Novel Approach of Deep Learning in Toxicity Prediction

Adhithiyan M, Karmel A

ABSTRACT--- Humans are always exposed to various harmful, harmless chemicals everyday. toxicity prediction is the method to find the toxicity of the chemicals , ie it is Toxic or Non toxic. among all the applications the toxicity prediction isvery much important as it involves large amount of expenses, chemicals, labour, etc. in the world of big data and artificial intelligence, toxicity prediction can be done effectively using machine learning and deep learning instead of drug evaluations in lab such as cellular, animal and clinical methods. in this paper we review machine learning methods to predict toxicity and extention of toxicity testing using deep learning such as DNN. we discuss about the molecular descriptors and certain endpoints and its relationship.

Index Terms – Toxicity prediction, machine learning, deep learning, molecular descriptors, endpoints.

1. INTRODUCTION

In our day to day scenario our human easily skin gets vulnerable to various chemical substances such as cosmetics, particles and regular harmful and harmless chemicals. but we doesn't know which chemical causes adverse effects and worse case like non acute and sub-acute poisoning which finally resulted in allergies. it may also leads to organ failure even deaths. this is occurring due to the toxic nature of that particular chemical. to avoid these issues every compound must be tested under certain experiments. but through modelling techniques like QSAR's we can predict the level of toxicity from the molecular descriptors of the chemicals developmental toxicity , acute toxicity are some examples of toxicity measures which can be predicted using the concept of Quantitative Structural Activity Relationship . usually the cost and time involved in testing of these compounds in test animals are very high. it will be much more effective if the scientists can predict the toxic levels and response of a compound using modelling approaches like QSAR

2. RELATED WORK

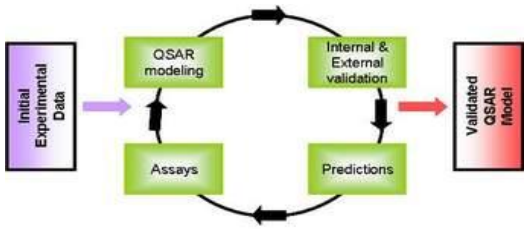
To identify the harmful effects caused by the chemicals it is very important to calculate toxicity levels . those chemicals create impacts on humans, plants and even animals. the toxicity prediction plays very crucial role in drug design. usually animal models are used for toxicity testing but in vivo animal tests are restricted by time, costing and some moralistic considerations.

Considering those factors scientists preferred computational methods instead of usual methods for predicting the levels of toxicity. the technology which enables to analyse, replicate, project and calculate the toxicity of the compounds through computations is called as Insilco testing. the main goal of Insilco technology is to improve the existing tests and to predicting the toxicity. In another hands prioritize the compounds and the and to reduce the final stage failures in new drug discovery there is always the continuous development in the technology of Insilco method by developing new model, upgrading the already existing models and removing certain models by validating it. but the problem is the model which work with very good accuracy for certain end point may not work better with different end point.

To overcome this problem one must have strong domain knowledge such as limitations, possibility and detailed analysis of that compound as well as end point. it is always necessary to choose the effective model to attain the best level of accuracy and improving it wisely.

3. METHODOLOGY

Quantitative Structural Activity Relationship (QSAR) modelling is found by Corvin Hansch and it is based on the belief that there is some interconnection between the structure of molecules and biological activity. due to that assumption QSAR tries to find out correlation between molecular properties and from the molecules and some experimental biological endpoints.



Qsar has been devoloped vastly and evolved from application of small compounds which belong to same genus using simple regression analysis to very large datasets consisting thousands of molecular structures which has large amount of moleular descriptors using statistical and machine learning models.

3a. Chemical Descriptors

In the paper we will find the word molecular descriptors. let us see what is molecular descriptors. it is the output of logical and statistical calculations which converts chemical informations which is encrypted within symbolic representation of molecule into usefull number of some experiment.

Revised Manuscript Received on February 11, 2019

Adhithiyan M, M.Tech SCSE, VIT University, Chennai Campus. (E-mail: adhithiyan@ymail.com)

Dr.Karmel A, SCSE at VIT University, Chennai Campus. (E-mail: karmel.a@vit.ac.in)

The biological activity of molecules is usually measured in assays to establish the level of inhibition of particular signal transduction or metabolic pathways. Drug discovery often involves the use of QSAR to identify chemical structures that could have good inhibitory effects on specific targets and have low toxicity (non-specific activity). Of special interest is the prediction of partition coefficient log P, which is an important measure used in identifying "druglikeness" according to Lipinski's Rule of Five.

While many quantitative structure activity relationship analyses involve the interactions of a family of molecules with an enzyme or receptor binding site, QSAR can also be used to study the interactions between the structural domains of proteins. Protein-protein interactions can be quantitatively analyzed for structural variations resulted from site-directed mutagenesis.

It is part of the machine learning method to reduce the risk for a SAR paradox, especially taking into account that only a finite amount of data is available.In general, all QSAR problems can be divided into coding and learning.

Due to the demands of time and the high cost of testing compounds for toxicity in test animals, it would be an advantage to be able to estimate the toxic response of chemical agents using theoretical approaches. Predicting whether a compound will be toxic or nontoxic is a classification problem and the methods of studying quantitative structure activity relationships (QSAR) can be used for this purpose [Hansch C. (1969) .

Here the term usefull gives two important meanings. as it can give gud information about chemical properties as wll as it can be the part of prediction model of some different molecules. molecular descriptors contains constitutional descriptor, chi connectivity indices, Topological Descriptor, Molecular Fragment, 2-D Molecular Properties, etc.

3b. Invivo Testing

The term in vitro, in contrast to in vivo, refers to a medical study or experiment which is done in the laboratory within the confines of a test tube or laboratory dish. Improvement over animal testing Most toxicologists believe that in vitro toxicity testing methods can be more useful, more time and cost-effective than toxicology studies in living animals (which are termed in vivo or "in life" methods). However, the extrapolation from in vitro to in vivo requires some careful consideration and is an active research area.

3b. Invitro Testing

The term in vivo refers to a medical test, experiment or procedure that is done on (or in) a living organism, such as a laboratory animal or human.

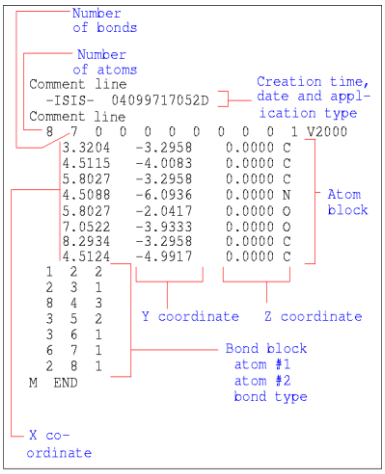
Both in vitro and in vivo methods can be used to predict the inherent hazard properties of chemical substances. However, results obtained from in vitro studies cannot often be used directly to predict biological responses of organisms to chemical exposure in vivo

3b. Insilico Testing

IN SILLICO methods, meaning "performed on computer or via computer simulation." This term was developed as an analogy to the Latin phrases in vivo and in vitro.

4. DATASET

The dataset we use will be in the .mol format (ie) molecule format. the information and values of like number of bonds, number of atoms, creation time and application type, atom block, x,y,z co-ordinates,etc will be in the plain text



This figure shows a connection table for methyl-cyanoacrylate (see 2.3) Additional redundant flags to lines of the *atom* and *bond blocks* have been omitted for clarity (see 2.5 for the full version).

5. 5. ENDPOINTS

Endpoints are the target in which the model to be built. For toxicity prediction we have certain endpoints such as

- 96-hours LC50 of fathead minnow (LC50 = lethal concentration at which 50% of population is killed in 96-hours exposure)
- 48-hour LC50 of daphnia magna ((LC50 = lethal concentration at which 50% daphnia magna is killed in 48-hours exposure
- IGC50 of Tetrahymena pyriformis (the concentration of substance that inhibits 50% of the growth)
- Oral rat LD50 (lethal dose which kills 50% of rats tested)
- Bioconcentration Factor (BCF)(ratio of the concentration of a chemical in an organism to the concentration of the chemical in the surrounding environment.)
- Developmental Toxicity (DevTox)
- Bacterial Reverse Mutation Assay (AMES Test (Mutagenicity)

6. MACHINE LEARNING TECHNIQUES

Hierarchical method

Hierarchical clustering is one of themachine learning algorithms which gives the average of multiple predictions from the various clusters. Every single model is being abstracted by using wards method by dividing thetraining set into group of structurally similar models. genetic algorithm is used to generate different models for clusters.

FDA method

This FDA method is used for predicting the test chemical by using the new model which fits to the chemical that is

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

mostly similar to the test chemical and the models are generated using run time.

Single model method

Single model model creates predctions with the help of multilinear regression which fits into training set by using chemical descriptors as independent variabales. it uses the approach of genetic algorithm and this model is developed before runtime.

Nearest Neighbor method

The toxicity prediction is done by calculating average of 3 chemicals from the training set which resembles same to the test compound

Consensus method:

Concensus model is the most accurate model till now as it is calculated by measuring the average of all the predicted toxicity values from all the QSAR methods

Random forest method:

This method is calculated under the concept of decision tree which converts the compounds into som toxicity score with the help of molecular descriptors which is set as decision variables. this method is only used for developmental toxicity endpoint

7. TOX RUNS USING MACHINE LEARNING TECHNIQUES & RESULTS

1. Molecule Name: Benzene

CAS: 71-43-2
End Point: Oral rat LD50

1a. Predicted Oral rat LD50 for 100-46-9 from Consensus method

This table consist of endpoint of ORAL RAT with the real time experimented laboratory value and model predicted value of Concensus method. The prediction interval is retrieved by summing and detucting the uncertainty from the predicted toxicity

Prediction results		
Endpoint	**Experimental value**	**Predicted value**
Oral rat LD_{50} - Log10(mol/kg)	N/A	2.39
Oral rat LD_{50} mg/kg	N/A	436.89

1b. Predicted Oral rat LD50 for 71-43-2 from Hierarchical clustering method.

Endpoint	**Experimental value (CAS= 71-43-2)**	**Predicted value[a]**	**Prediction interval**
Oral rat LD_{50} - Log10(mol/kg)	1.92	1.98	$1.47 \leq$ Tox ≤ 2.49
Oral rat LD_{50} mg/kg	930.60	820.45	$254.66 \leq$ Tox $\leq$ 2643.29

This table consist of endpoint of ORAL RAT with the real time experimented laboratory value and model predicted value of Heirarchical clustering method. The prediction interval is retrieved by summing and detucting the uncertainty from the predicted toxicity

Descriptors for 71-43-2 for cluster model#11653

The table consist of different molecular descriptors with its value and coefficients from single cluster. By adding the given values we get the predicted value.

Descriptor Values			
Descriptor	**Value**	**Coefficient**	**Value × Coefficient**
xch7	0.0000	9.2032	0.00
SdsCH_acnt	0.0000	0.0874	0.00
Qsv	0.9375	-1.2843	-1.20
BELe6	0.0000	0.6718	0.00
MAXDP	0.0000	-0.6089	0.00
MATS7m	0.0000	-0.3455	0.00
Model intercept	1.0000	3.18	3.18
Predicted value - Log10(mol/kg)			1.98

1c. Predicted Oral rat LD50 for 71-43-2 from FDA method

This table consist of endpoint of ORAL RAT with the real time experimented laboratory value and model predicted value from FDA method.

Prediction results			
Endpoint	**Experimental value (CAS= 71-43-2)**	**Predicted value[a]**	**Prediction interval**
Oral rat LD_{50} - Log10(mol/kg)	1.92	1.70	$1.12 \leq$ Tox ≤ 2.28
Oral rat LD_{50} mg/kg	930.60	1563.55	$411.10 \leq$ Tox $\leq$ 5946.76

Descriptors for 71-43-2 for FDA model

The table consist of different molecular descriptors with its value and coefficients from single cluster. By adding the given values we get the predicted value.

Descriptor Values			
Descriptor	**Value**	**Coefficient**	**Value × Coefficient**
SdCH2	0.0000	-0.0567	0.00
SsNH2	0.0000	0.2217	0.00
nH	6.0000	-0.1228	-0.74
nN	0.0000	0.5327	0.00
MWC10	8.7234	0.2846	2.48

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

CID2	1.9688	-2.0631	-4.06
-CH< [aliphatic attach]	0.0000	-0.4251	0.00
-CH< [aromatic attach]	0.0000	0.4814	0.00
=CH [aliphatic attach]	0.0000	0.2639	0.00
Model intercept	1.0000	4.01	4.01
Predicted value - Log10(mol/kg)			1.70

1d. Predicted Oral rat LD50 for 71-43-2 for Nearest neighbor method

This table consist of endpoint of ORAL RAT with the real time experimented laboratory value and model predicted value from Nearest Neighbour method.

Prediction results		
Endpoint	**Experimental value (CAS= 71-43-2)**	**Predicted value[a]**
Oral rat LD_{50} - Log10(mol/kg)	1.92	1.08
Oral rat LD_{50} mg/kg	930.60	6522.72

2. Molecule Name: Benzylamine

CAS: 100-46-9
End Point: Oral rat LD50

2a. Predicted Oral rat LD50 for 100-46-9 from Consensus method

This table consist of endpoint of ORAL RAT with the real time experimented laboratory value and model predicted value of Concensus method.

Prediction results		
Endpoint	**Experimental value**	**Predicted value**
Oral rat LD_{50} - Log10(mol/kg)	N/A	2.39
Oral rat LD_{50} mg/kg	N/A	436.89

2b. Predicted Oral rat LD50 for 100-46-9 from Hierarchical clustering method

This table consist of endpoint of ORAL RAT with the real time experimented laboratory value and model predicted value of Heirarchical clustering method. The prediction interval is obtained by adding and subtracting the uncertainty from the predicted toxicity

Prediction results			
Endpoint	**Experimental value**	**Predicted value**	**Prediction interval**
Oral rat LD_{50} - Log10(mol/kg)	N/A	2.20	$1.75 \leq$ Tox ≤ 2.65
Oral rat LD_{50} mg/kg	N/A	680.32	$241.20 \leq$ Tox ≤ 1918.86

Descriptors for 100-46-9 for cluster model#11454

The table consist of different molecular descriptors with its value and coefficients from single cluster. By adding the given values we get the predicted value.

Descriptor Values			
Descriptor	**Value**	**Coefficient**	**Value × Coefficient**
ic	4.0000	0.0493	0.20
MDEO12	0.0000	0.6077	0.00
BEHe7	1.3868	-0.4741	-0.66
ATS7m	0.0000	1.3283	0.00
MATS5v	-1.0000	-0.1288	0.13
MATS5e	-1.0000	-0.1753	0.18
GATS2v	0.4444	0.4164	0.19
-COOH [aromatic attach]	0.0000	-0.3917	0.00
Model intercept	1.0000	2.17	2.17
Predicted value -Log10(mol/kg)			2.20

2c. Predicted Oral rat LD50 for 100-46-9 from FDA method

This table consist of endpoint of ORAL RAT with the real time experimented laboratory value and model predicted value from FDA method. The prediction interval is obtained by adding and subtracting the uncertainty from the presumed toxicity

Prediction results			
Endpoint	**Experimental value**	**Predicted value**	**Prediction interval**
Oral rat LD_{50} - Log10(mol/kg)	N/A	2.58	$2.14 \leq$ Tox ≤ 3.03
Oral rat LD_{50} mg/kg	N/A	279.89	$99.94 \leq$ Tox ≤ 783.81

Descriptors for 100-46-9 for FDA model

The table consist of different molecular descriptors with its value and coefficients from single cluster. By adding the given values we get the predicted value.

Descriptor Values			
Descriptor	**Value**	**Coefficient**	**Value × Coefficient**
ka1	5.3211	-0.2312	-1.23
ARR	0.7500	0.6063	0.45
nN	1.0000	0.3400	0.34
-CH< [aromatic attach]	0.0000	0.4380	0.00
Model intercept	1.0000	3.02	3.02
Predicted value - Log10(mol/kg)			2.58

2d. Predicted Oral rat LD50 for 100-46-9 for Nearest neighbor method

This table consist of endpoint of ORAL RAT with the real time experimented laboratory value and model predicted value from Nearest Neighbour method. The prediction interval is obtained by adding and subtracting the uncertainty from the predicted toxicity

Prediction results		
Endpoint	**Experimental value**	**Predicted value**
Oral rat LD_{50} - Log10(mol/kg)	N/A	2.39
Oral rat LD_{50} mg/kg	N/A	437.93

8. TABLE INTERPRETATION

The above tox runs is done using two compounds Benzene and Benzylamine. After giving input the CAS number and setting the endpoint the output is obtained. The results has been produced by calculating the value and coefficient of the individual molecular descriptor.

Predicted results

The predicted results table consist of Endpoint, Experimental value, predicted value and prediction interval. The endpoint is the one we set before starting the experiment. Experimental value is the value we get from the laboratories which conducted the usual animal testing. The predicted value is the one we get from the trained model. At the end prediction interval is the limit in which the value can fall in between and it is calculated by summing and detucting the uncertainty from the presumed toxicity.

Descriptor Values

The descriptor value tables consist of descriptor name, value and coefficient and the product if value and coefficient. The descriptor name is the individual property of the single compound such as **SdCH2_acnt** which means Count of (= CH2) from that compound, **(nTB)** which means Number of triple bonds. The descriptor values were validated using softwares like MDL QSAR, Dragon and Molconn-z. the sum of value x coefficient gives the final output of model intercept.

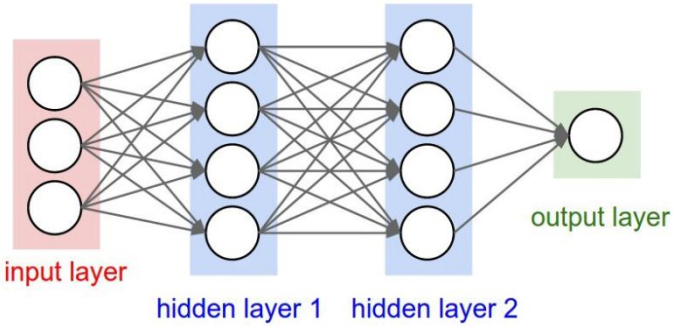
9. PROPOSED METHODOLOGY

9a. Toxicity Prediction based on Deep Learning.

Deep learning is been a novel algorithm and architectures for artificial neural networks as we have super computers with high processing speed to work with complex datasets. As we know deep learning discovers the various levels of distributed representations from the given datasets. from our dataset various molecular descriptors can be learnt from different compounds. moreover, deep learning usually enables the multitask learning that it learns multiple toxicities in single network so that it learns highly informative molecular features.

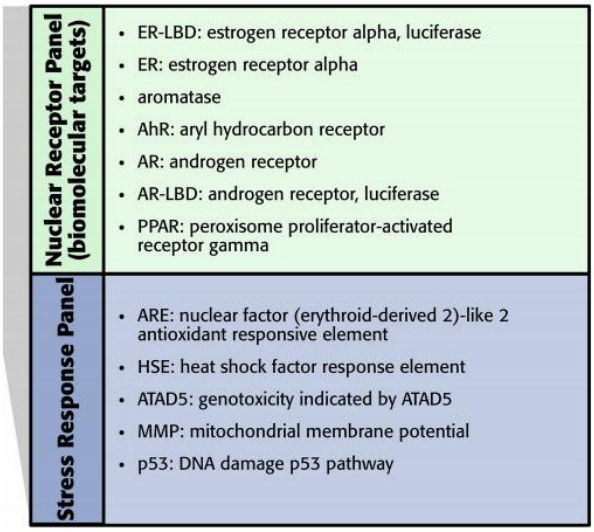
What is Deep Neural Networks

Basically DNN i.e. Deep neural network is the function which maps input vector towards output vector. weights plays major role in parameterizing the mapping and that are balanced and optimised in further learning. Usually the shallow networks will be having only 1 hidden layer and very less number of hidden neurons , but in DNN there will be many hidden layer and many number of neurons, DNN can have millions of neurons each layers but the major criteria is to capture all information from the given input.



Lets us see about neurons. basically neurons act as a abstracting feature with appropriate activation value that indicates the presence of that feature. basically the neuron activation is computed from the below layer neuron activation as a neuron is constructed from its previous layer. the 1st layer is called as input layer and last layer is called as output layer. the layers which is in-between called as hidden layers.

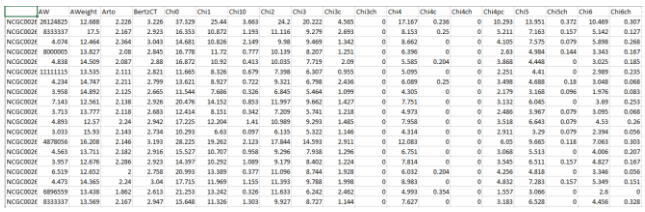
10. DATA SET END POINTS USED FOR DEEP LEARNING



11. END POINT EXPLANATION

- ANDROGEN RECEPTOR – Prostrate cancer and Androgen related disease
- ESTROGEN RECEPTOR – Disruption in normal endrocryne functions
- ARYL HYDROCARBON RECEPTOR – It is very crucial for adaptive response to environmental changes
- PEROXISOME PROLIFERATOR-ACTIVATED RECEPTOR GAMMA – It involves in regulation of glucose and lipid metabolism
- ANTI OXIDANT RESPONSE– It plays important role in amelioration of oxidative stress
- ATAD5 – It is used to identify the compounds that cause genetic stress
- HSE – Activation of heat shock response
- MMP – mitochondrial toxicity and mitochondria related disease
- P53 – Tumour suppressor protein is activated following cellular insult, DNA damage and cellular stress

12. DATASET



	AW	AWeight	Arto	BertzCT	Chi0	Chi1	Chi10	Chi2	Chi3	Chi3c	Chi3ch	Chi4	Chi4c	Chi4ch	Chi4pc	Chi5	Chi5ch	Chi6	Chi6ch
NCGC002(	39124825	12.688	2.226	3.226	37.329	25.44	3.863	24.2	20.222	4.565	0	17.167	0.236	0	10.293	13.951	0.372	10.409	0.307
NCGC002(	8333337	17.5	2.167	2.923	16.353	10.872	1.193	11.116	9.279	2.693	0	8.153	0.25	0	5.211	7.163	0.157	5.142	0.127
NCGC002(	4.074	12.464	2.364	3.043	14.681	10.826	2.149	9.98	9.469	1.342	0	8.662	0	0	4.105	7.575	0.079	5.898	0.268
NCGC002(	8000005	13.827	2.08	2.845	16.778	11.72	0.777	10.139	8.207	1.251	0	6.396	0	0	2.63	4.984	0.144	3.343	0.167
NCGC002(	4.838	14.509	2.087	2.88	16.872	10.92	0.413	10.035	7.719	2.09	0	5.585	0.204	0	3.868	4.448	0	3.025	0.185
NCGC002(	11111115	13.535	2.111	2.821	11.665	8.326	0.679	7.398	6.307	0.955	0	5.095	0	0	2.251	4.41	0	2.989	0.235
NCGC002(	4.234	14.747	2.211	2.799	13.621	8.927	0.722	9.321	6.798	2.436	0	6.089	0.25	0	3.498	4.688	0.18	3.048	0.068
NCGC002(	3.958	14.892	2.125	2.665	11.544	7.686	0.326	6.845	5.464	1.099	0	4.305	0	0	2.179	3.168	0.096	1.576	0.083
NCGC002(	7.143	12.561	2.138	2.926	20.476	14.152	0.853	11.997	9.662	1.427	0	7.751	0	0	3.132	6.045	0	3.69	0.253
NCGC002(	3.713	13.777	2.118	2.683	12.414	8.151	0.342	7.209	5.741	1.218	0	4.973	0	0	2.486	3.967	0.079	3.095	0.068
NCGC002(	4.893	12.57	2.24	2.942	17.225	12.204	1.41	10.989	9.293	1.485	0	7.958	0	0	3.516	6.643	0.079	4.53	0.26
NCGC002(	3.033	15.93	2.143	2.734	10.293	6.63	0.097	6.135	5.322	1.146	0	4.314	0	0	2.911	3.29	0.079	2.394	0.056
NCGC002(	4878056	16.208	2.146	3.193	28.225	19.262	2.123	17.844	14.593	2.911	0	12.083	0	0	6.05	9.665	0.118	7.063	0.303
NCGC002(	4.563	13.711	2.182	2.916	15.527	10.707	0.958	9.296	7.938	1.296	0	6.751	0	0	3.068	5.513	0	4.006	0.207
NCGC002(	3.957	12.676	2.286	2.923	14.397	10.292	1.089	9.179	8.402	1.224	0	7.814	0	0	3.545	6.511	0.157	4.827	0.167
NCGC002(	6.519	12.652	2	2.758	20.993	13.389	0.377	11.096	8.744	1.928	0	6.032	0.204	0	4.256	4.818	0	3.346	0.056
NCGC002(	4.473	14.365	2.24	3.04	17.715	11.969	1.155	11.393	9.788	1.998	0	8.983	0	0	4.832	7.283	0.157	5.349	0.151
NCGC002(	6896559	13.438	1.862	2.613	21.253	13.242	0.326	11.633	8.242	2.462	0	4.993	0.354	0	1.557	3.066	0	2.6	0
NCGC002(	8333337	13.569	2.167	2.947	15.648	11.126	1.303	9.927	8.727	1.144	0	7.627	0	0	3.183	6.528	0	4.456	0.328

13. PREPROCESSING THE DATASETS

This training set contains redundant compounds that appear multiple times within the data, but each time accompanied by carrier molecules such as water, salts or other solubles. Also, we observed compounds that actually consisted of two unrelated structures, but which for some unknown reason where encoded together. We semi automatically labeled these fragments, cleaning up contradictory and combining agreeing compounds.

This way we identified 8,695 distinct compound fragments. To further clean up the data, we made ran a standard clean-up routine for chemical compounds on the data using ChemAxon. This made all hydrogen atoms explicit, ensured that aromatic bonds and tautomers where coded consistently and unified the encoding of salts. We then calculated the input features

14. THRESHOLD TOXIC VALUES

Hazard	Value range (µg/kg bodyweight/day) if not given otherwise; values given per person were divided by 60 kg
General toxicity (genotoxic substances)	0.0025 - 2 (depending on duration)
General toxicity (non-genotoxic substances)	0.025 - 1 (depending on Cramer classes)
General toxicity organophosphates including carbamates, organohalogens and remaining Cramer class III substances	0.30 - 4
Repeat dose toxicity (oral)	0.63 - 60 (depending on Cramer classes and duration, OECD TG)
Repeat dose toxicity (inhalation)	0.07 - 23 (depending on Cramer classes and duration, OECD TG)
Genotoxicity	0.025 - 2
Carcinogenicity (genotoxic)	0.0025
Carcinogenicity (non-genotoxic)	0.025 - 0.75 (depending on Ames test and acute toxicity)
Acute toxicity (inhalation)	4 - 1,000 µg/m³
Neurotoxicity	0.3
Developmental toxicity	1[a] - 8 - 131 (depending on Cramer class) 0.5 - 1 µg/m³ (inhalation)
Reproductive toxicity	1 - 100
Estrogenic endocrine disruption	0.025
Immunotoxicity	0.15 - 1,000
Skin sensitization (dermal)	0.91 - 900 µg/cm²

15. VALIDATION

Q Squared and R Squared validation

$$R^2 = \left[\frac{N\Sigma XY - (\Sigma x)(\Sigma Y)}{\sqrt{([N\Sigma X^2 - (\Sigma X)^2][N\Sigma Y^2 - (\Sigma Y)^2])}} \right]^2$$

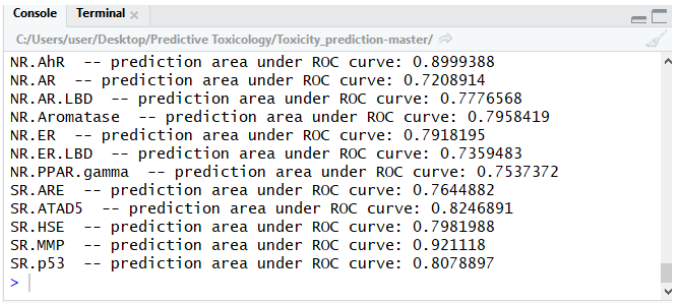
Here the X and Y values are taken from molecular descriptors.

$$PRESS = \sum_{i=1}^{N} (y_{pred,i} - y_i)^2$$

$$Q^2 = 1 - \frac{PRESS}{\sum_{i=1}^{N} (y_i - y_m)^2}$$

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

16. RESULT



17. CONCLUSION

These results has been brought using the methodology of QSAR combined with machine learning. In those previous tox tuns some of the machine learning algorithms used such as Hierarchical method, FDA method, Single model method, Group contribution method, Nearest neighbour method, Consensus method, Random forest method. Among these consensus method is having the highest accuracy. In those values are validated using r squared and q squared methods , but our aim is to implement this toxicity prediction in Deep learning.

18. REFERENCES

1. Ajmani, S., Jadhav, K., and Kulkarni, S. A. (2006). Three-dimensional QSAR using the k-nearest neighbor method and its interpretation. J. Chem. Inf. Model. 46, 24–31. doi: 10.1021/ci0501286
2. Andersen, M. E., and Krewski, D. (2009). Toxicity testing in the 21st century: bringing the vision to life. Toxicol. Sci. 107, 324–330. doi: 10.1093/toxsci/kfn255
3. Baldi, P., Sadowski, P., and Whiteson, D. (2014). Searching for exotic particles in high-energy physics with deep learning. Nat. Commun. 5:4308. doi: 10.1038/ ncomms5308
4. Bartkova, J., Hoˇrejší, Z., Koed, K., Krämer, A., Tort, F., Zieger, K., et al. (2005). DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis. Nature 434, 864–870. doi: 10.1038/nature 03482
5. Bender, A., Mussa, H., Glen, R. C., and Reiling, S. (2004). Molecular similarity searching using atom environments, information-based feature selection, and a naive Bayesian classifier. J. Chem. Inf. Comput. Sci. 44, 170–178. doi:10.1021/ci034207y
6. Chawla, A., Repa, J. J., Evans, R. M., and Mangelsdorf, D. J. (2001). Nuclear receptors and lipid physiology: opening the X-files. Science 294, 1866–1870. doi: 10.1126/science.294.5548.1866
7. Cire¸san, D. C., Meier, U., and Schmidhuber, J. (2012a). "Multi-column deep neural networks for image classification," in Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (Providence, RI), 3642–3649. doi: 10.1109/CVPR.2012.6248110
8. Cire¸san, D. C., Giusti, A., Gambardella, L. M., and Schmidhuber, J. (2013). "Mitosis detection in breast cancer histology images with deep neural networks," in 16th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2013), eds K. Mori, I. Sakuma, Y.Sato, C. Barillot, and N. Navab (Nagoya), 411–418. doi: 10.1007/978-3-642-40763-5_51
9. Cire¸san, D. C., Meier, U., Gambardella, L. M., and Schmidhuber, J. (2012b). "Deep big multilayer perceptrons for digit recognition," in Neural Networks: Tricks of the Trade, eds G. Montavon, G. B. Orr, and K.-R. Müller (Heidelberg: Springer), 581–598.
10. Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. J. Stat. Softw. 33, 1–22. doi: 10.18637/jss.v033.i01
11. Glorot, X., Bordes, A., and Bengio, Y. (2011). "Deep sparse rectifier neural networks," in Fourteenth International Conference on Artificial Intelligence and Statistics (AISTATS 2011), eds G. J. Gordon, D. B. Dunson, and M. Dudík (Fort Lauderdale, FL), 315–323.
12. Graves, A., Mohamed, A. R., and Hinton, G. E. (2013). "Speech recognition with deep recurrent neural networks," in Proceedings of the 2013 IEEE International Conference on

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication