7. BINDING AFFINITY PREDICTION USING DEEP NEURAL NETWORK ARCHITECTURES

The second phase of this research work is proceeded by adopting deep learning approach for building predictive models. The binding affinity prediction for SCA needs advanced system which can predict binding affinity more accurately. Binding affinity prediction through general machine learning techniques produces the desired result but still there is a need for a progressive system using deep learning. Machine learning techniques require structured data whereas deep learning relies on layers. The power of deep learning is representation learning wherein it learns the signals between the user defined features and creates a new feature set which gives a better prediction than common machine learning algorithms [93].

In this work, deep learning architectures are employed to predict the binding affinity of SCA as it is very efficient in learning the features and its signal. DNN works by passing the inputs to the input layer and it is passed to the hidden layer. Hidden layer performs mathematical computations on the inputs. The main challenge in defining neural network is to define the number of hidden layers and number of neurons for each layer. The name deep refers it can be more than one hidden layers. The neurons are associated with a weight and the initial weights are set randomly. The weight articulates the importance of input value [94]. In deep learning the model gives better prediction by having many hidden layers and tuning the hyper parameters. DNN models can be more influential in predicting binding affinity from normal complexes and mutation induced complexes as they are capable of discovering the complex signalling and interaction between features.

This chapter exhibits the modelling of binding affinity using deep learning architecture with three datasets such as PLD, PMLD and PPD. Three different architectures like sequential DNN, functional DNN and DNN with customized layers are used to build the predictive models. The development of the prediction models using variants of DNN with various optimizers are discussed briefly. The training of deep models performed through tuning the hyper parameters is also described. The performance evaluation of the models with various metrics and their comparison results with different optimizers are explained in this chapter. The comparative results of three deep neural network based affinity predictive models are also presented.

7.1 BINDING AFFINITY PREDICTION MODELS USING SEQUENTIAL DEEP NEURAL NETWORK

The basic DNN uses a stack of layers called sequential deep neural network and it is implemented serially. Sequential model is a plain stack of layers where each layer has one input tensor and one output tensor. This section explains the modelling of binding affinity prediction using sequential DNN with three independent datasets PLD, PMLD, PPD and three optimizers such as adam, RMSprop and Nadam. The problem is modelled as regression task and the predictive models are built by tuning the hyper parameters. Hyper parameters formulate the model to obtain accurate prediction of binding affinity.

Methodology

The sequential DNN is employed to construct binding affinity predictive models through self learning the features using three datasets with three optimizers. Various hyper parameters such as epochs, dropouts, learning rate, optimizers, loss function and activation function are defined appropriately to strengthen the architecture. The proposed framework of sequential DNN to build binding affinity predictive model is shown in Fig. 7.1.

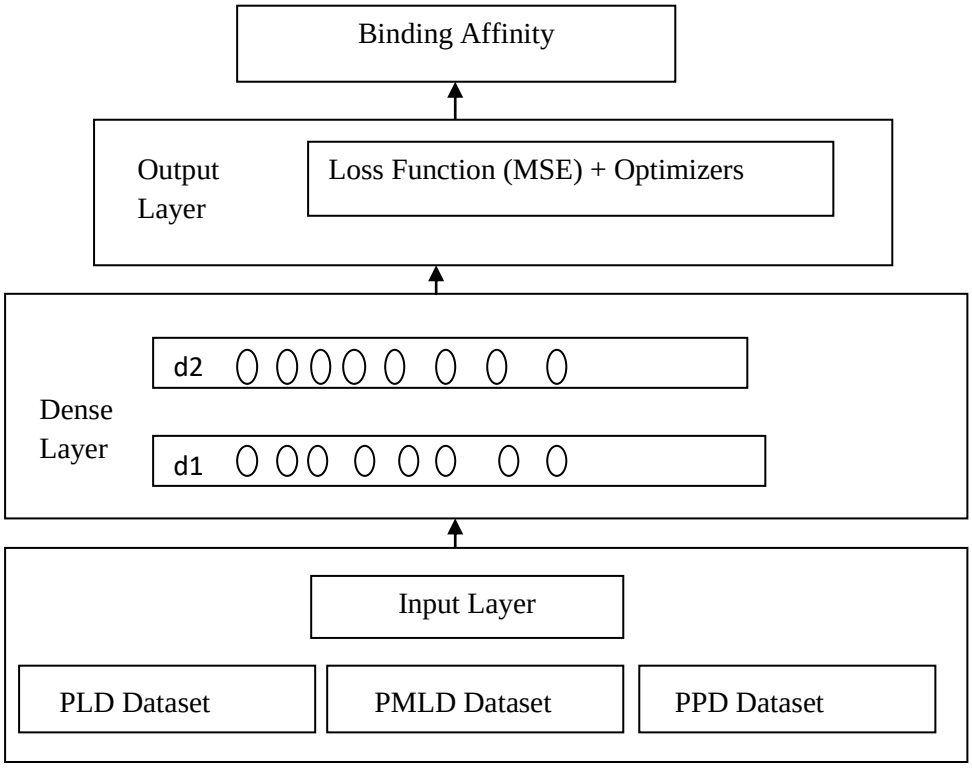

Fig. 7.1 Architecture of Sequential DNN Based Prediction Model

The input feature vectors created as described in chapter 4, 5 and 6 corresponding to three datasets such as PLD, PMLD and PPD are directly fed as input to the sequential DNN architecture. The proposed sequential DNN consists of one input layer, number of hidden layers 2 and one output layer. The input dataset is given as input dimension where output values are computed layer by layer in the network. When the input is given, it gets multiplied by weight value and the output is passed to the next layer. Initially random weights are applied to the neurons in each layer and adjusted when the error rate occurs. The error rate is the difference between actual and predicted value, the weights are adjusted to minimize the error and are propagated backward through the network after calculation. The hidden layer employs a rectified linear unit (relu) activation function which is applied to the sum of the weighted values to compute the output values. The inputs in the layer are modified by layer wise to get the new abstractions. Here the fully interconnected dense layer with one neuron in output layer is used for prediction.

In the binding affinity prediction problem, the model consists of one neuron in the output layer whereas 20 neurons in first hidden layer and 10 neurons in second hidden layer. The output layer is defined with relu activation function that allows the network to learn and output the predicted values. The layers are added and to improve the prediction rate various hyper parameters such as epochs, dropouts, learning rate, batch size and optimizers etc., are examined. The hyper parameters are defined as given below.

Loss Function: The network used the mean squared loss function while training, suitable for prediction problems. Mean squared error is calculated as the average of the squared differences between the predicted and actual values. The MSE loss reaches its minimum value at prediction. The range is 0 to ∞.

Activation Function: It is used to introduce nonlinearity to models, which allows deep learning models to learn nonlinear prediction boundaries. Relu activation function is applied to the sum of the weighted values to compute output values.

Optimizers: The network uses three different optimizers Adam, RMSprop and Nadam. The efficient optimization algorithm is Adam. This optimizer is extension of stochastic gradient optimization algorithm and Adam optimization algorithm computationally updates network weights iterative based on training data. It is well suited for problems that are large in terms of data and typically require little tuning of hyper parameters. Instead of adapting the parameter learning rates based on the average first moment, Adam also makes use of the average of the second moments of the gradients. Specifically, the algorithm calculates an exponential moving average of the gradient and the squared gradient, and the

parameter beta control the decay rates of these moving averages. The prediction metrics are reported for each training epoch to give an idea of the skill of the model in addition to the loss. RMSprop optimizer is an optimizer that utilizes the magnitude of recent gradients to normalize the gradients. Learning rate will be tuned in this optimizer. Nadam optimizer is Nesterov Adam optimizer. Optimization algorithm enables to minimize error function and update the network weight. Three optimizers like Adam, RMSProp and Nadam are used here to change the parameters in the neural network such as weights and learning rate to reduce the losses.

Learning Rate: Learning rate is the parameter that indicates the optimizer to maneuver the weights within the direction opposite of the gradient for a mini-batch. Different values of 0.1, 0.01 and 0.001 are tested to search out the one that offers the simplest loss. Training with a smaller learning rate decreases the loss within the first few iteration. Hence 0.01 was mounted as the learning rate.

Dropout: It is a regularization parameter that randomly skips neurons during training, forcing others in the layer to pick up the slack. The contribution of the detached neurons to the activation of downstream neurons is temporally removed on the forward pass and any weight updates are not applied to the neuron on the backward pass. Dropout is implemented by randomly selecting nodes to be dropped-out with a given probability in each weight update cycle. The model was tested with different dropout percentages varying from 20- 50% and the results are recorded.

Epochs: The epochs are the number of times the DNN will work through the training data set. Various epochs size is used in this network are 50, 100, 200, 400, 500.

The above hyper parameters are defined in order to achieve better learning and to produce accurate prediction rate. The network is trained using three optimizers such as Adam, RMSprop, Nadam with three different data sets PLD, PMLD, PPD.

Finally, 10-fold cross validation technique is used to evaluate the performance of the models based on various metrics such as explained variance score, mean squared error, root mean squared error, mean absolute error, median absolute error and R2 score.

Experiments and Results

The experiments have been carried out by employing sequential DNN using keras as front end which is high level API and tensorflow as backend in jupyter notebook and it is coded in python. Since the backend is tensorflow the libraries of tensorflow are used while implementing the code. Keras is chosen because it extends for custom building blocks to create new ideas. Keras is user –friendly and modular where the models are made by

connecting blocks to each other. It is easy to use, where keras offers simple APIs that helps in minimizing the number of user actions. Keras in deep learning allows easy and fast prototyping and it can run on many backends like theano, CNTK etc. The values set for hyper parameters in sequential DNN architecture are given in Table 7.1.

Table 7.1 Values of Hyperparameters for Sequential DNN

Hyperparameters	**Values**
Optimizers	Adam, RMSprop, Nadam
Learning Rate	0.01
Dropout	0.2-0.5
Epochs	50-500
Activation Function	Relu

Sequential models are trained with the above parameter settings using three datasets PLD, PMLD and PPD and various models are built for predicting the binding affinity. The datasets are split into training and testing set wherein the 90% percent of data is used for training and 10% of data is used for testing. The 10-fold cross validation technique is used to estimate prediction performance of the model. Independent models have been evaluated using important metrics such as explained variance score and mean squared error to monitor the consistency of the models where as other metrics such as mean squared error, root mean squared error, R2 score, mean absolute error and median absolute error are also calculated.

Results for Protein-Ligand Dataset

The results of sequential predictive models based on protein-ligand dataset with three optimizers such as adam, RMSprop, Nadam for various dropouts are presented below. The results of sequential DNN binding affinity predictive model with adam optimizer is presented in Table 7.2.

Table 7.2 Results of Sequential DNN Based Predictive Model with Adam Optimizer for Protein-Ligand Dataset

Dropout	**Epochs**	Explained variance score	R2 score	Mean squared error	Root mean squared error	Mean absolute error	Median absolute error
0.2	50	0.68	0.68	0.45	0.67	0.49	0.34
	100	0.70	0.70	0.32	0.57	0.35	0.23
	200	0.75	0.75	0.30	0.59	0.39	0.27
	400	0.80	0.80	0.20	0.44	0.28	0.19

	500	**0.94**	**0.94**	**0.13**	**0.36**	**0.15**	**0.14**
0.3	50	0.70	0.70	0.32	0.57	0.35	0.23
	100	0.75	0.75	0.30	0.59	0.39	0.27
	200	0.82	0.82	0.20	0.44	0.22	0.15
	400	0.90	0.90	0.12	0.34	0.15	0.6
	500	**0.95**	**0.95**	**0.1**	**0.32**	**0.12**	**0.4**
0.4	50	0.75	0.75	0.30	0.59	0.39	0.27
	100	0.82	0.82	0.20	0.44	0.22	0.15
	200	0.86	0.86	0.2	0.44	0.22	0.15
	400	0.89	0.89	0.2	0.44	0.24	0.10
	500	**0.90**	**0.90**	**0.15**	**0.34**	**0.15**	**0.6**
0.5	50	0.75	0.75	0.30	0.59	0.39	0.27
	100	0.82	0.82	0.20	0.44	0.22	0.15
	200	0.90	0.90	0.12	0.34	0.15	0.6
	400	0.80	0.80	0.20	0.44	0.28	0.19
	500	**0.92**	**0.92**	**0.14**	**0.32**	**0.12**	**0.4**

The experiment is carried out for protein-ligand dataset with different epochs and dropouts and the results showed that sequential DNN based prediction model with adam optimizer achieved the prediction rate and error rate for the dropout 0.3 and epoch 500 is 0.95 and 0.1 respectively. The results of mean absolute error, median absolute error, R2 score and root mean squared error are 0.12, 0.4, 0.95 and 0.32 respectively. Prediction rate and error rate obtained for the dropout 0.2 and epoch 500 is of 0.94 and 0.13 respectively which is better when compared to other dropout values and epochs. The results of root means squared error, mean absolute error, median absolute error and R2 score are 0.36, 0.15, 0.14 and 0.94 respectively. The results of sequential DNN with RMSprop optimizer for protein-ligand dataset are presented in Table 7.3.

Table 7.3 Results of Sequential DNN Based Predictive Model with RMSprop Optimizer for Protein-Ligand Dataset

Dropout	**Epochs**	Explained variance score	R2 score	Mean squared error	Root mean squared error	Mean absolute error	Median absolute error
0.2	50	0.68	0.68	0.45	0.67	0.49	0.34
	100	0.70	0.70	0.32	0.57	0.35	0.23
	200	0.72	0.72	0.32	0.54	0.34	0.24
	400	0.76	0.76	0.30	0.52	0.30	0.22
	500	**0.85**	**0.85**	**0.20**	**0.44**	**0.25**	**0.15**
0.3	50	0.70	0.70	0.32	0.57	0.35	0.23
	100	0.75	0.75	0.30	0.59	0.39	0.27

	200	0.82	0.82	0.20	0.44	0.22	0.15
	400	0.85	0.85	0.20	0.44	0.25	0.15
	500	**0.92**	**0.92**	**0.14**	**0.32**	**0.12**	**0.4**
0.4	50	0.75	0.75	0.30	0.59	0.39	0.27
	100	0.82	0.82	0.20	0.44	0.22	0.15
	200	0.86	0.86	0.2	0.44	0.22	0.15
	400	0.89	0.89	0.2	0.44	0.24	0.10
	500	**0.90**	**0.90**	**0.15**	**0.34**	**0.15**	**0.6**
0.5	50	0.75	0.75	0.30	0.59	0.39	0.27
	100	0.82	0.82	0.20	0.44	0.22	0.15
	200	0.90	0.90	0.12	0.34	0.15	0.6
	400	0.85	0.85	0.20	0.44	0.25	0.15
	500	**0.89**	**0.89**	**0.18**	**0.35**	**0.16**	**0.8**

The experimental result reveals that the prediction rate and error rate obtained for the dropout 0.3 and epoch 500 is 0.92 and 0.14 respectively. The results of error metrics acquired for root mean squared error, mean absolute error, median absolute error, R2 score are 0.32, 0.12, 0.4 and 0.92 respectively. The next highest prediction rate obtained for the dropout 0.4 and epoch 500 is 0.90. The results for mean squared error, root mean squared error, mean absolute error, median absolute error, R2 score are 0.15, 0.34, 0.15, 0.6 and 0.90 obtained for the dropout 0.4 and epoch 500 respectively. The result of sequential DNN with Nadam optimizer for protein-ligand dataset is given in Table 7.4.

Table 7.4 Results of Sequential DNN Based Predictive Model with Nadam Optimizer for Protein-Ligand Dataset

Dropout	**Epochs**	Explained variance score	R2 score	Mean squared error	Root mean squared error	Mean absolute error	Median absolute error
0.2	50	0.68	0.68	0.45	0.67	0.49	0.34
	100	0.70	0.70	0.32	0.57	0.35	0.23
	200	0.72	0.72	0.32	0.54	0.34	0.24
	400	0.76	0.76	0.30	0.52	0.30	0.22
	500	**0.85**	**0.85**	**0.20**	**0.44**	**0.25**	**0.15**
0.3	50	0.70	0.70	0.32	0.57	0.35	0.23
	100	0.75	0.75	0.30	0.59	0.39	0.27
	200	0.82	0.82	0.20	0.44	0.22	0.15
	400	0.85	0.85	0.20	0.44	0.25	0.15
	500	**0.90**	**0.90**	**0.15**	**0.34**	**0.15**	**0.6**

0.4	50	0.75	0.75	0.30	0.59	0.39	0.27
	100	0.82	0.82	0.20	0.44	0.22	0.15
	200	0.86	0.86	0.2	0.44	0.22	0.15
	400	0.89	0.89	0.2	0.44	0.24	0.10
	500	**0.90**	**0.90**	**0.15**	**0.34**	**0.15**	**0.6**
0.5	50	0.75	0.75	0.30	0.59	0.39	0.27
	100	0.82	0.82	0.20	0.44	0.22	0.15
	200	0.90	0.90	0.12	0.34	0.15	0.6
	400	0.85	0.85	0.20	0.44	0.25	0.15
	500	**0.89**	**0.89**	**0.18**	**0.35**	**0.16**	**0.8**

Binding affinity predictive model based on sequential deep neural network with Nadam optimizer attains better results for the dropout 0.3 and epoch 500. The model attains the same explained variance score of 0.90 with both 0.3 and 0.4 dropouts. The results of mean squared error, root mean squared error, mean absolute error, median absolute error, R2 score obtained are 0.15, 0.34, 0.15, 0.6 and 0.90 respectively. The results of sequential DNN models obtained for different dropouts and epochs are illustrated from Fig. 7.2 to Fig. 7.4.

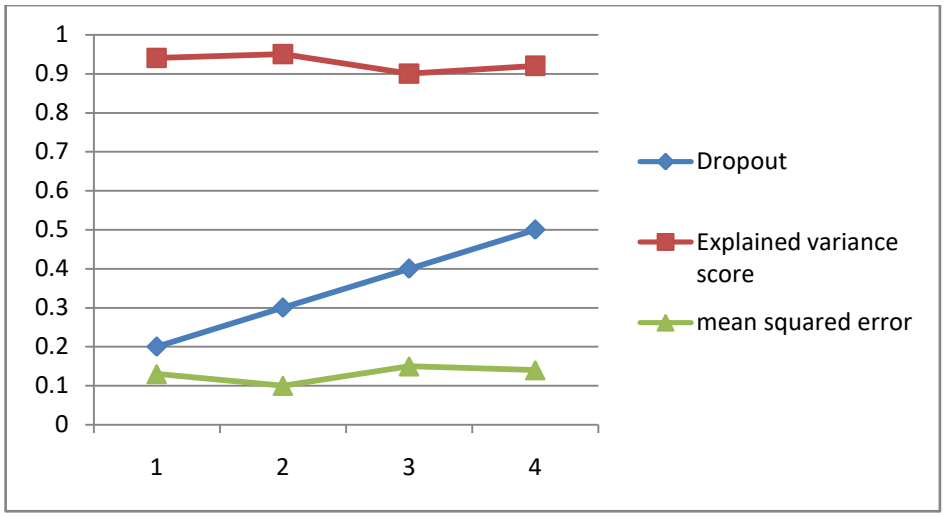

Fig. 7.2 Results of Sequential DNN Based Predictive Model with Adam Optimizer at Various Dropouts for Protein-Ligand Dataset

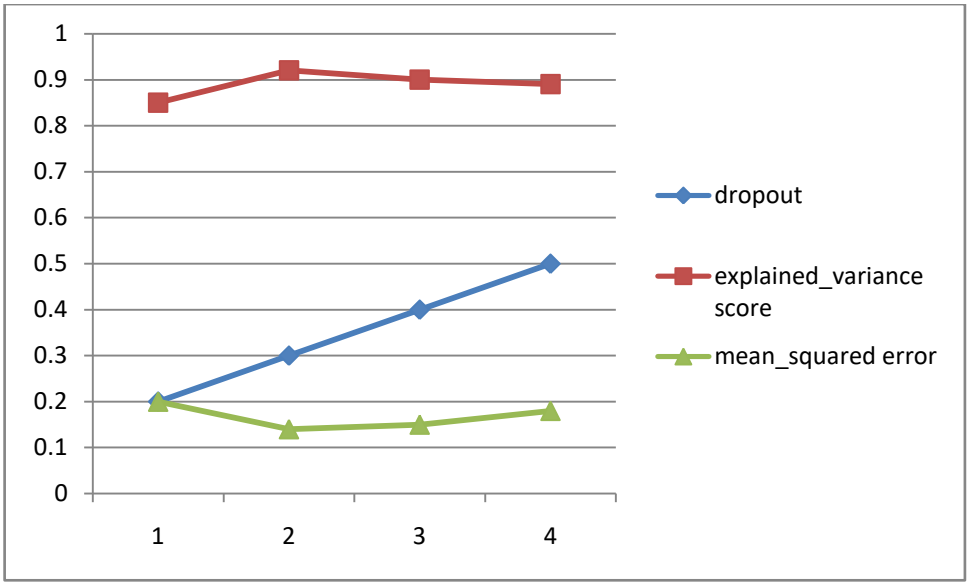

Fig. 7.3 Results of Sequential DNN Based Predictive Model with RMSprop Optimizer at Various Dropouts for Protein-Ligand Dataset

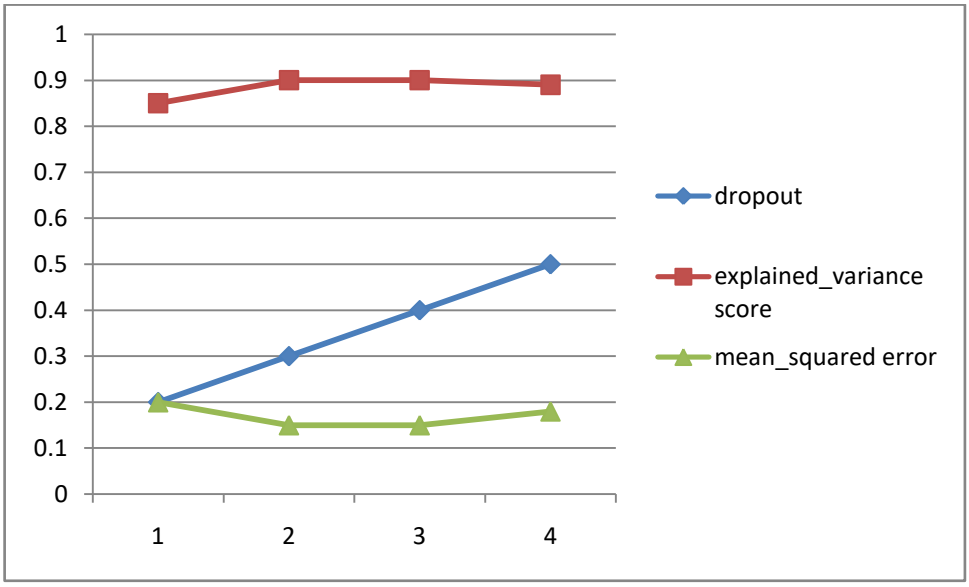

Fig. 7.4 Results of Sequential DNN Based Predictive Model with Nadam Optimizer at Various Dropouts for Protein-Ligand Dataset

The charts from Fig. 7.2 to Fig. 7.4 confirms that the predictive models with three different optimizers such as adam, RMSprop and Nadam obtained the highest prediction rate 0.95, 0.92 and 0.90 respectively. The sequential DNN models obtained the lowest error rate of 0.1, 0.14 and 0.15 at dropout of 0.3 and epoch 500 for the three optimizers. The explained variance score of sequential DNN with RMSprop optimizer is higher than the model with Nadam optimizer. When comparing the sequential models the dropout 0.4 and 0.5 of three different optimizers are inferior to the dropout 0.3. The error rate of sequential models with

three optimizers 0.2, 0.4 and 0.5 is superior to the dropout 0.3. This reveals that the sequential models based on protein-ligand docking dataset with adam optimizer performs better than RMSprop and Nadam optimizers.

Results for Protein-Mutated-Ligand Dataset

The results of sequential predictive models based on protein-mutated-ligand dataset with three optimizers such as adam, RMSprop, Nadam for various dropouts are presented below. The results of sequential predictive models with adam optimizer for protein-mutated-ligand dataset is tabulated in Table 7.5.

Table 7.5 Results of Sequential DNN Based Predictive Model with Adam Optimizer for Protein-Mutated-Ligand Dataset

Dropout	**Epochs**	Explained variance score	R2 score	Mean squared error	Root mean squared error	Mean absolute error	Median absolute error
0.2	50	0.68	0.68	0.45	0.67	0.49	0.34
	100	0.70	0.70	0.32	0.57	0.35	0.23
	200	0.75	0.75	0.30	0.59	0.39	0.27
	400	0.80	0.80	0.20	0.44	0.28	0.19
	500	**0.82**	**0.82**	**0.20**	**0.44**	**0.22**	**0.15**
0.3	50	0.70	0.70	0.32	0.57	0.35	0.23
	100	0.75	0.75	0.30	0.59	0.39	0.27
	200	0.82	0.82	0.20	0.44	0.22	0.15
	400	0.75	0.75	0.30	0.59	0.39	0.27
	500	**0.85**	**0.85**	**0.20**	**0.44**	**0.25**	**0.15**
0.4	50	0.75	0.75	0.30	0.59	0.39	0.27
	100	0.82	0.82	0.20	0.44	0.22	0.15
	200	0.82	0.82	0.20	0.44	0.22	0.15
	400	0.80	0.80	0.20	0.44	0.28	0.19
	500	**0.85**	**0.85**	**0.20**	**0.44**	**0.25**	**0.15**
0.5	50	0.75	0.75	0.30	0.59	0.39	0.27
	100	0.82	0.82	0.20	0.44	0.22	0.15
	200	0.82	0.82	0.20	0.44	0.22	0.15
	400	0.80	0.80	0.20	0.44	0.28	0.19
	500	**0.82**	**0.82**	**0.20**	**0.44**	**0.22**	**0.15**

Sequential predictive model for protein-mutated-ligand dataset with adam optimizer obtains the prediction rate and error rate for dropout of 0.3, 0.4 and epoch 500 is 0.85 and 0.20 respectively. The results for root mean squared error, mean absolute error, median absolute error and R2 score obtained are 0.44, 0.25, 0.15 and 0.85 respectively for both the

dropouts. The explained variance score and error rate for dropout 0.2, 0.5 are 0.82 and 0.20 respectively. The results for root mean squared error, mean absolute error, median absolute error and R2 score acquired are 0.44, 0.22, 0.15 and 0.82 respectively. The results of sequential predictive models with RMSprop optimizer for protein-mutated-ligand dataset are tabulated in Table 7.6.

Table 7.6 Results of Sequential DNN Based Predictive Model with RMSprop Optimizer for Protein-Mutated-Ligand Dataset

Dropout	**Epochs**	Explained variance score	R2 score	Mean squared error	Root mean squared error	Mean absolute error	Median absolute error
0.2	50	0.68	0.68	0.45	0.67	0.49	0.34
	100	0.70	0.70	0.32	0.57	0.35	0.23
	200	0.75	0.75	0.30	0.59	0.39	0.27
	400	0.75	0.75	0.30	0.59	0.39	0.27
	500	**0.76**	**0.76**	**0.29**	**0.57**	**0.35**	**0.25**
0.3	50	0.70	0.70	0.32	0.57	0.35	0.23
	100	0.68	0.68	0.45	0.67	0.49	0.34
	200	0.70	0.70	0.32	0.57	0.35	0.23
	400	0.70	0.70	0.32	0.57	0.35	0.23
	500	**0.78**	**0.78**	**0.30**	**0.59**	**0.39**	**0.27**
0.4	50	0.75	0.75	0.30	0.59	0.39	0.27
	100	0.78	0.78	0.30	0.59	0.39	0.27
	200	0.78	0.78	0.30	0.59	0.39	0.27
	400	0.75	0.75	0.30	0.59	0.39	0.27
	500	**0.78**	**0.78**	**0.30**	**0.59**	**0.39**	**0.27**
0.5	50	0.75	0.75	0.30	0.59	0.39	0.27
	100	0.75	0.75	0.30	0.59	0.39	0.27
	200	0.75	0.75	0.30	0.59	0.39	0.27
	400	0.78	0.78	0.30	0.59	0.39	0.27
	500	**0.78**	**0.78**	**0.30**	**0.59**	**0.39**	**0.27**

The results of sequential predictive model with RMSprop optimizer reveal that the various dropouts 0.3, 0.4, 0.5 and epoch 500 produced the same result for the prediction rate and the error rate as 0.78 and 0.30 respectively. The results for root mean squared error, mean absolute error, median absolute error and R2 score acquired are 0.59, 0.39, 0.27 and 0.78

respectively. The prediction rate and error rate for the dropout 0.2 is 0.76 and 0.29 respectively. The results of root mean squared error, mean absolute error, median absolute error and R2 score for the dropout 0.2 obtained is 0.57, 0.35, 0.25 and 0.76 respectively. The results of sequential predictive models with Nadam optimizer for protein-mutated-ligand dataset are summarized in Table 7.7.

Table 7.7 Results of Sequential DNN Based Predictive Model and Nadam Optimizer for Protein-Mutated-Ligand Dataset

Dropout	**Epochs**	Explained variance score	R2 score	Mean squared error	Root mean squared error	Mean absolute error	Median absolute error
0.2	50	0.68	0.68	0.45	0.67	0.49	0.34
	100	0.70	0.70	0.32	0.57	0.35	0.23
	200	0.75	0.75	0.30	0.59	0.39	0.27
	400	0.75	0.75	0.30	0.59	0.39	0.27
	500	**0.76**	**0.76**	**0.29**	**0.57**	**0.35**	**0.25**
0.3	50	0.70	0.70	0.32	0.57	0.35	0.23
	100	0.68	0.68	0.45	0.67	0.49	0.34
	200	0.70	0.70	0.32	0.57	0.35	0.23
	400	0.70	0.70	0.32	0.57	0.35	0.23
	500	**0.78**	**0.78**	**0.30**	**0.59**	**0.39**	**0.27**
0.4	50	0.75	0.75	0.30	0.59	0.39	0.27
	100	0.78	0.78	0.30	0.59	0.39	0.27
	200	0.78	0.78	0.30	0.59	0.39	0.27
	400	0.75	0.75	0.30	0.59	0.39	0.27
	500	**0.78**	**0.78**	**0.30**	**0.59**	**0.39**	**0.27**
0.5	50	0.75	0.75	0.30	0.59	0.39	0.27
	100	0.75	0.75	0.30	0.59	0.39	0.27
	200	0.75	0.75	0.30	0.59	0.39	0.27
	400	0.78	0.78	0.30	0.59	0.39	0.27
	500	**0.78**	**0.78**	**0.30**	**0.59**	**0.39**	**0.27**

The results of sequential DNN with Nadam optimizer for PMLD dataset produce the same results of RMSprop optimizer. The results obtained for different dropouts and epochs are illustrated from Fig. 7.5 to Fig. 7.7.

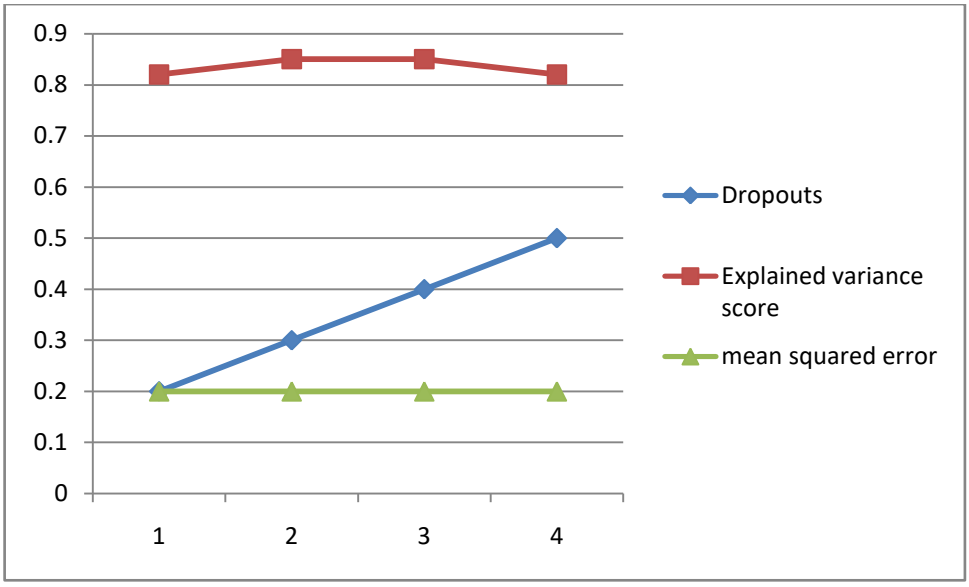

Fig. 7.5 Results of Sequential DNN Based Predictive Model with Adam Optimizer at Various Dropouts for Protein-Mutated-Ligand Dataset

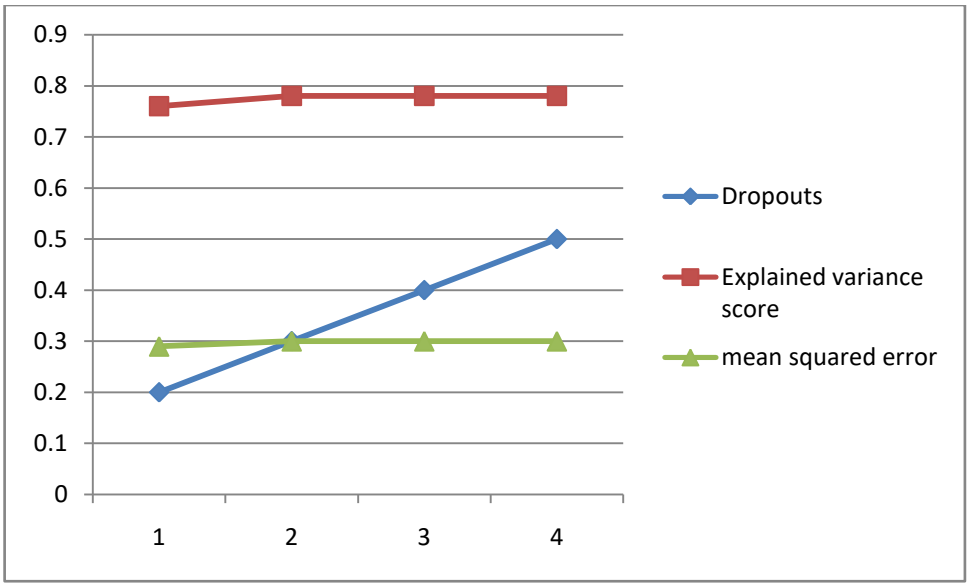

Fig. 7.6 Results of Sequential DNN Based Predictive Model with RMSprop Optimizer at Various Dropouts for Protein-Mutated-Ligand Dataset

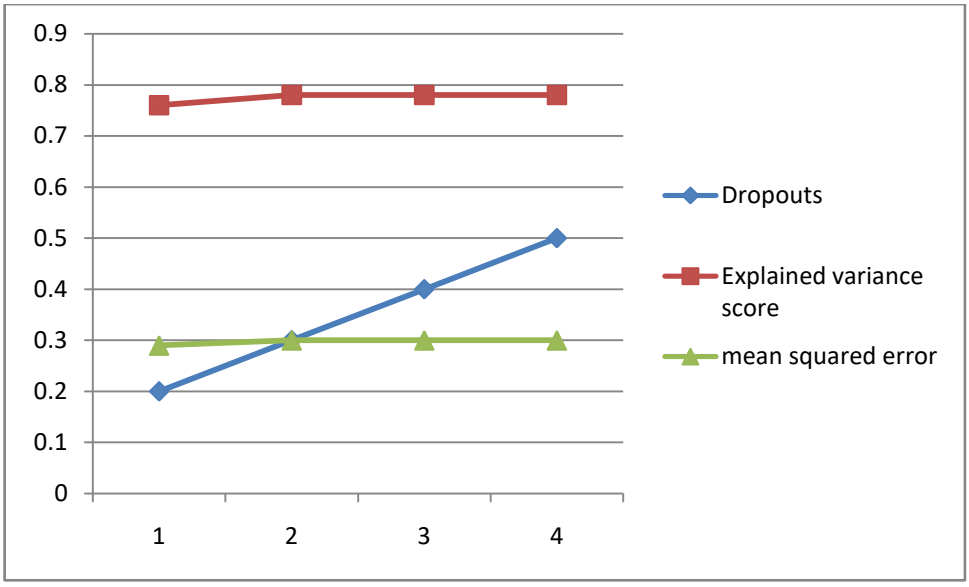

Fig 7.7 Results of Sequential DNN Based Predictive Model with Nadam Optimizer at Various Dropouts for Protein-Mutated-Ligand Dataset

The charts from Fig. 7.5 to Fig. 7.7 demonstrates that the predictive models with sequential DNN for protein-mutated-ligand dataset with three different optimizers such as adam, RMSprop and Nadam obtained the prediction rate 0.85, 0.78 and 0.78 respectively. The lowest error rate obtained for the three optimizers are 0.2, 0.3, 0.3 at dropout 0.3 and epoch 500. Here the Nadam optimizer obtains the same result as that of RMSprop optimizer. The result shows that the sequential DNN models based on protein-mutated-ligand docking dataset with adam optimizer achieves better prediction rate than RMSprop and Nadam optimizers.

Results for Protein-protein Dataset

The results of sequential predictive models based on protein-protein dataset with three optimizers such as adam, RMSprop, Nadam for various dropouts are presented below. The results of sequential DNN predictive models with adam optimizer for protein-protein dataset is given in Table 7.8.

Table 7.8 Results of Sequential DNN Based Predictive Model with Adam Optimizer for Protein-Protein Dataset

Dropout	**Epochs**	Explained variance score	R2 score	Mean squared error	Root mean squared error	Mean absolute error	Median absolute error
0.2	50	0.68	0.68	0.45	0.67	0.49	0.34
	100	0.70	0.70	0.32	0.57	0.35	0.23

	200	0.75	0.75	0.30	0.59	0.39	0.27
	400	0.80	0.80	0.20	0.44	0.28	0.19
	500	**0.94**	**0.94**	**0.13**	**0.36**	**0.15**	**0.14**
0.3	50	0.70	0.70	0.32	0.57	0.35	0.23
	100	0.75	0.75	0.30	0.59	0.39	0.27
	200	0.82	0.82	0.20	0.44	0.22	0.15
	400	0.90	0.90	0.12	0.34	0.15	0.6
	500	**0.95**	**0.95**	**0.1**	**0.32**	**0.12**	**0.4**
0.4	50	0.75	0.75	0.30	0.59	0.39	0.27
	100	0.82	0.82	0.20	0.44	0.22	0.15
	200	0.86	0.86	0.2	0.44	0.22	0.15
	400	0.89	0.89	0.2	0.44	0.24	0.10
	500	**0.90**	**0.90**	**0.15**	**0.34**	**0.15**	**0.6**
0.5	50	0.75	0.75	0.30	0.9	0.39	0.27
	100	0.82	0.82	0.20	0.44	0.22	0.15
	200	0.90	0.90	0.12	0.34	0.15	0.6
	400	0.80	0.80	0.20	0.44	0.28	0.19
	500	**0.92**	**0.92**	**0.14**	**0.32**	**0.12**	**0.4**

The results of sequential DNN predictive models with adam optimizer for protein-protein dataset illustrates that the prediction rate and error rate for the the dropout 0.3 and epoch 500 is 0.95 and 0.1 respectively. The results for root mean squared error, mean absolute error, median absolute error and R2 score obtained are 0.32, 0.12, 0.4 and 0.95 respectively. The prediction rate and error rate obtained for dropout 0.2 is 0.94 and 0.13 respectively. The results of root mean squared error, mean absolute error, median absolute error and R2 score acquired for dropout 0.2 are 0.36, 0.15, 0.14 and 0.94 respectively. The prediction rate and error rate obtained for the dropout 0.5 is 0.92 and 0.14 respectively. The results for root mean squared error, mean absolute error, median absolute error and R2 score obtained are 0.32, 0.12, 0.4 and 0.92 respectively. The least prediction rate and error rate obtained for dropout 0.5 is 0.92 and 0.14 respectively. The resulst for root mean squared error, mean absolute error, median absolute error and R2 score for the dropout 0.5 and epoch 500 is 0.32, 0.12, 0.4 and 0.95 respectively. The results of sequential DNN predictive models with RMSprop optimizer for protein-protein dataset are given in Table 7.9.

Table 7.9 Results of Sequential DNN Based Predictive Model with RMSprop Optimizer for Protein-Protein Dataset

Dropout	**Epochs**	Explained variance score	R2 score	Mean squared error	Root mean squared error	Mean absolute error	Median absolute error
0.2	50	0.68	0.68	0.45	0.67	0.49	0.34
	100	0.70	0.70	0.32	0.57	0.35	0.23
	200	0.72	0.72	0.32	0.54	0.34	0.24
	400	0.76	0.76	0.30	0.52	0.30	0.22
	500	**0.85**	**0.85**	**0.20**	**0.44**	**0.25**	**0.15**
0.3	50	0.70	0.70	0.32	0.57	0.35	0.23
	100	0.72	0.72	0.30	0.59	0.39	0.27
	200	0.80	0.80	0.20	0.44	0.22	0.15
	400	0.83	0.83	0.20	0.44	0.25	0.15
	500	**0.90**	**0.90**	**0.14**	**0.32**	**0.12**	**0.4**
0.4	50	0.75	0.75	0.30	0.59	0.39	0.27
	100	0.82	0.82	0.20	0.44	0.22	0.15
	200	0.84	0.84	0.2	0.44	0.22	0.15
	400	0.85	0.85	0.2	0.44	0.24	0.10
	500	**0.89**	**0.89**	**0.15**	**0.34**	**0.15**	**0.6**
0.5	50	0.75	0.75	0.30	0.59	0.39	0.27
	100	0.82	0.82	0.20	0.44	0.22	0.15
	200	0.82	0.82	0.20	0.44	0.22	0.15
	400	0.85	0.85	0.20	0.44	0.25	0.15
	500	**0.89**	**0.89**	**0.18**	**0.35**	**0.16**	**0.8**

Affinity predictive model with sequential DNN using RMSprop optimizer acquired the prediction rate and error rate for the dropout 0.3 is 0.90 and 0.14. The results for root mean squared error, mean absolute error, median absolute error and R2 score for the dropout 0.3 is 0.32, 0.12, 0.4 and 0.90 respectively. The next highest prediction rate obtained is 0.89 for dropout of 0.4 and 0.5. The results for mean squared error, root mean squared error, mean absolute error, median absolute error and R2 score for for the dropout 0.4, 0.5 and epoch 500 is 0.18, 0.35, 0.16, 0.8 and 0.89 respectively. The results of sequential DNN predictive models with Nadam optimizer for protein-protein dataset are given in Table 7.10.

Table 7.10 Results of Sequential DNN Based Predictive Model with Nadam Optimizer for Protein-Protein Dataset

Dropout	**Epochs**	Explained variance score	R2 score	Mean squared error	Root mean squared error	Mean absolute error	Median absolute error
0.2	50	0.68	0.68	0.45	0.67	0.49	0.34
	100	0.70	0.70	0.32	0.57	0.35	0.23
	200	0.72	0.72	0.32	0.54	0.34	0.24
	400	0.76	0.76	0.30	0.52	0.30	0.22
	500	**0.78**	**0.78**	**0.24**	**0.48**	**0.27**	**0.17**
0.3	50	0.70	0.70	0.32	0.57	0.35	0.23
	100	0.72	0.72	0.30	0.59	0.39	0.27
	200	0.76	0.76	0.30	0.52	0.30	0.22
	400	0.80	0.80	0.20	0.44	0.25	0.15
	500	**0.82**	**0.82**	**0.20**	**0.44**	**0.22**	**0.15**
0.4	50	0.75	0.75	0.30	0.59	0.39	0.27
	100	0.82	0.82	0.20	0.44	0.22	0.15
	200	0.84	0.84	0.2	0.44	0.22	0.15
	400	0.85	0.85	0.2	0.44	0.24	0.10
	500	**0.85**	**0.85**	**0.2**	**0.44**	**0.24**	**0.10**
0.5	50	0.75	0.75	0.30	0.59	0.39	0.27
	100	0.80	0.80	0.20	0.44	0.25	0.15
	200	0.82	0.82	0.20	0.44	0.22	0.15
	400	0.85	0.85	0.20	0.44	0.25	0.15
	500	**0.85**	**0.85**	**0.20**	**0.44**	**0.25**	**0.15**

The results of sequential DNN predictive model with Nadam optimizer based on protein-protein dataset obtains the prediction rate and error rate for the dropout 0.4, 0.5 is 0.85 and 0.20 respectively. The results for root mean squared error, mean absolute error, median absolute error and R2 score acquired are 0.44, 0.25, 0.15 and 0.85 respectively. The next highest prediction rate and error rate obtained for the dropout 0.3 and epoch 500 is 0.82 and 0.2 respectively. The results for root mean squared error, mean absolute error, median absolute error and R2 score obtained are 0.44, 0.22, 0.15 and 0.82 respectively. The least prediction rate and error rate is for the dropout 0.2 is 0.78 and 0.24 respectively. The results for root mean squared error, mean absolute error, median absolute error and R2 score for the dropout is 0.48, 0.27, 0.17 and 0.78 respectively. The results obtained for different dropouts and epochs are illustrated from Fig. 7.8 to Fig. 7.10.

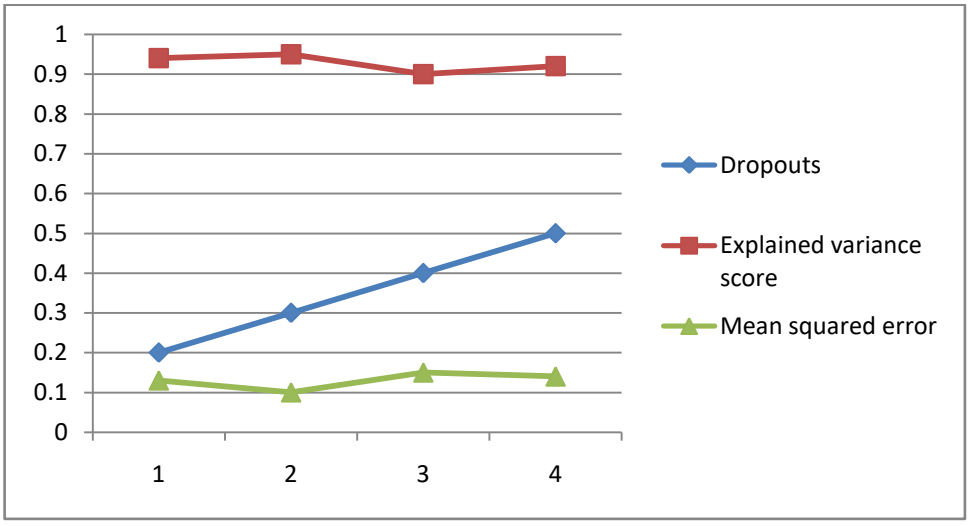

Fig. 7.8 Results of Sequential DNN Based Predictive Model with Adam Optimizer at Various Dropouts for Protein-Protein Dataset

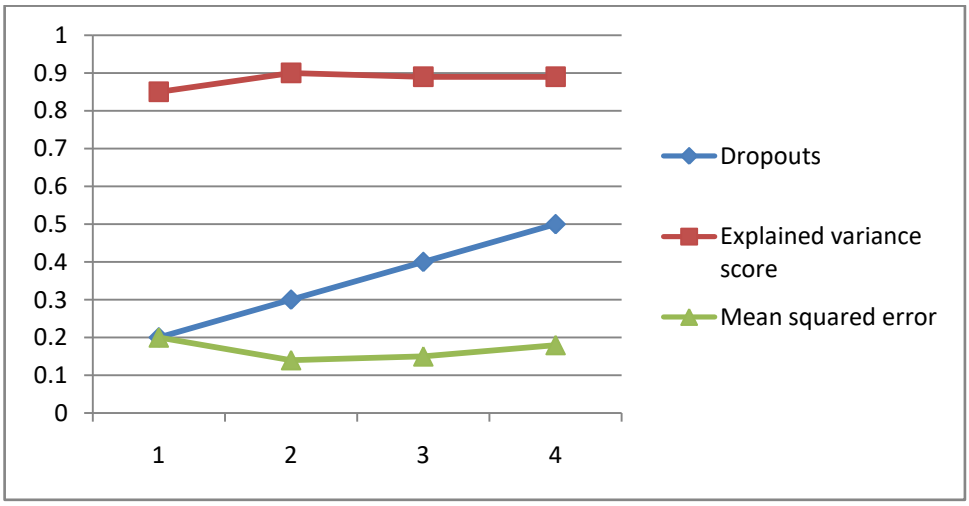

Fig. 7.9 Results of Sequential DNN Based Predictive Model with RMSprop Optimizer at Various Dropouts for Protein-Protein Dataset

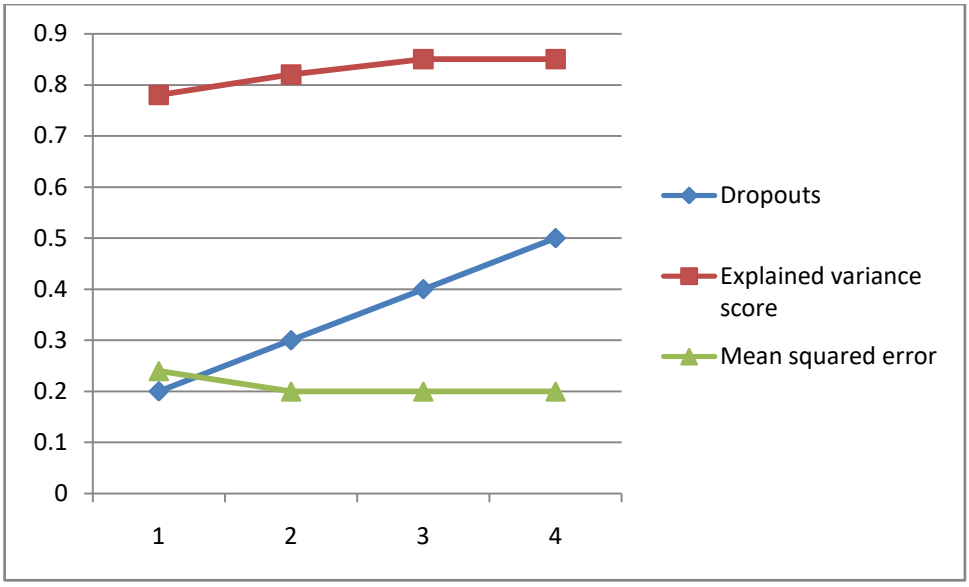

Fig. 7.10 Results of Sequential DNN Based Predictive Model with Nadam Optimizer at Various Dropouts for Protein-Protein Dataset

The charts from Fig.7.8 to Fig. 7.10 illustrates that the prediction rate for adam optimizer obtained for dropout 0.3 is 0.95 where as the prediction rate of RMSprop and Nadam optimizers for the dropout 0.3 is 0.90 and 0.82 respectively. The mean squared error obtained for the adam optimizer is 0.1 which is low than other two optimizers whereas the mean squared error for the optimizers RMSprop and Nadam is 0.14 and 0.20 which is superior to adam optimizer. The Nadam optimizer attained the highest prediction rate at the dropout 0.4 and 0.5 as 0.85. The results of three datasets show that the adam optimizer based model performs better than other two optimizers. The adam optimizer based model obtained the prediction rate 0.95 for PLD dataset, 0.85 for PMLD dataset, 0.95 for PPD dataset. These prediction rates are obtained for the dropout 0.3 and epoch 500 whereas the error rate obtained for these datasets are 0.1, 0.20 and 0.1. Among these datasets PLD and PPD datasets has the superior prediction rate and the inferior mean squared error.

In the next section the development of binding affinity predictive model based on functional DNN is discussed.

Findings

The benefit of deep architecture based on sequential deep neural network to predict binding affinity prediction for SCA is confirmed compared to traditional regression. It is evident that the performance of sequential DNN is relatively higher than the regression algorithms for all the three datasets developed based on protein-ligand docking, protein-mutated-ligand docking and protein-protein interaction. The high prediction rate of sequential

DNN model proves that the network is able to learn efficiently with two separate dense layers by learning the user defined features through representation learning which aids in accurate prediction of binding affinity. The sequential DNN predictive model with appropriate settings of hyper parameters improves the ability of the model and strength of prediction tasks. The model with protein-protein interaction properties achieves the highest prediction rate and minimum error rate than other two docking approaches. It is ascertain that the architecture of sequential DNN and the representation learning of user defined features contribute to the overall performance in predicting binding affinity.

7.2 BINDING AFFINITY PREDICTION MODELS USING FUNCTIONAL DEEP NEURAL NETWORK

The sequential DNN models implemented on the three different datasets with three optimizers presented in the previous section is good at learning signals and hidden features from the training data but it does not share layers and do not possess any connection with past inputs. Since there is no connection between the layers, sequential DNN has no connection with the past input and layers. So, here the variant of sequential DNN architecture called functional DNN is used which connects with past input by sharing layers and predicts the prediction rate better. In functional DNN layers are connected with each layer where the input of one layer is passed as output to the next layer. It allows to define multiple input or output models and models are defined by creating instances of layers and connecting them directly to each other in pairs, then defining a model that specifies the layers to act as the input and output to the model. Binding affinity predictive models with functional DNN is good at building complex models like multi-input/multi-output models and models with shared layers. In this work, models with shared layers are used for implementing the affinity predictive models. This section illustrates the development of functional DNN based models to predict binding affinity for SCA with three different datasets such as PLD, PMLD and PPD datasets.

Methodology

The functional DNN is utilized to construct binding affinity predictive models through self learning the features using three datasets. The predictive models are constructed with three datasets such as PLD, PMLD, PPD with three optimizers such as adam, RMSprop and Nadam. The hyper parameters used in the previous case are defined here also to reinforce the functional DNN architecture. The proposed framework of functional DNN based binding affinity prediction model is shown in Fig. 7.11.

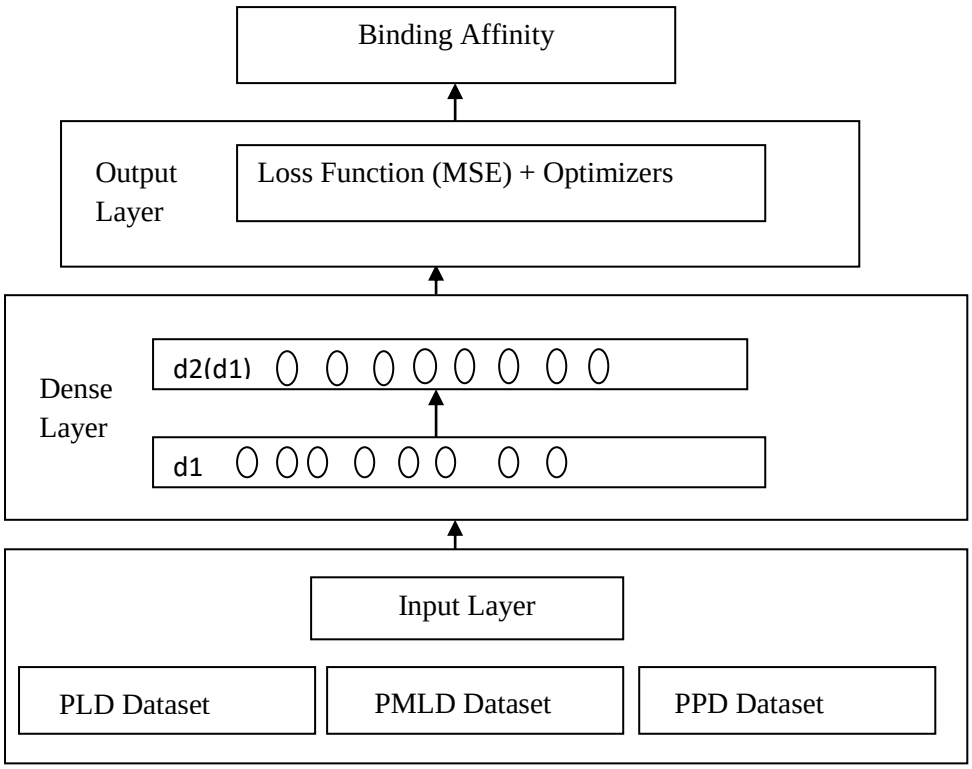

Fig. 7.11 Architecture of Functional DNN Based Prediction Model

The input feature vectors created as described in chapter 4, 5 and 6 corresponding to three datasets such as PLD, PMLD and PPD are directly fed as input to the functional DNN architecture. The proposed functional DNN consists of one input layer, 2 hidden layers with 8 memory units and one output layer. The input layer is defined with input shape and layers are connected. Here in the functional DNN weights are assigned randomly to the neurons in each layer and adjusted according to the error.

In this technique of binding affinity predictive models, it has one neuron in the output layer whereas 30 neurons in first hidden layer and 20 neurons in second hidden layer. The output layer is defined with relu which transforms the summed weighted input from the node into the output. The layers are connected by tensors to carry forward the computed tensor as input to the next layer to improve the prediction rate. Various hyper parameters such as epochs, dropouts, learning rate and optimizers are utilized.

At last, 10-fold cross validation is used to validate the performance of the models using various metrics such as explained variance score, mean squared error, root mean squared error, mean absolute error, median absolute error and R2 score.

Experiments and Results

Experiments have been carried out by implementing functional DNN using keras in python with three datasets as inputs to build the prediction models. The network is trained with varying epochs of 50, 100, 200, 400, 500 and dropout values as 0.2, 0.3, 0.4, 0.5 are considered. The learning rate is fixed as 0.01 and the three optimizers like adam, RMSprop, Nadam with 2 hidden layers. Three optimizers are used to check the variations in prediction rate. With these parameter settings, three datasets are trained and the predictive models are built. The recognition rate of these predictive models is evaluated using 10-fold cross validation technique. The performance of the models are evaluated on metrics like explained variance score, mean squared error, root mean squared error, mean absolute error, median absolute error and R2 score.

Results for Protein-Ligand Dataset

The results of functional DNN based binding affinity predictive models built using protein-ligand dataset with three optimizers such as adam, RMSprop, Nadam and for various dropouts are presented below. The results of functional DNN based binding affinity predictive model with adam optimizer is presented in Table 7.11.

Table 7.11 Results of Functional DNN Based Predictive Model with Adam Optimizer for Protein-Ligand Dataset

Dropout	**Epochs**	Explained variance score	R2 score	Mean squared error	Root mean squared error	Mean absolute error	Median absolute error
0.2	50	0.68	0.68	0.45	0.67	0.49	0.34
	100	0.70	0.70	0.32	0.57	0.35	0.23
	200	0.75	0.75	0.30	0.59	0.39	0.27
	400	0.80	0.80	0.20	0.44	0.28	0.19
	500	**0.94**	**0.94**	**0.13**	**0.36**	**0.15**	**0.14**
0.3	50	0.70	0.70	0.32	0.57	0.35	0.23
	100	0.75	0.75	0.30	0.59	0.39	0.27
	200	0.82	0.82	0.20	0.44	0.22	0.15
	400	0.90	0.90	0.12	0.34	0.15	0.6
	500	**0.97**	**0.97**	**0.1**	**0.32**	**0.12**	**0.4**
0.4	50	0.75	0.75	0.30	0.59	0.39	0.27
	100	0.82	0.82	0.20	0.44	0.22	0.15
	200	0.86	0.86	0.2	0.44	0.22	0.15

	400	0.89	0.89	0.2	0.44	0.24	0.10
	500	**0.90**	**0.90**	**0.15**	**0.34**	**0.15**	**0.6**
0.5	50	0.75	0.75	0.30	0.59	0.39	0.27
	100	0.82	0.82	0.20	0.44	0.22	0.15
	200	0.90	0.90	0.12	0.34	0.15	0.6
	400	0.80	0.80	0.20	0.44	0.28	0.19
	500	**0.92**	**0.92**	**0.14**	**0.32**	**0.12**	**0.4**

The results of functional DNN with adam optimizer for PLD dataset obtains the prediction rate and error rate for the dropout 0.3 is 0.97 and 0.1 respectively. The results for root mean squared error, mean absolute error, median absolute error and R2 score acquired are 0.32, 0.12, 0.4 and 0.97 respectively. The next highest prediction rate and error rate obtained for the dropout 0.2 is 0.94 and 0.13 respectively. The results for root mean squared error, mean absolute error, median absolute error and R2 score as 0.36, 0.15, 0.14 and 0.94 respectively. The least prediction rate and the error rate obtained for the dropout 0.4 is 0.90 and 0.15 respectively. The results for root mean squared error, mean absolute error, median absolute error and R2 score are 0.34, 0.15, 0.6 and 0.90 respectively. The results of functional DNN predictive models with RMSprop optimizer for protein-ligand dataset are given in Table 7.12.

Table 7.12 Results of Functional DNN Based Predictive Model with RMSprop Optimizer for Protein-Ligand Dataset

Dropout	**Epochs**	Explained variance score	R2 score	Mean squared error	Root mean squared error	Mean absolute error	Median absolute error
0.2	50	0.70	0.70	0.32	0.57	0.35	0.23
	100	0.70	0.70	0.32	0.57	0.35	0.23
	200	0.72	0.72	0.32	0.54	0.34	0.24
	400	0.76	0.76	0.30	0.52	0.30	0.22
	500	**0.80**	**0.80**	**0.20**	**0.44**	**0.25**	**0.15**
0.3	50	0.70	0.70	0.32	0.57	0.35	0.23
	100	0.80	0.80	0.20	0.44	0.25	0.15
	200	0.80	0.80	0.20	0.44	0.25	0.15
	400	0.82	0.82	0.20	0.44	0.22	0.15
	500	**0.90**	**0.90**	**0.15**	**0.34**	**0.15**	**0.6**
0.4	50	0.75	0.75	0.30	0.59	0.39	0.27

	100	0.82	0.82	0.20	0.44	0.22	0.15
	200	0.84	0.84	0.2	0.44	0.22	0.15
	400	0.85	0.85	0.2	0.44	0.24	0.10
	500	**0.85**	**0.85**	**0.2**	**0.44**	**0.24**	**0.10**
0.5	50	0.75	0.75	0.30	0.59	0.39	0.27
	100	0.80	0.80	0.20	0.44	0.25	0.15
	200	0.82	0.82	0.20	0.44	0.22	0.15
	400	0.85	0.85	0.20	0.44	0.25	0.15
	500	**0.85**	**0.85**	**0.20**	**0.44**	**0.25**	**0.15**

The predictive model with RMSprop optimizer for PLD dataset attains the prediction rate and error rate for the dropout 0.3 is 0.90 and 0.15 respectively. The results for root mean squared error, mean absolute error, median absolute error and R2 score acquired are 0.34, 0.15, 0.6 and 0.90 respectively. The model yields the prediction rate and error rate for the dropout 0.4, 0.5 is 0.85 and 0.20 respectively. The results for root mean squared error, mean absolute error and median absolute error produced are 0.44, 0.25 and 0.15 respectively. The least prediction rate and error rate for the dropout 0.2 obtained is 0.80 and 0.20 respectively. The results for root mean squared error, mean absolute error and median absolute error as 0.44, 0.25, 0.15 respectively. The results of functional DNN predictive models with Nadam optimizer for protein-ligand dataset are given in Table 7.13.

Table 7.13 Results of Functional DNN Based Predictive Model with Nadam Optimizer for Protein-Ligand Dataset

Dropout	**Epochs**	Explained variance score	R2 score	Mean squared error	Root mean squared error	Mean absolute error	Median absolute error
0.2	50	0.70	0.70	0.32	0.57	0.35	0.23
	100	0.70	0.70	0.32	0.57	0.35	0.23
	200	0.72	0.72	0.32	0.54	0.34	0.24
	400	0.76	0.76	0.30	0.52	0.30	0.22
	500	**0.80**	**0.80**	**0.20**	**0.44**	**0.25**	**0.15**
0.3	50	0.70	0.70	0.32	0.57	0.35	0.23
	100	0.80	0.80	0.20	0.44	0.25	0.15
	200	0.80	0.80	0.20	0.44	0.25	0.15
	400	0.82	0.82	0.20	0.44	0.22	0.15
	500	**0.90**	**0.90**	**0.15**	**0.34**	**0.15**	**0.6**
0.4	50	0.75	0.75	0.30	0.59	0.39	0.27

	100	0.82	0.82	0.20	0.44	0.22	0.15
	200	0.84	0.84	0.2	0.44	0.22	0.15
	400	0.85	0.85	0.2	0.44	0.24	0.10
	500	**0.85**	**0.85**	**0.2**	**0.44**	**0.24**	**0.10**
0.5	50	0.75	0.75	0.30	0.59	0.39	0.27
	100	0.80	0.80	0.20	0.44	0.25	0.15
	200	0.82	0.82	0.20	0.44	0.22	0.15
	400	0.85	0.85	0.20	0.44	0.25	0.15
	500	**0.85**	**0.85**	**0.20**	**0.44**	**0.25**	**0.15**

The predictive model with Nadam optimizer for PLD dataset attains the same results as of RMSprop optimizer with the highest prediction rate and error rate for the dropout 0.3 is 0.90 and 0.15 respectively. The results obtained for different dropouts and epochs are illustrated from Fig. 7.12 to Fig. 7.14.

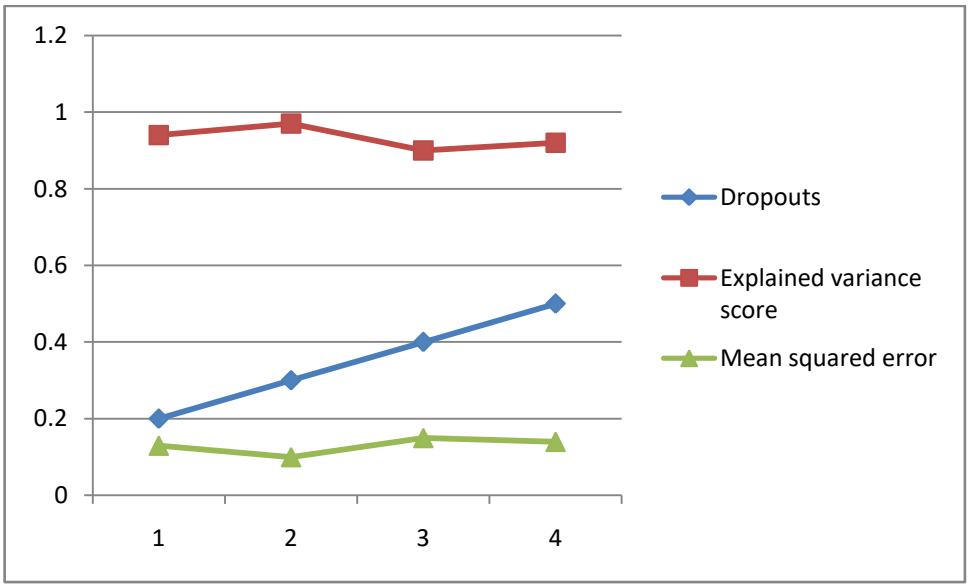

Fig. 7.12 Results of Functional DNN Based Predictive Model with Adam Optimizer at Various Dropouts for Protein-Ligand Dataset

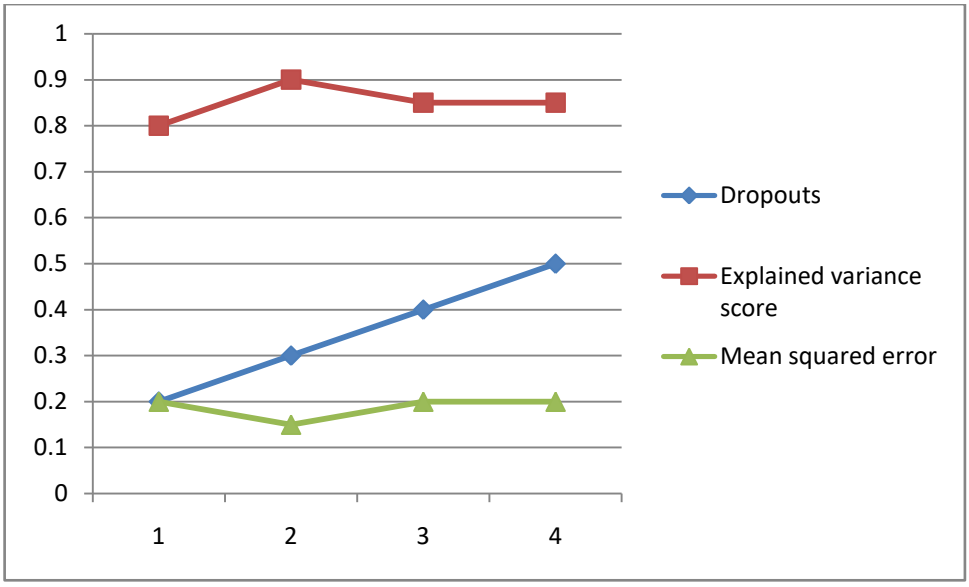

Fig. 7.13 Results of Functional DNN Based Predictive Model with RMSprop Optimizer at Various Dropouts for Protein-Ligand Dataset

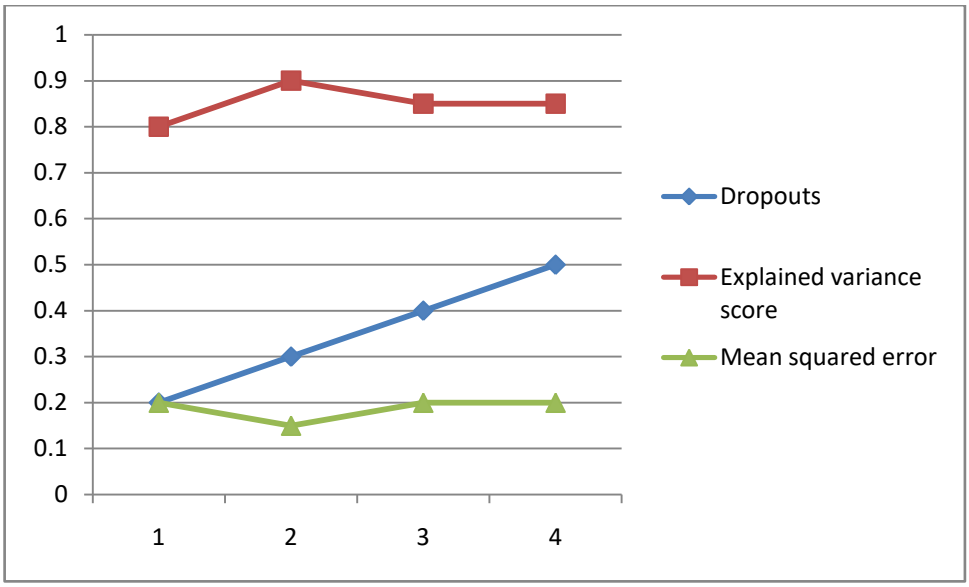

Fig. 7.14 Results of Functional DNN Based Predictive Model with Nadam Optimizer at Various Dropouts for Protein-Ligand Dataset

The charts from Fig. 7.12 to Fig. 7.14 illustrates that the predictive models based on PLD dataset showed that adam optimizer performs better when compared with RMSprop and Nadam optimizer. The adam optimizer showed the prediction rate of 0.97 and the minimized error rate of 0.1 which is inferior to other two optimizers. The RMSprop and Nadam optimizer produces the same result.

Results for Protein-Mutated-Ligand Dataset

The results of functional DNN based binding affinity predictive models built using protein-mutated-ligand dataset with three optimizers such as adam, RMSprop, Nadam for various dropouts are presented below. The results of functional predictive models with adam optimizer for protein-mutated-ligand dataset is tabulated in Table 7.14.

Table 7.14 Results of Functional DNN Based Predictive Model with adam Optimizer for Protein-Mutated-Ligand Dataset

Dropout	**Epochs**	Explained variance score	R2 score	Mean squared error	Root mean squared error	Mean absolute error	Median absolute error
0.2	50	0.70	0.70	0.32	0.57	0.35	0.23
	100	0.70	0.70	0.32	0.57	0.35	0.23
	200	0.76	0.76	0.30	0.52	0.30	0.22
	400	0.76	0.76	0.30	0.52	0.30	0.22
	500	**0.82**	**0.82**	**0.20**	**0.44**	**0.22**	**0.15**
0.3	50	0.70	0.70	0.32	0.57	0.35	0.23
	100	0.80	0.80	0.20	0.44	0.25	0.15
	200	0.80	0.80	0.20	0.44	0.25	0.15
	400	0.85	0.85	0.2	0.44	0.24	0.10
	500	**0.90**	**0.90**	**0.15**	**0.34**	**0.15**	**0.6**
0.4	50	0.75	0.75	0.30	0.59	0.39	0.27
	100	0.82	0.82	0.20	0.44	0.22	0.15
	200	0.84	0.84	0.2	0.44	0.22	0.15
	400	0.85	0.85	0.2	0.44	0.24	0.10
	500	**0.85**	**0.85**	**0.2**	**0.44**	**0.24**	**0.10**
0.5	50	0.75	0.75	0.30	0.59	0.39	0.27
	100	0.80	0.80	0.20	0.44	0.25	0.15
	200	0.82	0.82	0.20	0.44	0.22	0.15
	400	0.82	0.82	0.20	0.44	0.22	0.15
	500	**0.85**	**0.85**	**0.20**	**0.44**	**0.25**	**0.15**

The functional DNN based predictive model with adam optimizer for PMLD dataset reveals that the prediction rate and error rate obtained for the dropout of 0.3 and epoch 500 is 0.90 and 0.15 respectively. The results for for root mean squared error, mean absolute error and median absolute error as 0.34, 0.15, 0.6 respectively. The least prediction rate acquired for the dropout 0.4, 0.5 is 0.85 and 0.20 respectively. The results for root mean squared error,

mean absolute error and median absolute error acquired are 0.44, 0.25, 0.15 respectively The results of functional predictive models with RMSprop optimizer for protein-mutated-ligand dataset is tabulated in Table 7.15.

Table 7.15 Results of Functional DNN Based Predictive Model with RMSprop Optimizer for Protein-Mutated-Ligand Dataset

Dropout	**Epochs**	Explained variance score	R2 score	Mean squared error	Root mean squared error	Mean absolute error	Median absolute error
0.2	50	0.70	0.70	0.32	0.57	0.35	0.23
	100	0.70	0.70	0.32	0.57	0.35	0.23
	200	0.76	0.76	0.30	0.52	0.30	0.22
	400	0.76	0.76	0.30	0.52	0.30	0.22
	500	**0.82**	**0.82**	**0.20**	**0.44**	**0.22**	**0.15**
0.3	50	0.70	0.70	0.32	0.57	0.35	0.23
	100	0.80	0.80	0.20	0.44	0.25	0.15
	200	0.82	0.82	0.20	0.44	0.22	0.15
	400	0.84	0.84	0.2	0.44	0.22	0.15
	500	**0.85**	**0.85**	**0.2**	**0.44**	**0.24**	**0.10**
0.4	50	0.75	0.75	0.30	0.59	0.39	0.27
	100	0.80	0.80	0.20	0.44	0.25	0.15
	200	0.80	0.80	0.20	0.44	0.25	0.15
	400	0.80	0.80	0.20	0.44	0.25	0.15
	500	**0.80**	**0.80**	**0.20**	**0.44**	**0.25**	**0.15**
0.5	50	0.75	0.75	0.30	0.59	0.39	0.27
	100	0.80	0.80	0.20	0.44	0.25	0.15
	200	0.82	0.82	0.20	0.44	0.22	0.15
	400	0.82	0.82	0.20	0.44	0.22	0.15
	500	**0.85**	**0.85**	**0.20**	**0.44**	**0.25**	**0.15**

The predictive model with RMSprop optimizer for PMLD dataset demonstrates that the prediction rate and error rate obtained for the dropout of 0.3, 0.5 is 0.85 and 0.20 respectively. The results for root mean squared error, mean absolute error and median absolute error acquired are 0.44, 0.25, 0.15 respectively. The least prediction rate produced for the dropout 0.4 is 0.80 and the error rate is 0.20. The results for root mean squared error, mean absolute error and median absolute error attained are 0.44, 0.25, 0.15 respectively. The

results of functional DNN models with Nadam optimizer for protein-mutated-ligand dataset are summarized in Table 7.16.

Table 7.16 Results of Functional DNN Based Predictive Model with Nadam Optimizer for Protein-Mutated-Ligand Dataset

Dropout	**Epochs**	Explained variance score	R2 score	Mean squared error	Root mean squared error	Mean absolute error	Median absolute error
0.2	50	0.70	0.70	0.32	0.57	0.35	0.23
	100	0.70	0.70	0.32	0.57	0.35	0.23
	200	0.72	0.72	0.30	0.50	0.29	0.20
	400	0.72	0.72	0.30	0.50	0.29	0.20
	500	**0.74**	**0.74**	**0.29**	**0.48**	**0.20**	**0.12**
0.3	50	0.70	0.70	0.32	0.57	0.35	0.23
	100	0.72	0.72	0.30	0.50	0.29	0.20
	200	0.72	0.72	0.30	0.50	0.29	0.20
	400	0.72	0.72	0.30	0.50	0.29	0.20
	500	**0.75**	**0.75**	**0.30**	**0.59**	**0.39**	**0.27**
0.4	50	0.68	0.68	0.45	0.67	0.49	0.34
	100	0.68	0.68	0.45	0.67	0.49	0.34
	200	0.70	0.70	0.32	0.57	0.35	0.23
	400	0.70	0.70	0.32	0.57	0.35	0.23
	500	**0.70**	**0.70**	**0.32**	**0.57**	**0.35**	**0.23**
0.5	50	0.75	0.75	0.30	0.59	0.39	0.27
	100	0.75	0.75	0.30	0.59	0.39	0.27
	200	0.75	0.75	0.30	0.59	0.39	0.27
	400	0.75	0.75	0.30	0.59	0.39	0.27
	500	**0.75**	**0.75**	**0.30**	**0.59**	**0.39**	**0.27**

Table 7.16 shows that the prediction rate and error rate with Nadam optimizer produced for the dropout 0.3, 0.5 is 0.75 and 0.30 respectively. The results for root mean squared error, mean absolute error and median absolute error as 0.59, 0.39, 0.27 respectively. The least prediction rate and the error rate for the dropout 0.4 is 0.70 and 0.32 respectively. The results for root mean squared error, mean absolute error and median absolute error as 0.57, 0.35 and 0.23 respectively. The results obtained for different dropouts and epochs are illustrated from Fig. 7.15 to Fig. 7.17.

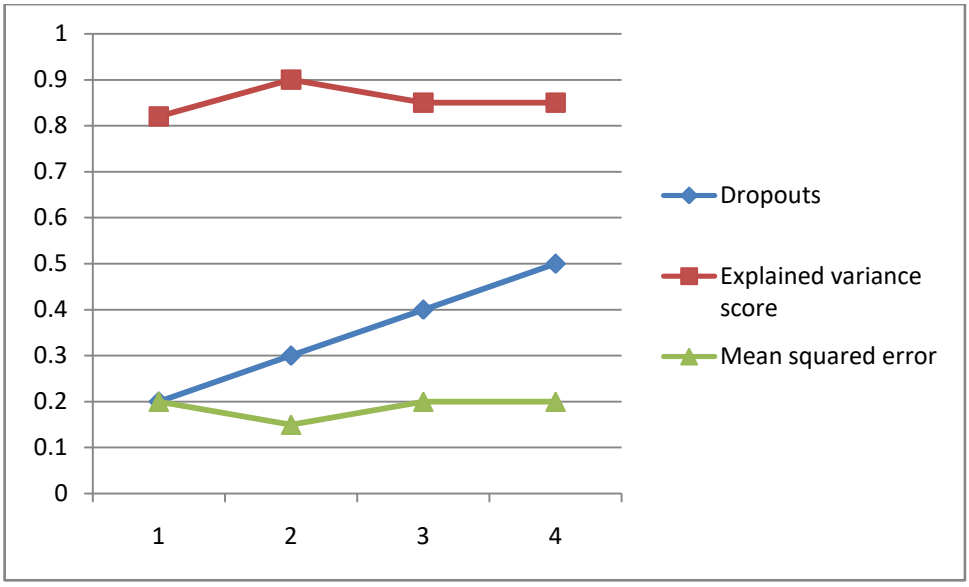

Fig. 7.15 Results of Functional DNN Based Predictive Model with Adam Optimizer at Various Dropouts for Protein-Mutated-Ligand Dataset

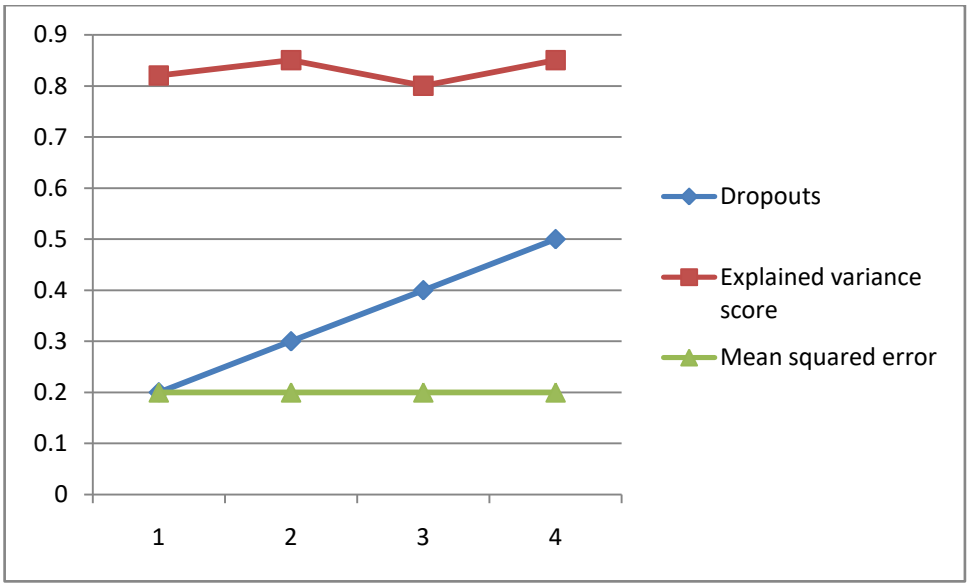

Fig. 7.16 Results of Functional DNN Based Predictive Model with RMSprop Optimizer at Various Dropouts for Protein-Mutated-Ligand Dataset

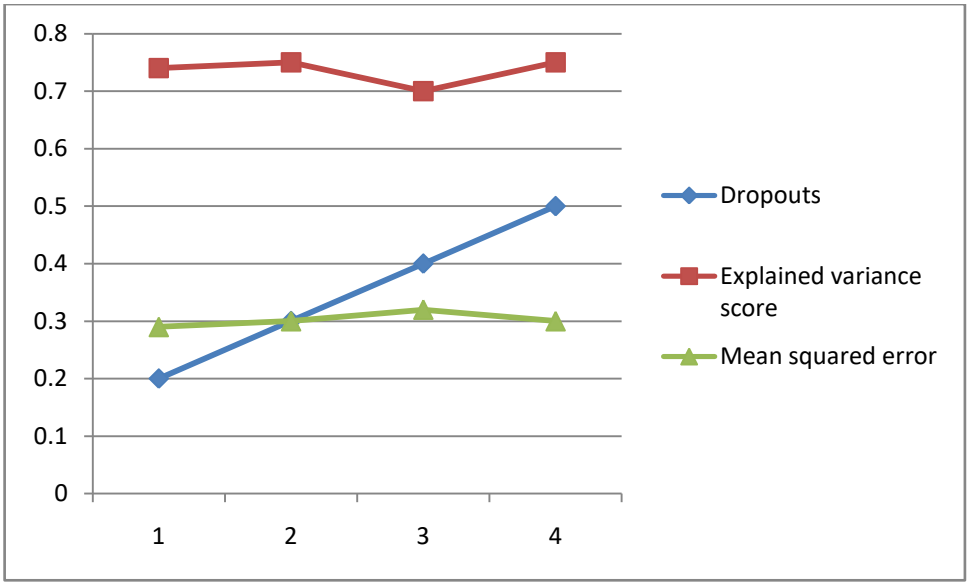

Fig. 7.17 Results of Functional DNN Based Predictive Model with Nadam Optimizer at Various Dropouts for Protein-Mutated-Ligand Dataset

The charts from Fig. 7.15 to Fig. 7.17 of varying dropouts for PMLD dataset shows that the predictive model with adam optimizer achieved prediction rate of 0.90 and the error rate of 0.15 where the prediction rate is superior to other two optimizers RMSprop, Nadam optimizer and the error rate is inferior than two optimizers. This shows that adam optimizer achieves better prediction rate and it is suitable for the predicting binding affinity.

Results for Protein-protein Dataset

The results of functional DNN models based on protein-protein dataset with three optimizers such as adam, RMSprop, Nadam and for various dropouts are presented below. The results of functional DNN based predictive models with adam optimizer for protein-protein dataset is given in Table 7.17.

Table 7.17 Results of Functional DNN Based Predictive Model with Adam Optimizer for Protein-Protein Dataset

Dropout	**Epochs**	Explained variance score	R2 score	Mean squared error	Root mean squared error	Mean absolute error	Median absolute error
0.2	50	0.76	0.76	0.30	0.52	0.30	0.22
	100	0.76	0.76	0.30	0.52	0.30	0.22
	200	0.82	0.82	0.20	0.44	0.22	0.15
	400	0.82	0.82	0.20	0.44	0.22	0.15

	500	**0.84**	**0.84**	**0.30**	**0.48**	**0.20**	**0.12**
0.3	50	0.84	0.75	0.30	0.48	0.20	0.12
	100	0.84	0.75	0.30	0.48	0.20	0.12
	200	0.85	0.85	0.2	0.44	0.24	0.10
	400	0.90	0.90	0.15	0.34	0.15	0.6
	500	**0.97**	**0.97**	**0.1**	**0.32**	**0.12**	**0.4**
0.4	50	0.85	0.85	0.2	0.44	0.24	0.10
	100	0.85	0.85	0.2	0.44	0.24	0.10
	200	0.85	0.85	0.2	0.44	0.24	0.10
	400	0.90	0.90	0.15	0.34	0.15	0.6
	500	**0.90**	**0.90**	**0.15**	**0.34**	**0.15**	**0.6**
0.5	50	0.85	0.85	0.2	0.44	0.24	0.10
	100	0.85	0.85	0.2	0.44	0.24	0.10
	200	0.85	0.85	0.2	0.44	0.24	0.10
	400	0.90	0.90	0.15	0.34	0.15	0.6
	500	**0.90**	**0.90**	**0.15**	**0.34**	**0.15**	**0.6**

Table 7.17 shows that the prediction rate and error rate with adam optimizer for the dropout 0.3, 0.5 is 0.97 and 0.1 respectively. The results for root mean squared error, mean absolute error and median absolute error as 0.32, 0.12 and 0.4 respectively. The least prediction rate and the error rate for the dropout 0.2 is 0.84 and 0.30 respectively. The results for root mean squared error, mean absolute error and median absolute error obtained are 0.48, 0.20 and 0.12 respectively. The results of functional DNN models with RMSprop optimizer for protein-protein dataset are summarized in Table 7.18.

Table 7.18 Results of Functional DNN Based Predictive Model with RMSprop Optimizer for Protein-Protein Dataset

Dropout	**Epochs**	Explained variance score	R2 score	Mean squared error	Root mean squared error	Mean absolute error	Median absolute error
0.2	50	0.76	0.76	0.30	0.52	0.30	0.22
	100	0.82	0.82	0.20	0.44	0.22	0.15
	200	0.82	0.82	0.20	0.44	0.22	0.15
	400	0.84	0.75	0.30	0.48	0.20	0.12
	500	**0.84**	**0.84**	**0.30**	**0.48**	**0.20**	**0.12**
0.3	50	0.84	0.75	0.30	0.48	0.20	0.12
	100	0.84	0.75	0.30	0.48	0.20	0.12
	200	0.85	0.85	0.2	0.44	0.24	0.10

	400	0.90	0.90	0.15	0.34	0.15	0.6
	500	**0.92**	**0.92**	**0.15**	**0.34**	**0.15**	**0.6**
0.4	50	0.84	0.75	0.30	0.48	0.20	0.12
	100	0.84	0.75	0.30	0.48	0.20	0.12
	200	0.85	0.85	0.2	0.44	0.24	0.10
	400	0.90	0.90	0.15	0.34	0.15	0.6
	500	**0.90**	**0.90**	**0.15**	**0.34**	**0.15**	**0.6**
0.5	50	0.85	0.85	0.2	0.44	0.24	0.10
	100	0.85	0.85	0.2	0.44	0.24	0.10
	200	0.85	0.85	0.2	0.44	0.24	0.10
	400	0.90	0.90	0.15	0.34	0.15	0.6
	500	**0.90**	**0.90**	**0.15**	**0.34**	**0.15**	**0.6**

Table 7.18 shows that the prediction rate and error rate with RMSprop optimizer for the dropout of 0.3 is 0.92 and 0.15 respectively. The results for root mean squared error, mean absolute error and median absolute error attained are 0.34, 0.15 and 0.6 respectively. The least prediction rate and error rate for the dropout 0.2 is 0.84 and 0.30 respectively. The results for root mean squared error, mean absolute error and median absolute error as 0.48, 0.20 and 0.12 respectively. The results of functional DNN models with Nadam optimizer for protein-protein dataset are summarized in Table 7.19.

Table 7.19 Results of Functional DNN Based Predictive Model with Nadam Optimizer for Protein-Protein Dataset

Dropout	**Epochs**	Explained variance score	R2 score	Mean squared error	Root mean squared error	Mean absolute error	Median absolute error
0.2	50	0.76	0.76	0.30	0.52	0.30	0.22
	100	0.82	0.82	0.20	0.44	0.22	0.15
	200	0.82	0.82	0.20	0.44	0.22	0.15
	400	0.84	0.75	0.30	0.48	0.20	0.12
	500	**0.84**	**0.84**	**0.30**	**0.48**	**0.20**	**0.12**
0.3	50	0.84	0.75	0.30	0.48	0.20	0.12
	100	0.84	0.75	0.30	0.48	0.20	0.12
	200	0.85	0.85	0.2	0.44	0.24	0.10
	400	0.90	0.90	0.15	0.34	0.15	0.6
	500	**0.90**	**0.90**	**0.15**	**0.34**	**0.15**	**0.6**
0.4	50	0.84	0.75	0.30	0.48	0.20	0.12
	100	0.84	0.75	0.30	0.48	0.20	0.12

	200	0.85	0.85	0.2	0.44	0.24	0.10
	400	0.85	0.85	0.2	0.44	0.24	0.10
	500	**0.85**	**0.85**	**0.2**	**0.44**	**0.24**	**0.10**
0.5	50	0.84	0.75	0.30	0.48	0.20	0.12
	100	0.84	0.75	0.30	0.48	0.20	0.12
	200	0.85	0.85	0.2	0.44	0.24	0.10
	400	0.85	0.85	0.2	0.44	0.24	0.10
	500	**0.85**	**0.85**	**0.2**	**0.44**	**0.24**	**0.10**

Table 7.19 shows that the prediction rate and error rate with Nadam optimizer for the dropout 0.3 is 0.90 and 0.15 respectively. The results for root mean squared error, mean absolute error and median absolute error as 0.34, 0.15 and 0.6 respectively. The least prediction rate and error rate for the dropout 0.2 is 0.84 and 0.30 respectively. The results for root mean squared error, mean absolute error and median absolute error as 0.48, 0.20 and 0.12 respectively. The results obtained for PPD dataset at different dropouts and epochs are illustrated from Fig. 7.18 to Fig. 7.20.

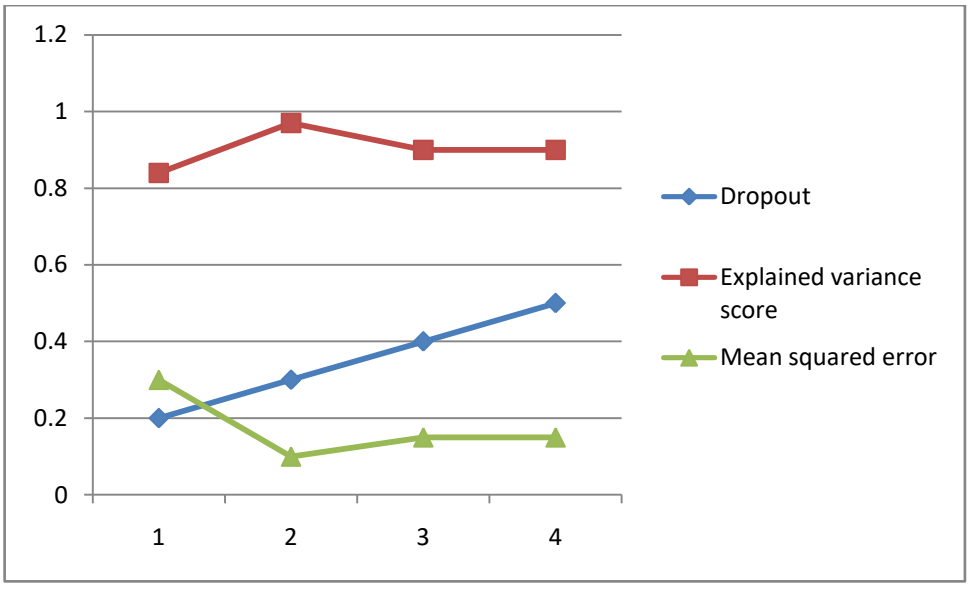

Fig. 7.18 Results of Functional DNN Based Predictive Model with Adam Optimizer at Various Dropouts for Protein-Protein Dataset

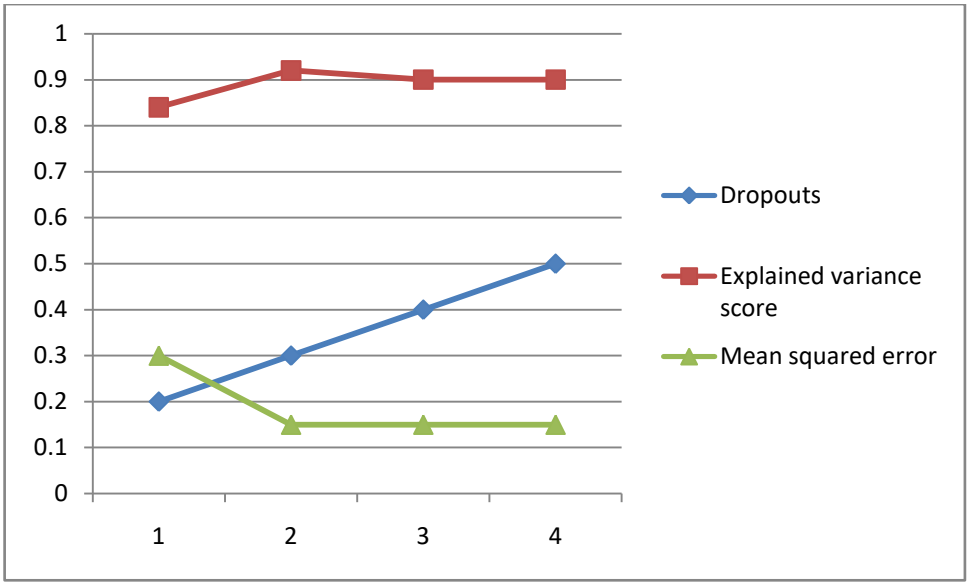

Fig. 7.19 Results of Functional DNN Based Predictive Model with RMSprop Optimizer at Various Dropouts for Protein-Protein Dataset

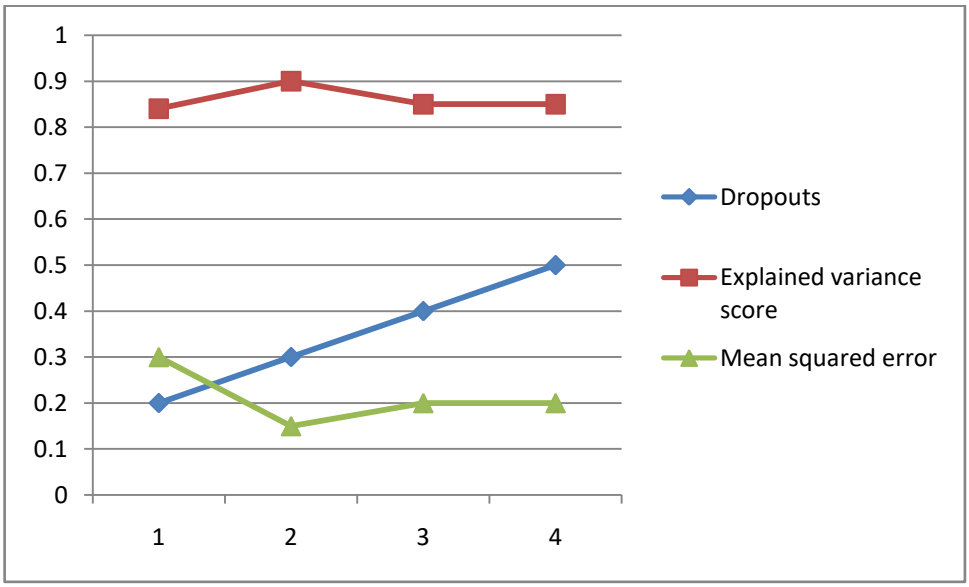

Fig. 7.20 Results of Functional DNN Based Predictive Model with Nadam Optimizer at Various Dropouts for Protein-Protein Dataset

The charts from Fig. 7.18 to Fig. 7.20 demonstrate that the adam optimizer based prediction model at dropout 0.3 achieved the prediction rate 0.97 and error rate 0.1. When compared to other optimizer adam optimizer outperforms in case of functional DNN also. The RMSprop optimizer based prediction model obtained the prediction rate of 0.92 at dropout 0.3 with the error rate of 0.15. The Nadam optimizer based prediction model achieved the prediction rate of 0.90 at dropout 0.3 and the mean squared error is 0.15. Among

these independent models, the model built using PPD dataset with adam optimizer performed well and achieved better prediction rate.

In the next section, the development of binding affinity predictive models built with customized layers in DNN is discussed.

Findings

The benefit of deep architecture based on functional deep neural network to predict binding affinity prediction for SCA is confirmed. It is evident that the performance of funcitonal DNN is relatively higher than the sequential DNN for all the three predictive models based on protein-ligand, protein-mutated-ligand docking and protein-protein interaction. The high prediction rate of functional DNN model proves that the network is able to learn efficiently with two separate dense layers remembering the past data by sharing the layers. The output of previous layer is fed as input to the next layer which shares the information that leads to achieve better prediction rate. The functional DNN predictive models with appropriate settings of hyper parameters improves the ability of the model and strength of prediction tasks. The model with protein-protein interaction properties achieves the highest prediction rate and minimum error rate than other two docking approaches. It is determined that the architecture of functional DNN and the representation learning of user defined features contribute to the overall performance in predicting binding affinity.

7.3 BINDING AFFINITY PREDICTION MODELS USING CUSTOMIZED LAYERS WITH DEEP NEURAL NETWORK

The functional deep models implemented on the three different datasets with three optimizers presented in the previous section are better at remembering the past inputs and connects with input through layer. But the weights not are trained in functional DNN due to which the error rate is not minimized. Since the layers are connected the modifications to any layer affects the whole model. The customized layers in DNN are used to solve this issue where the weights are pre-trained and the error rate is minimized. The weights are trained and passed to the layers as pre-trained weights that reduce cost function. The functional DNN does not possess the pre-trained weights and so that the error rate minimization is not possible. This section demonstrates the modelling of binding affinity prediction through DNN with customized layers and various optimizers using three different datasets PLD, PMLD and PPD.

Methodology

The DNN with customized layers is utilized to construct binding affinity predictive models through self learning the features of three datasets. The predictive models are constructed using three datasets PLD, PMLD and PPD. Three optimizers such as adam, RMSprop and Nadam are used with DNN architecture. The hyper parameters loss function and activation function are utilized here in addition to the hyper parameters used in previous cases. The loss function used here is sparse categorical crossentropy and the activation function used is softmax. The proposed framework of DNN with customized layer to build binding affinity predictive model is shown in Fig. 7.21.

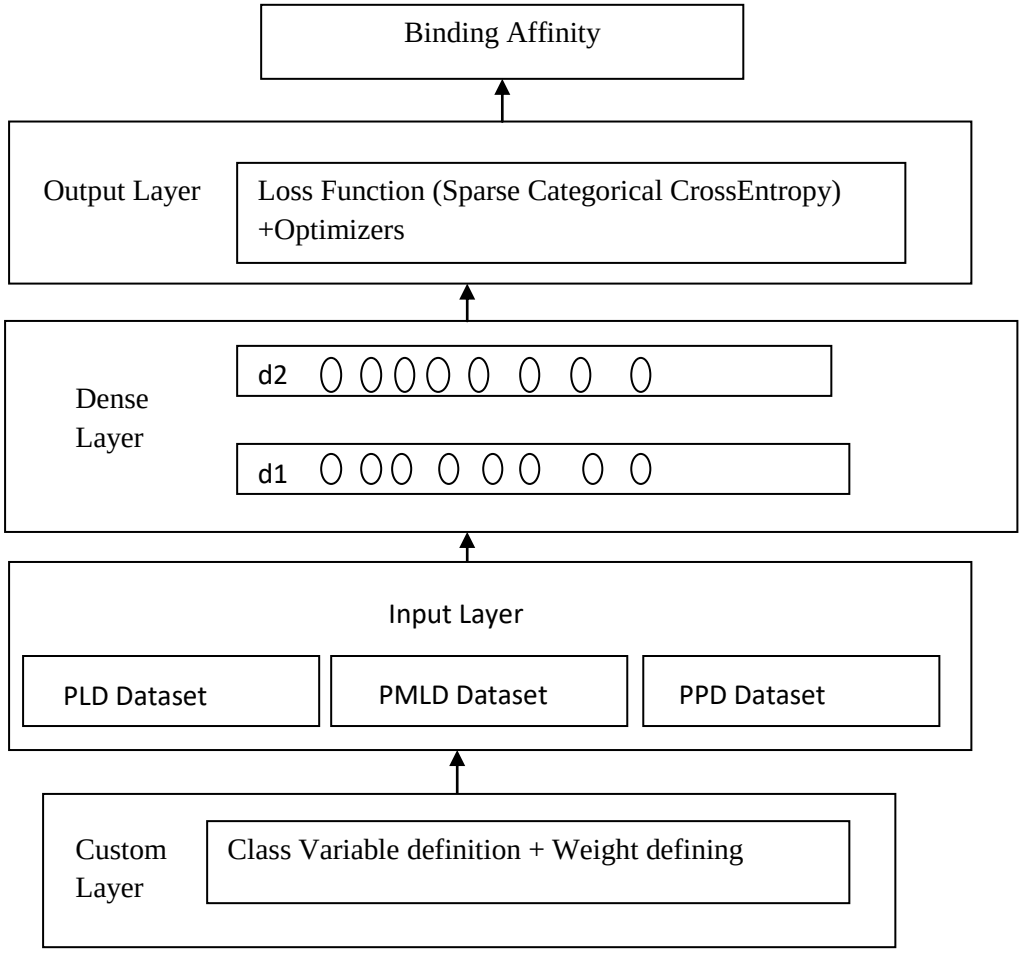

Fig. 7.21 Architecture of DNN with Customized Layer Based Prediction Model

The input feature vectors created as described in chapter 4, 5 and 6 corresponding to three datasets such as PLD, PMLD and PPD are directly fed as input to the customized layers with DNN architecture. Here custom layer is built by defining the class variable and super class variable. The trainable weights are defined and the custom layers are created subsequent

to variable definition. The weight is defined corresponding to input pattern and set in the kernel. The kernel is the custom functionality of the layer and it creates the weight using normal initializer.

In this technique, the input pattern is defined with number of inputs and 2 hidden layers where one layer has 20 neurons and other has 10 neurons. The output layer has one neuron which is used for prediction. The output layer is defined with softmax activation function and it is used to calculate the probabilities of each binding affinity value over all possible targets where it is computed as the ratio of the exponential of the input value and the sum of exponential values. This activation function is used to achieve the high probability. The loss function sparse categorical crossentropy loss function is used, since the targets are integer variables and suitable for prediction problems.

Finally, 10-fold cross validation is used to evaluate the performance of the models using various metrics such as explained variance score, mean squared error, root mean squared error, mean absolute error, median absolute error and R2 score.

Experiments and Results

Experiments have been carried out by implementing DNN with customized layers using keras in python with three datasets as inputs to build the prediction models. The network is trained with varying epochs of 50, 100, 200, 400, 500 and dropout values as 0.2, 0.3, 0.4, 0.5 are considered. The learning rate is fixed as 0.01 and the three optimizers like adam, RMSprop, Nadam with 2 hidden layers. Three optimizers are used to check the variations in prediction rate. With these parameter settings, the predictive models are built. The input dimension specified is 56 where the weight value specified initially is 56 and the weight value keeps on updating for each iteration.

Pre-trained weights are shared among the layers in which the hidden layer is 30 with single output layer. The performance of these predictive models is tested using 10-fold cross validation technique and the performance of the models are evaluated on metrics such as explained variance score, mean squared error, root mean squared error, mean absolute error, median absolute error and R2 score.

Results for Protein-Ligand Dataset

The results of binding affinity predictive models implemented through DNN architecture with customized layers and three optimizers such as adam, RMSprop, Nadam for various dropouts are presented below. The results of predictive model based on DNN with customized layers and adam optimizer is presented in Table 7.20.

Table 7.20 Results of DNN with Customized Layers Based Predictive Model and Adam Optimizer for Protein-Ligand Dataset

Dropout	**Epochs**	Explained variance score	R2 score	Mean squared error	Root mean squared error	Mean absolute error	Median absolute error
0.2	50	0.68	0.68	0.45	0.67	0.49	0.34
	100	0.70	0.70	0.32	0.57	0.35	0.23
	200	0.72	0.72	0.32	0.54	0.34	0.24
	400	0.76	0.76	0.30	0.52	0.30	0.22
	500	**0.85**	**0.85**	**0.20**	**0.44**	**0.25**	**0.15**
0.3	50	0.70	0.70	0.32	0.57	0.35	0.23
	100	0.75	0.75	0.30	0.59	0.39	0.27
	200	0.82	0.82	0.20	0.44	0.22	0.15
	400	0.85	0.85	0.20	0.44	0.25	0.15
	500	**0.95**	**0.95**	**0.11**	**0.30**	**0.12**	**0.4**
0.4	50	0.75	0.75	0.30	0.59	0.39	0.27
	100	0.82	0.82	0.20	0.44	0.22	0.15
	200	0.86	0.86	0.2	0.44	0.22	0.15
	400	0.89	0.89	0.2	0.44	0.24	0.10
	500	**0.90**	**0.90**	**0.15**	**0.34**	**0.15**	**0.6**
0.5	50	0.75	0.75	0.30	0.59	0.39	0.27
	100	0.82	0.82	0.20	0.44	0.22	0.15
	200	0.90	0.90	0.12	0.34	0.15	0.6
	400	0.85	0.85	0.20	0.44	0.25	0.15
	500	**0.89**	**0.89**	**0.18**	**0.35**	**0.16**	**0.8**

Table 7.20 shows that the predictive model built with adam optimizer achieves the better score. Prediction rate and error rate for the dropout 0.3 and epoch 500 is 0.95 and 0.11 respectively. The results for metrics such as root mean squared error, mean absolute error and median absolute error are 0.30, 0.12 and 0.4 respectively. The results of predictive model based on DNN with customized layers and RMSprop optimizer is presented in Table 7.21.

Table 7.21 Results of DNN with Customized Layers Based Predictive Model and RMSprop Optimizer for Protein-Ligand Dataset

Dropout	**Epochs**	Explained variance score	R2 score	Mean squared error	Root mean squared error	Mean absolute error	Median absolute error
0.2	50	0.68	0.68	0.45	0.67	0.49	0.34
	100	0.70	0.70	0.32	0.57	0.35	0.23
	200	0.72	0.72	0.32	0.54	0.34	0.24
	400	0.76	0.76	0.30	0.52	0.30	0.22
	500	**0.85**	**0.85**	**0.20**	**0.44**	**0.25**	**0.15**
0.3	50	0.70	0.70	0.32	0.57	0.35	0.23
	100	0.75	0.75	0.30	0.59	0.39	0.27
	200	0.82	0.82	0.20	0.44	0.22	0.15
	400	0.85	0.85	0.20	0.44	0.25	0.15
	500	**0.85**	**0.85**	**0.20**	**0.44**	**0.25**	**0.15**
0.4	50	0.75	0.75	0.30	0.59	0.39	0.27
	100	0.82	0.82	0.20	0.44	0.22	0.15
	200	0.70	0.70	0.32	0.57	0.35	0.23
	400	0.75	0.75	0.30	0.59	0.39	0.27
	500	**0.82**	**0.82**	**0.20**	**0.44**	**0.22**	**0.15**
0.5	50	0.75	0.75	0.30	0.59	0.39	0.27
	100	0.72	0.72	0.32	0.54	0.34	0.24
	200	0.68	0.68	0.45	0.67	0.49	0.34
	400	0.72	0.72	0.32	0.54	0.34	0.24
	500	**0.75**	**0.75**	**0.30**	**0.59**	**0.39**	**0.27**

Table 7.21, illustrate that the predictive model built with RMSprop optimizer achieves the prediction rate and error rate for the dropout 0.3 and epoch 500 is 0.85 and 0.20 respectively. The results for metrics such as root mean squared error, mean absolute error and median absolute error are 0.44, 0.25 and 0.15 respectively. The least prediction rate and error rate for the dropout 0.5 and epoch 500 is 0.75 and 0.30 respectively. The results for metrics such as root mean squared error, mean absolute error and median absolute error are 0.59, 0.39 and 0.27 respectively. The results of predictive model based on DNN with customized layers and Nadam optimizer is presented in Table 7.22.

Table 7.22 Results of DNN with Customized Layers Based Predictive Model and Nadam Optimizer for Protein-Ligand Dataset

Dropout	**Epochs**	Explained variance score	R2 score	Mean squared error	Root mean squared error	Mean absolute error	Median absolute error
0.2	50	0.68	0.68	0.45	0.67	0.49	0.34
	100	0.70	0.70	0.32	0.57	0.35	0.23
	200	0.72	0.72	0.32	0.54	0.34	0.24
	400	0.76	0.76	0.30	0.52	0.30	0.22
	500	**0.82**	**0.82**	**0.20**	**0.44**	**0.22**	**0.15**
0.3	50	0.70	0.70	0.32	0.57	0.35	0.23
	100	0.75	0.75	0.30	0.59	0.39	0.27
	200	0.82	0.82	0.20	0.44	0.22	0.15
	400	0.85	0.85	0.20	0.44	0.25	0.15
	500	**0.85**	**0.85**	**0.20**	**0.44**	**0.25**	**0.15**
0.4	50	0.75	0.75	0.30	0.59	0.39	0.27
	100	0.82	0.82	0.20	0.44	0.22	0.15
	200	0.86	0.86	0.2	0.44	0.22	0.15
	400	0.75	0.75	0.30	0.59	0.39	0.27
	500	**0.82**	**0.82**	**0.20**	**0.44**	**0.22**	**0.15**
0.5	50	0.75	0.75	0.30	0.59	0.39	0.27
	100	0.68	0.68	0.45	0.67	0.49	0.34
	200	0.70	0.70	0.32	0.57	0.35	0.23
	400	0.72	0.72	0.32	0.54	0.34	0.24
	500	**0.82**	**0.82**	**0.20**	**0.44**	**0.22**	**0.15**

The Table 7.22 demonstrates that the predictive model with Nadam optimizer produces the same result as that of RMSprop optimizer for the dropout of 0.3 and epoch of 500. Prediction rate and error rate obtained for the dropout rate of 0.3 and epoch 500 is 0.85 and 0.20 respectively. From the results, it is obvious that the predictive model learned with features of energy calculations in PLD dataset and network weight updation by adam optimizer, outperforms the predictive model fabricated with RMSprop and nadam optimizers. The results obtained for different dropouts and epochs for PLD dataset are illustrated from Fig. 7.22 to Fig. 7.24.

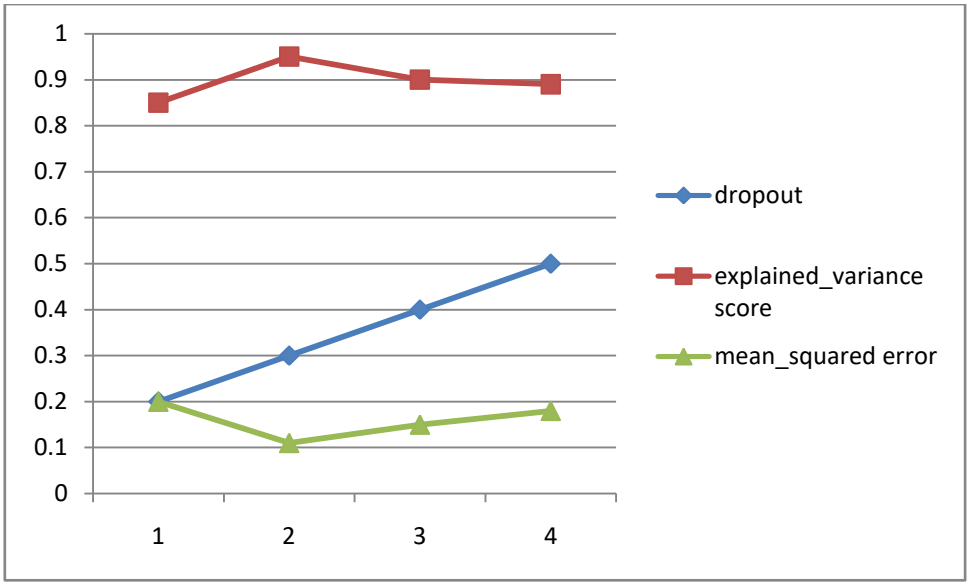

Fig. 7.22 Results of DNN with Customized Layers Based Predictive Model and Adam Optimizer at Various Dropouts for Protein-Ligand Dataset

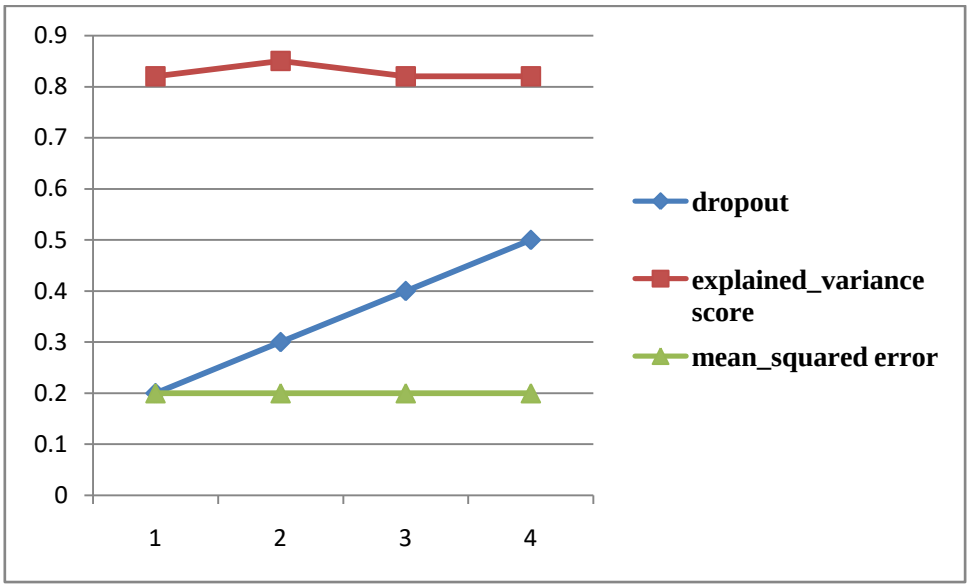

Fig. 7.23 Results of DNN with Customized Layers Based Predictive Model and RMSprop Optimizer at Various Dropouts for Protein-Ligand Dataset

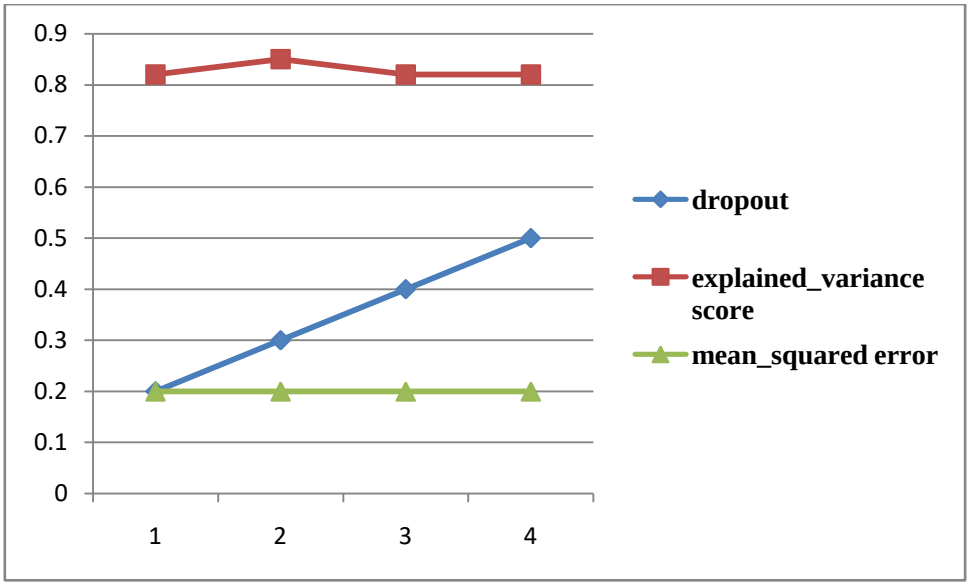

Fig. 7.24 Results of DNN with Customized Layers Based Predictive Model and Nadam Optimizer at Various Dropouts for Protein-Ligand Dataset

The charts from Fig. 7.22 to Fig. 7.24 reveal that the binding affinity predictive models based on PMLD dataset with adam optimizer outperforms the other two optimizers. The adam optimizer achieves the highest prediction rate of 0.95 and error rate of 0.11. The RMSprop and Nadam optimizer are achieved the inferior prediction rate and superior error rate. When comparing the predictive model with PLD and PMLD datasets, the predictive model with adam optimizer achieves higher prediction rate for PLD dataset.

Results for Protein-Mutated-Ligand Dataset

The results of DNN with customized layers based predictive models for protein-mutated-ligand dataset with three optimizers adam, RMSprop, Nadam for various dropouts are presented below. The results of predictive model based on DNN with customized layers and adam optimizer for protein-mutated-ligand dataset is tabulated in Table 7.23.

Table 7.23 Results of DNN with Customized Layers Based Predictive Model and adam Optimizer for Protein-Mutated-Ligand Dataset

Dropout	**Epochs**	Explained variance score	R2 score	Mean squared error	Root mean squared error	Mean absolute error	Median absolute error
0.2	50	0.68	0.68	0.45	0.67	0.49	0.34
	100	0.70	0.70	0.32	0.57	0.35	0.23
	200	0.72	0.72	0.32	0.54	0.34	0.24

	400	0.76	0.76	0.30	0.52	0.30	0.22
	500	**0.85**	**0.85**	**0.20**	**0.44**	**0.25**	**0.15**
0.3	50	0.70	0.70	0.32	0.57	0.35	0.23
	100	0.75	0.75	0.30	0.59	0.39	0.27
	200	0.82	0.82	0.20	0.44	0.22	0.15
	400	0.85	0.85	0.20	0.44	0.25	0.15
	500	**0.92**	**0.92**	**0.14**	**0.32**	**0.12**	**0.4**
0.4	50	0.75	0.75	0.30	0.59	0.39	0.27
	100	0.82	0.82	0.20	0.44	0.22	0.15
	200	0.86	0.86	0.2	0.44	0.22	0.15
	400	0.89	0.89	0.2	0.44	0.24	0.10
	500	**0.90**	**0.90**	**0.15**	**0.34**	**0.15**	**0.6**
0.5	50	0.75	0.75	0.30	0.59	0.39	0.27
	100	0.82	0.82	0.20	0.44	0.22	0.15
	200	0.90	0.90	0.12	0.34	0.15	0.6
	400	0.85	0.85	0.20	0.44	0.25	0.15
	500	**0.89**	**0.89**	**0.18**	**0.35**	**0.16**	**0.8**

Table 7.23, demonstrates that the predictive model with adam optimizer for PMLD dataset is higher, where the features like scoring functions and sequence descriptors along with energy calculations helps in achieving binding affinity precisely. The explained variance score and mean squared error is 0.92 and 0.14 obtained for the dropout 0.3. The results for root mean squared error, mean absolute error and median absolute error are 0.32, 0.12 and 0.4 respectively. The results of predictive model based on DNN with customized and RMSprop optimizer for protein-mutated-ligand dataset are tabulated in Table 7.24.

Table 7.24 Results of Customized Layers with DNN Based Predictive Model and RMSprop Optimizer for Protein-Mutated-Ligand Dataset

Dropout	**Epochs**	Explained variance score	R2 score	Mean squared error	Root mean squared error	Mean absolute error	Median absolute error
0.2	50	0.68	0.68	0.45	0.67	0.49	0.34
	100	0.70	0.70	0.32	0.57	0.35	0.23
	200	0.72	0.72	0.32	0.54	0.34	0.24
	400	0.76	0.76	0.30	0.52	0.30	0.22
	500	**0.85**	**0.85**	**0.20**	**0.44**	**0.25**	**0.15**
0.3	50	0.70	0.70	0.32	0.57	0.35	0.23

	100	0.75	0.75	0.30	0.59	0.39	0.27
	200	0.82	0.82	0.20	0.44	0.22	0.15
	400	0.85	0.85	0.20	0.44	0.25	0.15
	500	**0.90**	**0.90**	**0.15**	**0.34**	**0.15**	**0.6**
0.4	50	0.75	0.75	0.30	0.59	0.39	0.27
	100	0.82	0.82	0.20	0.44	0.22	0.15
	200	0.86	0.86	0.2	0.44	0.22	0.15
	400	0.89	0.89	0.2	0.44	0.24	0.10
	500	**0.90**	**0.90**	**0.15**	**0.34**	**0.15**	**0.6**
0.5	50	0.75	0.75	0.30	0.59	0.39	0.27
	100	0.82	0.82	0.20	0.44	0.22	0.15
	200	0.90	0.90	0.12	0.34	0.15	0.6
	400	0.85	0.85	0.20	0.44	0.25	0.15
	500	**0.89**	**0.89**	**0.18**	**0.35**	**0.16**	**0.8**

Table 7.24 reveals that the prediction rate is lower in the model built with rmsprop optimizer than the model constructed with adam optimizer, where the adam optimizer updates the learning rate of each parameter. The value of explained variance score is 0.90 and mean squared error is 0.15 with the dropout rate of 0.3 and at the 500 epoch. The results for root mean squared error, mean absolute error and median absolute error are 0.34, 0.15 and 0.6 respectively. The results of predictive model based on DNN with customized layers and Nadam optimizer for protein-mutated-ligand dataset are tabulated in Table 7.25.

Table 7.25 Results of Customized Layers with DNN Based Predictive Model and Nadam Optimizer for Protein-Mutated-Ligand Dataset

Dropout	**Epochs**	Explained variance score	R2 score	Mean squared error	Root mean squared error	Mean absolute error	Median absolute error
0.2	50	0.68	0.68	0.45	0.67	0.49	0.34
	100	0.70	0.70	0.32	0.57	0.35	0.23
	200	0.72	0.72	0.32	0.54	0.34	0.24
	400	0.76	0.76	0.30	0.52	0.30	0.22
	500	**0.85**	**0.85**	**0.20**	**0.44**	**0.25**	**0.15**
0.3	50	0.70	0.70	0.32	0.57	0.35	0.23
	100	0.75	0.75	0.30	0.59	0.39	0.27
	200	0.82	0.82	0.20	0.44	0.22	0.15
	400	0.85	0.85	0.20	0.44	0.25	0.15
	500	**0.90**	**0.90**	**0.15**	**0.34**	**0.15**	**0.6**

0.4	50	0.75	0.75	0.30	0.59	0.39	0.27
	100	0.82	0.82	0.20	0.44	0.22	0.15
	200	0.86	0.86	0.2	0.44	0.22	0.15
	400	0.89	0.89	0.2	0.44	0.24	0.10
	500	**0.90**	**0.90**	**0.15**	**0.34**	**0.15**	**0.6**
0.5	50	0.75	0.75	0.30	0.59	0.39	0.27
	100	0.82	0.82	0.20	0.44	0.22	0.15
	200	0.90	0.90	0.12	0.34	0.15	0.6
	400	0.85	0.85	0.20	0.44	0.25	0.15
	500	**0.89**	**0.89**	**0.18**	**0.35**	**0.16**	**0.8**

Table 7.25 illustrates that the prediction rate of the model trained with nadam optimizer is 0.90 and the mean squared error is 0.90 at the dropout rate of 0.3 and at the epoch of 500. From the above table, it is evident that the predictive model trained with the features of scoring functions, sequence descriptors aids in achieving better prediction rate. The learning rate updation for each parameter by adam optimizer, outperforms the models built with rmsprop and nadam optimizers. The results obtained for different dropouts and epochs are illustrated from Fig. 7.25 to Fig. 7.27.

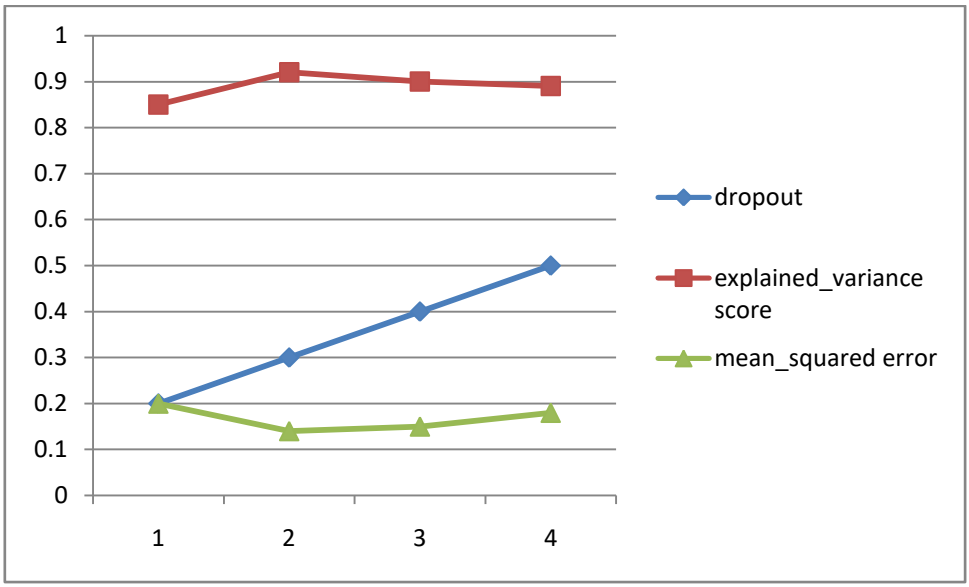

Fig. 7.25 Results of DNN with Customized Layers Based Predictive Model and Adam Optimizer at Various Dropouts for Protein-Mutated-Ligand Dataset

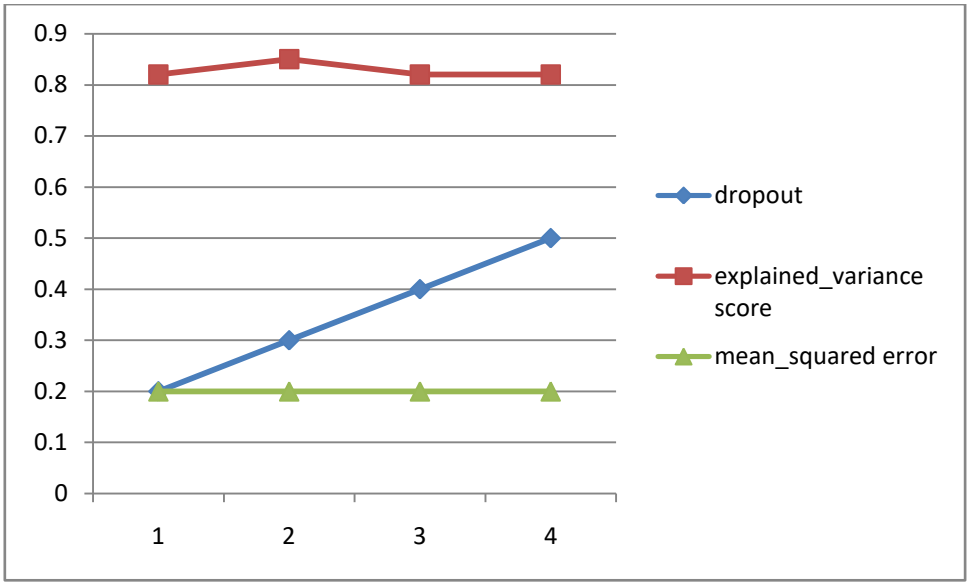

Fig. 7.26 Results of DNN with Customized Layers Based Predictive Model and RMSprop Optimizer at Various Dropouts for Protein-Mutated-Ligand Dataset

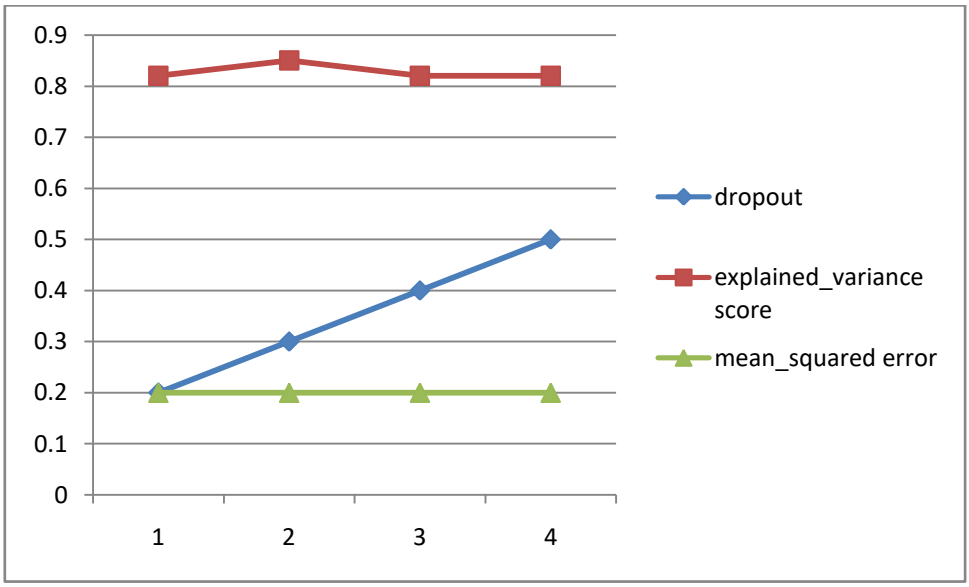

Fig. 7.27 Results of DNN with Customized Layers Based Predictive Model and Nadam Optimizer at Various Dropouts for Protein-Mutated-Ligand Dataset

The charts from Fig. 7.25 to Fig. 7.27 reveal that the binding affinity predictive models for PMLD dataset with adam optimizer outperforms the other two optimizers. The adam optimizer achieves the highest prediction rate of 0.92 and error rate of 0.14. The RMSprop and Nadam optimizer are achieved the inferior prediction rate and superior error rate. When comparing the PLD and PMLD datasets, PLD dataset achieves higher prediction rate when trained the network with adam optimizer.

Results for Protein-protein Dataset

The results of predictive models based on DNN with customized layers for protein-protein dataset with three optimizers such as adam, RMSprop, Nadam for various dropouts are presented below. The results of DNN with customized layers based model and adam optimizer for protein-protein dataset is given in Table 7.26.

Table 7.26 Results of DNN with Customized Layers and adam Optimizer for Protein-Protein Dataset

Dropout	**Epochs**	Explained variance score	R2 score	Mean squared error	Root mean squared error	Mean absolute error	Median absolute error
0.2	50	0.68	0.68	0.45	0.67	0.49	0.34
	100	0.70	0.70	0.32	0.57	0.35	0.23
	200	0.75	0.75	0.30	0.59	0.39	0.27
	400	0.80	0.80	0.20	0.44	0.28	0.19
	500	**0.94**	**0.94**	**0.13**	**0.36**	**0.15**	**0.14**
0.3	50	0.70	0.70	0.32	0.57	0.35	0.23
	100	0.75	0.75	0.30	0.59	0.39	0.27
	200	0.82	0.82	0.20	0.44	0.22	0.15
	400	0.90	0.90	0.12	0.34	0.15	0.6
	500	**0.98**	**0.98**	**0.01**	**0.29**	**0.09**	**0.3**
0.4	50	0.75	0.75	0.30	0.59	0.39	0.27
	100	0.82	0.82	0.20	0.44	0.22	0.15
	200	0.86	0.86	0.2	0.44	0.22	0.15
	400	0.89	0.89	0.2	0.44	0.4	0.10
	500	**0.90**	**0.90**	**0.15**	**0.34**	**0.15**	**0.6**
0.5	50	0.75	0.75	0.30	0.59	0.39	0.27
	100	0.82	0.82	0.20	0.44	0.22	0.15
	200	0.90	0.90	0.12	0.34	0.15	0.6
	400	0.80	0.80	0.20	0.44	0.28	0.19
	500	**0.92**	**0.92**	**0.14**	**0.32**	**0.12**	**0.4**

Table 7.26 proves that the prediction rate and error rate for the dropout 0.3 and epoch 500 is 0.98 and 0.01 respectively. The results for root mean squared error, median absolute error, mean absolute error and R2 score attained at the dropout 0.3 are 0.29, 0.09, 0.3 and 0.98 respectively. When the dropout is 0.2 and epoch 500; the prediction rate and error rate

obtained is 0.94 and 0.13 respectively. The predictive model acquired the better score in dropout 0.3 and epoch 500. The results of predictive models based on DNN with customized layers and RMSprop optimizer for protein-protein dataset are given in Table 7.27.

Table 7.27 Results of DNN with Customized Layers Based Predictive Model and RMSprop optimizer for Protein-Protein Dataset

Dropout	**Epochs**	Explained variance score	R2 score	Mean squared error	Root mean squared error	Mean absolute error	Median absolute error
0.2	50	0.68	0.68	0.45	0.67	0.49	0.34
	100	0.70	0.70	0.32	0.57	0.35	0.23
	200	0.75	0.75	0.30	0.59	0.39	0.27
	400	0.80	0.80	0.20	0.44	0.28	0.19
	500	**0.85**	**0.85**	**0.20**	**0.44**	**0.28**	**0.19**
0.3	50	0.75	0.75	0.30	0.59	0.39	0.27
	100	0.82	0.82	0.20	0.44	0.22	0.15
	200	0.86	0.86	0.2	0.44	0.22	0.15
	400	0.89	0.89	0.2	0.44	0.24	0.10
	500	**0.90**	**0.90**	**0.15**	**0.34**	**0.15**	**0.6**
0.4	50	0.75	0.75	0.30	0.59	0.39	0.27
	100	0.82	0.82	0.20	0.44	0.22	0.15
	200	0.90	0.90	0.12	0.34	0.15	0.6
	400	0.85	0.85	0.20	0.44	0.25	0.15
	500	**0.89**	**0.89**	**0.18**	**0.35**	**0.16**	**0.8**
0.5	50	0.75	0.75	0.30	0.59	0.39	0.27
	100	0.82	0.82	0.20	0.44	0.22	0.15
	200	0.90	0.90	0.12	0.4	0.15	0.6
	400	0.80	0.80	0.20	0.44	0.28	0.19
	500	**0.89**	**0.89**	**0.18**	**0.3**	**0.16**	**0.8**

Table 7.27 shows that the prediction rate and error rate for the dropout 0.3 and for the epoch 500 is 0.90 and 0.15 respectively. The results for root mean squared error, median absolute error, mean absolute error and R2 score attained at the dropout 0.3 are 0.34, 0.15, 0.6 and 0.90 respectively. The dropout 0.4, 0.5 and epoch 500 obtained the prediction rate and error rate of 0.89 and error rate of 0.18. This predictive model acquired the better score

for dropout is 0.3 and epoch is 500. The results of predictive models based on DNN with customized layers and Nadam optimizer for protein-protein dataset are given in Table 7.28.

Table 7.28 Results of DNN with Customized Layers Based Predictive Model and Nadam optimizer for Protein-Protein Dataset

Dropout	**Epochs**	Explained variance score	R2 score	Mean squared error	Root mean squared error	Mean absolute error	Median absolute error
0.2	50	0.68	0.68	0.45	0.67	0.49	0.34
	100	0.70	0.70	0.32	0.57	0.35	0.23
	200	0.72	0.72	0.32	0.54	0.34	0.24
	400	0.76	0.76	0.30	0.52	0.30	0.22
	500	**0.85**	**0.85**	**0.20**	**0.44**	**0.25**	**0.15**
0.3	50	0.70	0.70	0.32	0.57	0.35	0.23
	100	0.75	0.75	0.30	0.59	0.39	0.27
	200	0.82	0.82	0.20	0.44	0.22	0.15
	400	0.85	0.85	0.20	0.44	0.25	0.15
	500	**0.90**	**0.90**	**0.15**	**0.34**	**0.15**	**0.6**
0.4	50	0.75	0.75	0.30	0.59	0.39	0.27
	100	0.82	0.82	0.20	0.44	0.22	0.15
	200	0.86	0.86	0.2	0.44	0.22	0.15
	400	0.89	0.89	0.2	0.44	0.24	0.10
	500	**0.90**	**0.90**	**0.15**	**0.34**	**0.15**	**0.6**
0.5	50	0.75	0.75	0.30	0.59	0.39	0.27
	100	0.82	0.82	0.20	0.44	0.22	0.15
	200	0.90	0.90	0.12	0.34	0.15	0.6
	400	0.85	0.85	0.20	0.44	0.25	0.15
	500	**0.89**	**0.89**	**0.18**	**0.35**	**0.16**	**0.8**

Table 7.28 reveals that DNN trained with Nadam optimizer yields the same score as that of network with RMSprop optimizer for the dropout value of 0.3 and epoch value of 500. It is evident, that adam optimizer based prediction model outperforms other optimizers where explained variance score and mean squared error for the dropout 0.3 and epoch 500 is 0.98 and 0.01 respectively. The results for R2 score, mean absolute error, median absolute error and root mean squared error obtained are 0.9, 0.09, 0.3 and 0.01 respectively where the error

rate of the model is minimum and prediction rate is high. The results of different dropouts and epochs obtained for PPD dataset are illustrated from Fig. 7.28 to Fig. 7.30.

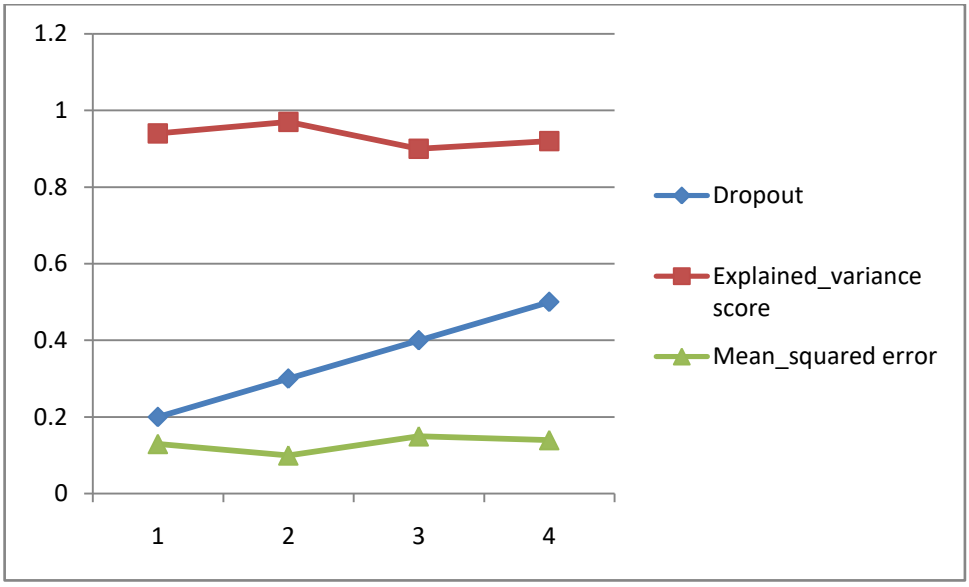

Fig. 7.28 Results of Binding Affinity Predictive Model Based on DNN with Customized Layers and Adam optimizer at Various Dropouts for Protein-Protein Dataset

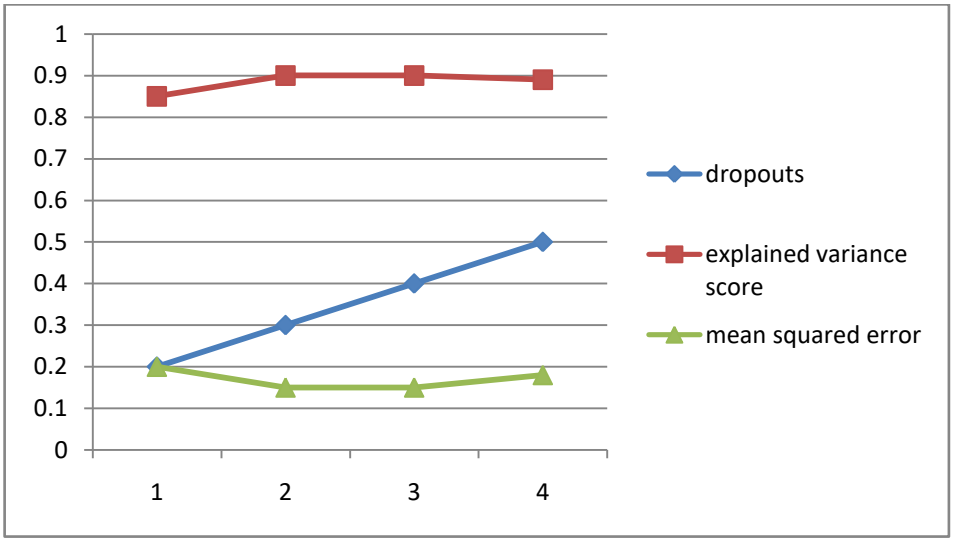

Fig. 7.29 Results of Binding Affinity Predictive Model Based on DNN with Customized Layers and RMSprop Optimizer at Various Dropouts for Protein-Protein Dataset

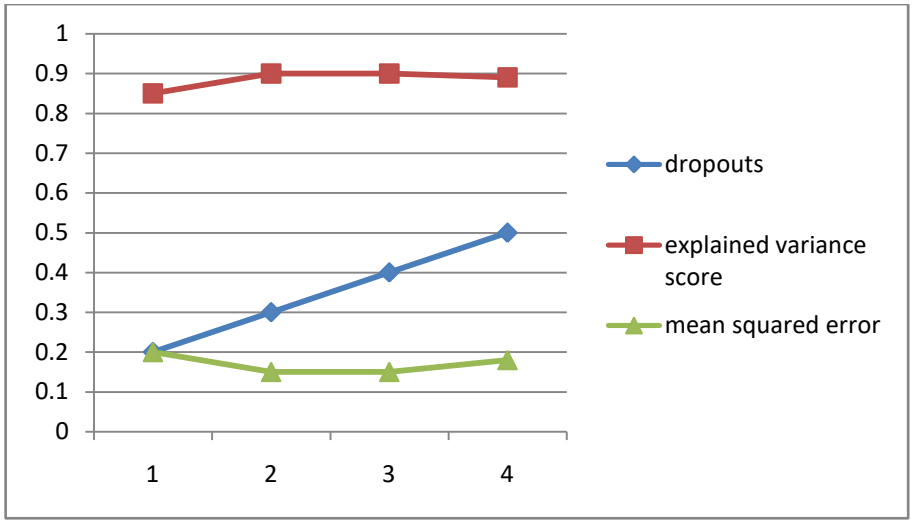

Fig. 7.30 Results of Binding Affinity Predictive Model Based on DNN with Customized Layers and Nadam Optimizer at Various Dropouts for Protein-Protein Dataset

The charts from Fig. 7.28 to Fig. 7.30 reveal that the binding affinity predictive models trained with PPD dataset and adam optimizer outperforms the other two optimizers. The adam optimizer achieves the highest prediction rate of 0.98 and error rate of 0.01. The RMSprop and Nadam optimizer are achieved the inferior prediction rate and superior error rate. When comparing three datasets, PPD dataset achieves higher prediction rate with adam optimizer.

Findings

The deep architecture based on deep neural network with customized layers to predict binding affinity prediction for SCA is evident than the functional DNN. It is obvious that the performance of DNN with customized layers is relatively higher than the functional DNN for all the three predictive models based on protein-ligand docking, protein-mutated-ligand docking and protein-protein interaction. The high prediction rate of model based on DNN with customized layers proves that the network is able to learn efficiently with two separate dense layers by having the the custom layer with pre trained weights. The pre trained weights to the input layer minimizes the error rate and improves the prediction rate. The predictive model based on DNN with customized layers with appropriate settings of hyper parameters improves the ability of the model and strength of prediction tasks. The predictive model based on DNN with customized layers based on protein-protein interaction properties achieves the highest prediction rate and minimum error rate. It is determined that the

architecture of DNN with customized layers and the representation learning of user defined features contribute to the overall performance in predicting binding affinity.

7.4 COMPARATIVE ANALYSIS

The comparative analysis is divided into two phases. In the first phase, the comparison of binding affinity models based on three deep neural network architectures, sequential DNN, functional DNN, DNN with customized layers is done. In the second phase, the comparison of deep models with regression based binding affinity models is performed. In case of deep learning, DNN with adam optimizer performs better when dropout is 0.3 and the epoch is 500 than other two optimizers for all the three datasets and hence the efficiency of such DNN models in predicting binding affinity is compared and analyzed.

Comparison of DNN Based Predictive Models

The comparative results with respect to various performance metrics such as explained variance score, mean squared error, root mean squared error, R2 score, median absolute error and mean absolute error, of DNN based predictive models implemented with three datasets PLD, PMLD, PPD is reported in Table 7.29.

Table 7.29 Comparative Results of DNN Based Binding Affinity Predictive Models

Metrics/Algorithms	**Sequential DNN**			**Functional DNN**			**Customized Layers with DNN**		
Datasets	**PLD**	**PMLD**	**PPD**	**PLD**	**PMLD**	**PPD**	**PLD**	**PMLD**	**PPD**
Explained Variance Score	0.95	0.85	**0.95**	0.97	0.90	**0.97**	0.95	0.92	**0.98**
R2 Score	0.95	0.85	**0.95**	0.97	0.90	**0.97**	0.95	0.92	**0.98**
Mean Squared Error	0.1	0.20	**0.1**	0.1	0.15	**0.1**	0.1	0.14	**0.01**
Root Mean Squared Error	0.32	0.44	**0.32**	0.32	0.34	**0.32**	0.29	0.32	**0.29**
Mean Absolute Error	0.12	0.25	**0.12**	0.12	0.15	**0.12**	0.09	0.12	**0.09**
Median Absolute Error	0.4	0.15	**0.4**	0.4	0.6	**0.4**	0.3	0.4	**0.3**

From the Table 7.29 it is evident that the predictive models based on DNN with customized layers performed better than the other two architectures sequential DNN and functional DNN for all the three datasets. The prediction rate obtained for model based on DNN with customized layers and PLD dataset is 0.95 and the error rate is 0.1 and the results for root mean squared error, mean absolute error, median absolute error and R2 score are 0.29, 0.09, 0.3 and 0.95 respectively. The same prediction rate is obtained with sequential DNN based model also. The prediction rate obtained for model based on DNN with customized layers and PMLD dataset is 0.92 and the error rate is 0.14 and the results for root mean squared error, mean absolute error, median absolute error and R2 score are 0.32, 0.12, 0.4 and 0.92 respectively.

The prediction rate obtained for model based on DNN with customized layers and PMLD dataset is 0.98 and the error rate is 0.01. The results for root mean squared error, mean absolute error, median absolute error and R2 score are 0.29, 0.09, 0.3 and 0.98 respectively. Among all the DNN architectures, DNN with customized layers for PPD dataset produced the better results when compared with the sequential DNN models and functional DNN models. The comparative results of DNN based predictive models are illustrated in Fig. 7.31.

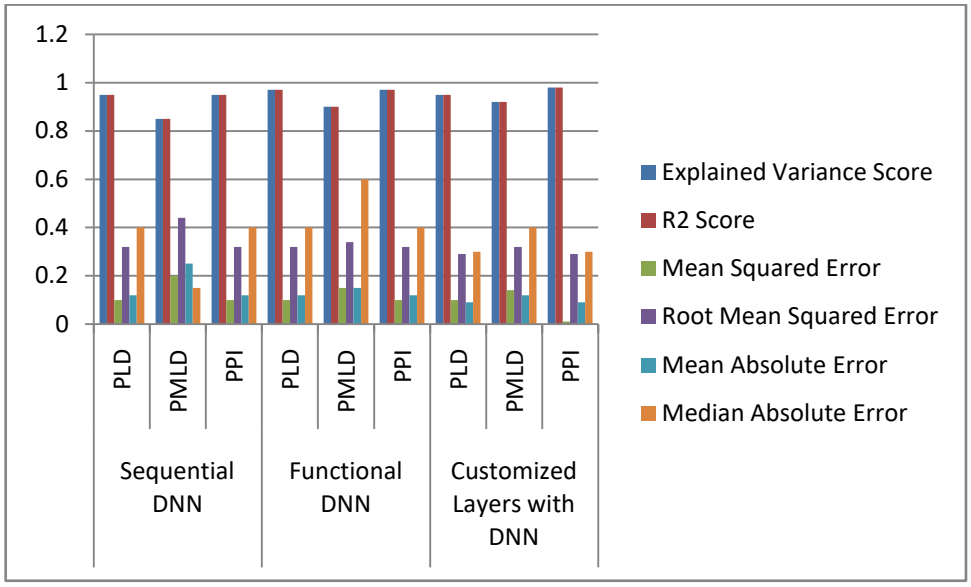

Fig. 7.31 Comparative Results of DNN Based Predictive Models

Fig. 7.31 reveals that the predictive model based on DNN with customized layers achieved the better prediction rate for all the three datasets. The predictive models based on DNN with customized layers for PLD, PMLD and PPD dataset achieved the prediction rate of 0.95, 0.92 and 0.98 respectively. The error rate obtained for three datasets using DNN with customized layers are 0.1, 0.14 and 0.01 respectively. It proves that DNN with customized

layers highly contribute in predicting binding affinity than the sequential DNN and functional DNN.

Comparison of DNN Based Predictive Models with Regression Models

In the second phase, the regression based predictive models are compared with DNN models and the results are analyzed. The performance results of predictive models are compared with respect to various metrics such as explained variance score, mean squared error, root mean squared error, R2 score, median absolute error and mean absolute error for all three datasets. The comparative results are presented in Table 7.30.

Table 7.30 Comparative Results of DNN Based Binding Affinity Predictive Models and Regression Models

Metrics		Explained Variance Score	R2 Score	Mean Squared Error	Root Mean Squared Error	Mean Absolute Error	Median Absolute Error
Algorithms	Datasets						
Sequential DNN	PLD	0.95	0.95	0.1	0.32	0.12	0.4
	PMLD	0.85	0.85	0.20	0.44	0.25	0.15
	PPD	0.95	0.95	0.1	0.32	0.12	0.4
Functional DNN	PLD	0.97	0.97	0.1	0.32	0.12	0.4
	PMLD	0.90	0.90	0.15	0.34	0.15	0.6
	PPD	0.97	0.97	0.1	0.32	0.12	0.4
Customized Layers with DNN	PLD	**0.95**	**0.95**	**0.1**	**0.29**	**0.09**	**0.3**
	PMLD	**0.92**	**0.92**	**0.14**	**0.32**	**0.12**	**0.4**
	PPD	**0.98**	**0.98**	**0.01**	**0.29**	**0.09**	**0.3**
Random Forest	PLD	0.85	0.85	0.20	0.44	0.15	0.25
	PMLD	0.87	0.87	0.2	0.4	0.15	0.22
	PPD	0.89	0.89	0.2	0.44	0.10	0.24
Linear Regression	PLD	0.70	0.70	0.32	0.57	0.23	0.35
	PMLD	0.68	0.68	0.45	0.67	0.34	0.49
	PPD	0.82	0.82	0.20	0.44	0.15	0.22
Support Vector Regression	PLD	0.76	0.76	0.30	0.57	0.22	0.30
	PMLD	0.70	0.70	0.32	0.57	0.23	0.35
	PPD	0.86	0.86	0.2	0.44	0.15	0.22
Artificial Neural	PLD	0.82	0.82	0.20	0.44	0.15	0.22
	PMLD	0.75	0.75	0.30	0.59	0.27	0.39

Network	PPD	0.76	0.76	0.30	0.52	0.22	0.30

From the Table 7.30 it is evident that the predictive model based on DNN with customized layers performed better for all the three datasets than the other architectures such as sequential DNN, functional DNN and regression algorithms such as linear regression, random forest, support vector regression and artificial neural network. The prediction rate for the model DNN with customized layers obtained for PLD Dataset is 0.95 whereas the prediction rate for the models with PLD and PMLD dataset are 0.97 and 0.98 respectively. The error rate for the model with PLD dataset is 0.1 whereas the error rate for the models with PMLD dataset and PPD dataset is 0.14 and 0.1 respectively. The results for the model with PLD dataset obtains the root mean squared error, mean absolute error, median absolute error and R2 score as 0.29, 0.09, 0.3 and 0.95 respectively. The predictive model with PMLD dataset obtains the root mean squared error, mean absolute error, median absolute error and R2 score as 0.32, 0.12, 0.4 and 0.92 respectively. The predictive model with PPD dataset obtains the root mean squared error, mean absolute error, median absolute error and R2 score as 0.29, 0.09, 0.3 and 0.98 respectively. Among the three datasets, PPD dataset achieves the highest prediction rate as it interacts with full molecule for interfacial contacts. The predictive model based on random forest algorithm produces better result than the other predictive models in machine learning. The random forest model obtained the prediction rate and error rate for PLD dataset is 0.85 and 0.20 respectively whereas the model for PMLD dataset obtained 0.87 and 0.2 respectively as the prediction rate and error rate. The model with PPD dataset attained the value of 0.89 and 0.2 respectively as prediction rate and error rate. The results for root mean squared error, mean absolute error, median absolute error and R2 score obtained for PLD dataset is 0.44, 0.15, 0.25 and 0.85 respectively whereas the results for root mean squared error, mean absolute error, median absolute error and R2 score obtained for PMLD dataset is 0.4, 0.15, 0.22 and 0.87 respectively. The results for root mean squared error, mean absolute error, median absolute error and R2 score obtained for PPD dataset is 0.44, 0.10, 0.24 and 0.89 respectively. The predictive model with PPD dataset achieves the highest prediction in machine learning also. This proves the deep models using DNN with customized layers are best suited for prediction problems. The comparative results of DNN based predictive models are illustrated in Fig. 7.32.

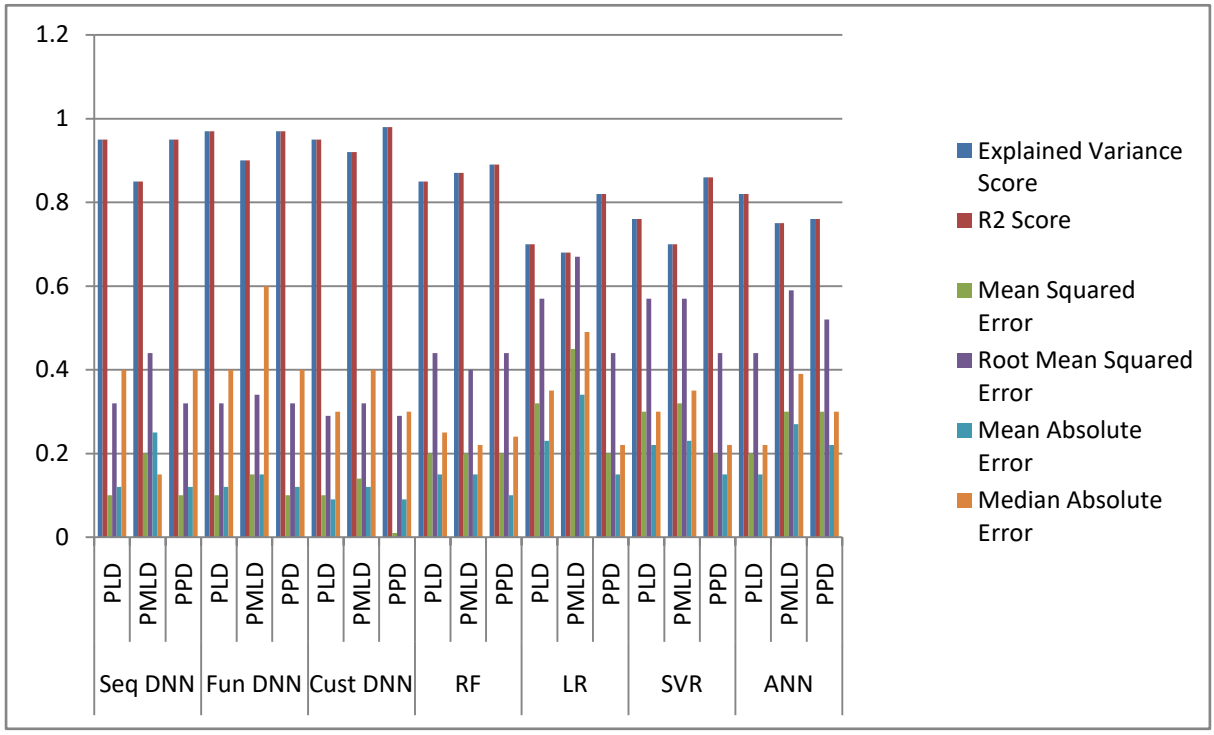

Fig. 7.32 Comparative Results of DNN Based Models and Regression Models

Fig. 7.32 confirms that the DNN with customized layer achieves the better prediction rate for all the datasets than other architectures. Among the regression algorithms, random forest performs better than the other algorithms with high prediction rate and low error rate. When comparing the random forest based model with the model based on DNN with customized layers, the model based on DNN with customized layers achieves high prediction rate and low error rate for all the three datasets. It shows that the predictive model based on DNN with customized layer is well suited for binding affinity prediction problems.

Findings

The comparative results show that the deep learning models perform better than shallow regression models. Network with user defined layers achieves the higher performance since it has pre-trained weights that are shared among the layers which helps in error minimization and improves the prediction rate. The network weight and learning rate updation by adam optimizer demonstrates the high prediction rate. Appropriate definition of hyper parameters facilitates in optimizing the models. The binding affinity predictive models built through protein-protein interaction using DNN with customized layers achieve the highest explained score than the other two deep architectures. This proves that deep learning can perform well on hand crafted features with representation learning. The comparison of

DNN predictive models with traditional regression algorithms shows that the DNN with customized layer is best suited in predicting binding affinity. The error rate produced for the predictive models through protein-protein interaction is less for DNN with customized layers and hence this approach is suitable for prediction of binding affinity for any disorders.

SUMMARY

This chapter illustrated the modelling of binding affinity prediction as regression task and its implementation with three kinds of DNN architectures such as sequential DNN, functional DNN, DNN with customized layers. The binding affinity predictive models built with three datasets have been explained in this chapter. The experimental results of three DNN architectures have been reported and the comparative analysis is presented. The comparison of DNN based models and regression based binding affinity predictive models with respect to various evaluation metrics has been illustrated with tables and charts. The conclusion of the thesis, summary of the findings and contributions made in this research will be presented in subsequent chapter.

Remarks

1. *The paper titled, Affinity Prediction of Spinocerebellar Ataxia Using Protein-Ligand and Protein-Protein Interactions with Functional Deep Learning has been published in International Journal of Engineering and Advanced Technology, Vol 8, No 5, pp 858-865, June 2019* ***(Scopus Indexed)***
2. *The paper titled, Affinity Prediction of Spinocerebellar Ataxia using Protein-Protein Interactions and Deep Neural Network with User-Defined Layer has been published in International Journal of Advanced science and Technology, Vol 28, No 13, PP 20-37, 2019* ***(Scopus Indexed)***
3. *The paper titled, Binding Affinity Predictions by Representation Learning of Various Docked Complexes in Spinocerebellar Ataxia has been published in Test Engineering and Management, Vol 83, PP 22571-22583, Issue- March-april 2020* ***(Scopus Indexed)***
4. *The paper titled, Binding Affinity Prediction of SCA using PLD and PMLD with Functional DNN and its Variants has been published in International Journal of Scientific and Technology Research Vol 9, No 6, pp 382-388, June 2020* ***(Scopus Indexed)***