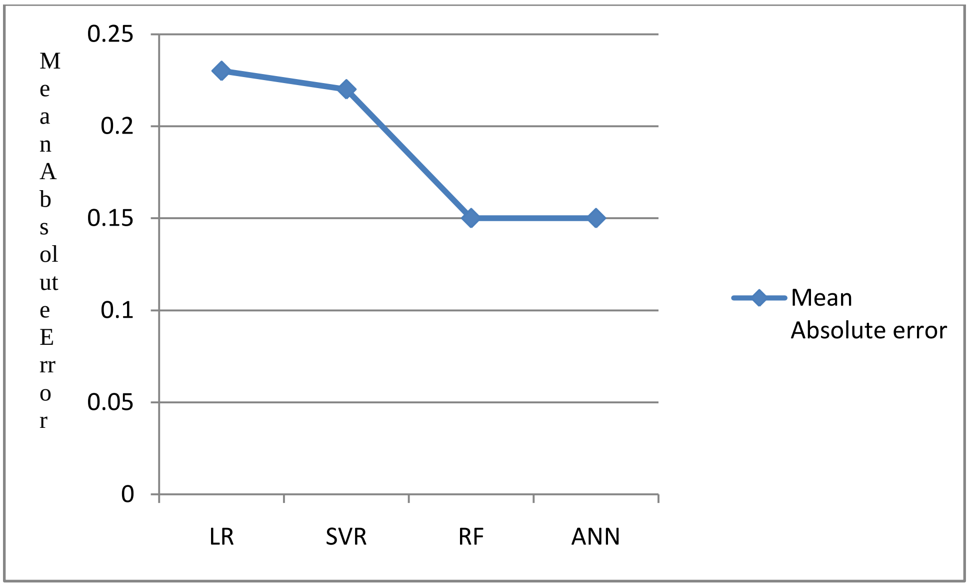


Fig. 4.11 Mean Absolute Error of Binding Affinity Predictive Models Based on Protein-Ligand Docking

From Fig. 4.6, it is observed that the random forest based predictive model achieves higher explained variance score than the other regression algorithms. The Fig. 4.7 reveals that the R2 score curve goes superior for random forest based predictive model and inferior for other predictive models. It is exposed from Fig. 4.8, the random forest based predictive model obtains the low error rate compared to the other regression algorithms. The other error metrics from Fig. 4.9 to Fig. 4.11 discloses that the curve for random forest based predictive model goes inferior in error rate wherein the other predictive models achieve higher error rate. This concludes that evaluation results of random forest based predictive model through protein-ligand docking outperform other predictive models based on linear regression, support vector regression and artificial neural network.

Findings

The experimental results reveals that the features extracted from the docked complexes are exceedingly commits in determining binding affinity. PFI shows the importance of each feature where the binding energy attains the high score. P-value of binding energy with binding affinity shows that the value is less than 0.05 and this reveals that the relationship is strong between binding energy and binding affinity. Random forest based predictive model built with features of protein-ligand docking reveals that explained variance score is higher and the mean squared is low than other regression algorithms. The error rate associated with binding affinity predictive models is less for random forest model and hence it is suitable for prediction of binding affinity for other disorders.

SUMMARY

This chapter illustrated the binding affinity predictive modelling using four different regression tasks. The implementation of various regression techniques for predicting the binding affinity based on protein-ligand dataset have been described in detail. Four independent models have been built and the performances of the models have been reported. The comparative analysis with respect to various evaluation metrics is also presented with tables and charts in this chapter. The development of predictive models to predict binding affinity based on mutations will be discussed in following chapter.

Remarks

The paper titled Binding Affinity Prediction Models for Spinocerebellar Ataxia Using Supervised Learning, has been presented in Second International Conference on Smart Trends in Information Technology and Computer Communications (SMARTCOM), Pune, August 18-19, 2017 and published in springer (CCIS) series, Vol 876, pp 145-152. ***(Scopus indexed)***