**Journal of Computational Methods in Molecular Design, 2015, 5 (4):24-32**

**Scholars Research Library**
*(http://scholarsresearchlibrary.com/archive.html)*

**ISSN : 2231- 3176**
**CODEN (USA): JCMMDA**

# QSAR based analysis of fatal drug induced renal toxicity

**Vasudha Satalkar[1], Sudhir Kulkarni[2] and Dattatraya Joshi[1]**

*[1]Dept. of Biotechnology, MGM Institute of Health Sciences, Kamothe, Navi Mumbai, India*
*[2]Nova Lead Pharma Pvt. Ltd, Baner Road, Pune, India*

---

**ABSTRACT**

*This study is aimed at finding Quantitative Structure Activity Relationships (QSAR) for drugs reported to result in fatal consequence due to kidney failure and categorized as Adverse Drug Reactions (ADR). Study is based on the reports from open source Canada Vigilance Adverse Reaction Online database. Biological toxicity of small molecules has been predicted as a function of molecular structural features represented by their molecular descriptors. QSAR methods used have identified the structural features of the drugs/molecules and predicted their toxicity. Drugs suspected to cause kidney failure as ADR were analyzed. The molecular descriptors of these drugs were obtained using DRAGON web interface. The structural characteristics that distinguish drugs reported to cause death due to kidney failure as ADR against drugs not causing death but causing kidney failure as ADR were checked. Three QSAR methods used to find the relationships were Simple Kmeans clustering, decision tree and linear regression analysis. The greater value of the descriptor MAXDP is favorable for preventing death has been illustrated by all three models. The 9-membered ring of the benzimidazole substructure can be inferred from Pubchem database to contribute positively towards death. The descriptor, T(N..P), sum of topological distances between N..P 2D atom pairs, would prevent death if its value is lowered. A decrease in value of the descriptor, PCR - ratio of multiple path count over path count, will result in a decrease in probability of fatal consequences. The study predicts that renal toxicity could be decreased if the above mentioned molecular descriptors are modified.*

**Keywords:** Adverse Drug Reactions, Decision tree, Kidney failure, K-means clustering, linear regression, Quantitative Structure Activity Relationship.

---

## INTRODUCTION

Adverse Drug Reactions (ADR) are a major cause because of which a drug could be withdrawn from market. The ADR related toxicity is known to be one of the top ten causes of death in the US. In addition ADR related toxicity can enhance annual health care cost several fold. [1]. Kidneys are a pair of organs in our body that filters waste materials that are subsequently eliminated in urine. The buildup of waste and fluid in the body is due to kidney failure. Acute kidney failure develops suddenly due to drugs or other reasons. Chronic kidney failure develops gradually over time. In the end stages of kidney failure the patient suffers from symptoms like anemia, high blood pressure, bone disease, heart failure, and poor mental functioning [2]. Mortality due to acute kidney failure is rising at a faster pace compared to mortality due to acute myocardial infarction. [3]. There is a need for understanding why some drugs lead to acute or chronic kidney failure.

Biological response in terms of activity or toxicities of a drug can be defined by its structure. A quantitative structure-activity relationship (QSAR) relates quantitative chemical structure attributes (molecular descriptors) to a

biological response. The chemical as well as biological properties of compounds are related to the structure of the compound. The chemical structures can be mathematically transformed into numerical values that characterize the molecule. [4]. Molecular descriptors are the numerical values that characterize properties of molecules. A lot of descriptors are now available through various sources. A method is essential to pick optimal subset of descriptors from many available sources that can explain the biological response. [5]

Various pattern recognition techniques like linear regression analysis, clustering (simple Kmeans clustering, KNN clustering, hierarchical clustering), decision tree etc. have been used by other researchers in finding relationships between biological response of compounds and their structural characteristics. A review of QSAR models and software for predicting reproductive toxicity has been provided by Piparo & Worth [6]. Fliri *et al*. have used hierarchical clustering to find relationships between percentage inhibition values and molecular structures of compounds. [7] Fliri et al. have also used hierarchical clustering to find relationships between drugs induced side effects and molecular structural characteristics. [8] Hong et al. have used tree based approach to find relationships between structural characteristics of chemicals and their estrogen binding capacity. [9] Rodgers et al. have used KNN clustering method to find QSAR relationships between drugs and liver related ADRs on a large scale. They used MolconnZ descriptors and Dragon descriptors for developing QSAR models and found Dragon descriptors more useful in predicting liver toxicity. [10]

The objective of this study was to find the structural characteristics defined by their molecular descriptors that distinguish drugs (small single drug molecules) reported to cause death due to kidney failure as ADR against drugs not causing death but causing kidney failure as ADR. The QSAR methods used to find the relationships were Simple Kmeans clustering, decision tree and linear regression analysis. Further, an open source Weka machine learning software [11] is used for selecting molecular descriptors and building various models. This study is intended to help drug designer avoid structural features that can cause fatality due to kidney failure.

## MATERIALS AND METHODS

The first task for this study was to identify drugs suspected to cause kidney failure as ADR. For this purpose in open source Canada Vigilance Adverse Reaction Online database [12] was searched. Structures of the drug molecules identified were retrieved from Pubchem database in SMILES format. The molecular descriptors for these drugs were calculated using DRAGON [13] web interface. Appropriate descriptors were selected using various attribute selection algorithms provided by Weka machine learning software and literature studies. Finally, three different models were created using the Weka machine learning software to identify the structural features of the drugs that may be responsible for fatal consequences. A graphical flow chart of the study is shown in Figure 1 below.

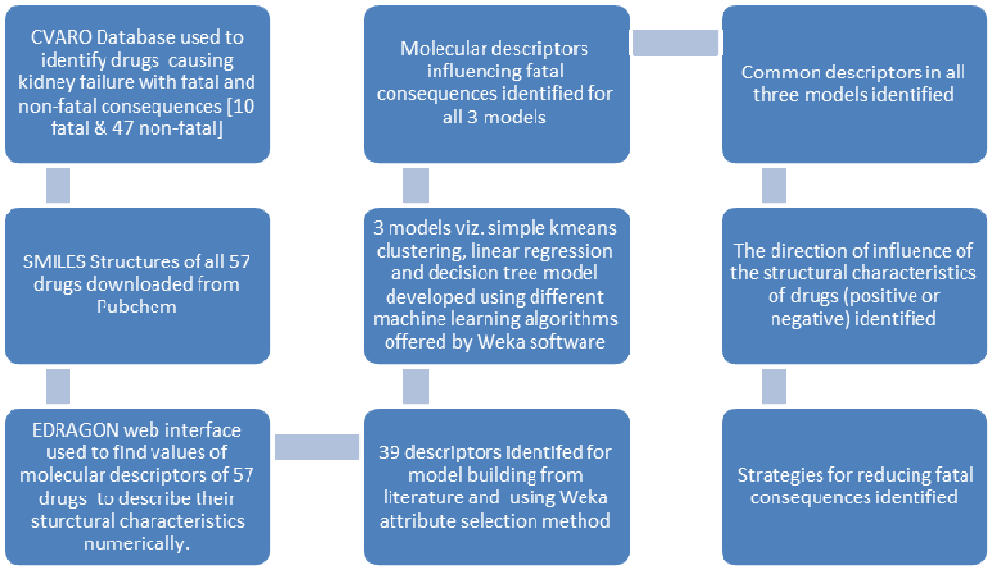


**Figure 1: Graphical flow chart of the study**

**2.1. Data collection and data cleaning**

The open source Canada Vigilance Adverse Reaction Online database was searched for all drugs using ADR terms "Renal Failure Acute (RFA)" and "Renal failure Chronic (RFC)" from 01-01-1965 to 30-09-2012 in chunks of 10 years.

All the records downloaded from Canada Vigilance Adverse Reaction Online Database and were imported into MySQL for cleaning purpose. Two separate tables were created for Acute Renal Failure and Chronic Renal Failure. The raw data consisted of a total of 55574 records for RFA and 8364 records for RFC. Only records reported as suspects for an adverse event Renal Failure Acute & Renal Failure Chronic were selected. Duplicate records were removed. Total of 1573 records of RFA suspects and 358 records of RFC suspects were found after removal of duplicates.

**2.1.1. *Drug Selection Criteria***

Criteria used for drug selection were

1. Since this study is based on structural characteristics, single drug molecules with known chemical structures in SMILES format in Pubchem database were included.

2. Biosimilars/biologics such as vaccines, monoclonal antibodies etc. are excluded as they lack precise SMILES structures in Pubchem.

**2.2. Data analysis**

A total of 577 drugs were reported to cause Renal Failure Acute and 147 drugs were found to cause Renal Failure Chronic. There were 276 incidences of drug induced death due to Renal Failure Acute (RFA) and 55 incidences of death due to Renal Failure Chronic (RFC) reported in the database. 57 drugs were common in Renal Failure Acute & Renal Failure Chronic database. Of the 57 drugs suspected to cause both Acute Renal Failure as well as Chronic Renal Failure, the drugs (small molecules) where death has been reported as patient outcome were found to be Azathioprine, Clozaril/Clozapine, Diclofenac Sodium, Diflucan/Fluconazole, Furosemide, Indomethacin, Metformin, Micardis/Telmisartan, Viread/Tenofovir, and Zyprexa/Olanzapine.

The final dataset used for this study consists of 57 drug molecules, 47 of which are not suspected to cause death as patient consequence due to kidney failure as ADR and 10 are suspected to cause death.

Simplified Molecular Input Line Entry Specification (SMILES) of structures of the 57 drugs were used to obtain molecular descriptor values through DRAGON web interface. All 266 possible topological descriptors that are implemented in the DRAGON (Milano Chemometrics and QSAR Research Group, Bicocca, Italy) software [13] were calculated for the 57 drugs.

**2.2. Development of Quantitative Structure Activity Relationship Models**

The biological consequence we are trying to study is death as a mathematical function of the numerical values describing structure of the drug molecules suspected to cause kidney failure. Three QSAR models were developed using simple k-means clustering, linear regression and decision tree analysis methods.

**2.2.1. *Selection of Molecular descriptors***

Of the 266 molecular descriptors downloaded from DRAGON interface [13] attributes with zero variance were deleted at the outset. 20 descriptors were selected from literature regarding descriptors with known toxicities from compound containing Nitrogen atoms as all drugs in our study found to cause death due to renal failure have nitrogen atom [14]. 18 attributes were selected using Cfssubseteval + BestFit attribute selection method offered by Weka open source data mining software. A total of 39 attributes (including death added from data analysis done earlier) were used find Quantitative Structure Activity Relationship (QSAR) of drugs suspected to cause death as consequence due to kidney failure.

***2.2.2. Simple K-means Clustering Model Development***

The 38 molecular descriptors selected previously were subjected to Principal Component Analysis which resulted in a set of 10 molecular descriptors SOK, X5Av, Mp, AAC, Lop, T(N..P), nR09, nR04, nR11, MAXDP. The attribute death was added to the descriptor set as reported in Canada Vigilance Adverse Reaction Online database with death as patient consequence due to kidney failure as suspected ADR. The attribute death was converted into binary

format where death was represented as 1 for death as patient consequence otherwise 0 if no death was reported. The set of 11 descriptors were imported into Weka machine learning software and Simple Kmeans clustering algorithm was chosen for model creation.

***2.2.3. Linear Regression Model Development***

For building this model the 39 descriptors selected earlier were subjected to genetic search method provided by Weka for better attribute selection. This method yielded 17 molecular descriptors viz. MSD, PHI, Mv, ARR, nCL, nR04, nR09, nR11, MAXDP, BLI, T(N..P), T (F...F), T (Cl..Cl), PCR, X1Av, X3Av, Death. These were exported to Weka machine learning software for creating linear regression model.

Since we had 57 compounds to start with, we created training and test set for internal validation. This was done by randomly selecting 8 compounds from death =0 group and selecting 2 compounds from death = 1 group. Linear regression model was created again using remaining 47 compounds and checked how the generated equation predicts death parameter of 10 compounds in test set that were not used to build the regression equation.

***2.2.4 Decision Tree model Development***

All the 38 attributes selected earlier were imported into Weka. The attribute death added earlier was converted into nominal attribute: y =yes and n=no. The classifier used was weka.classifiers.trees.J48 -C 0.25 -M 2. A decision tree is a simple data mining algorithm that creates a collection of "if → then" conditional rules for assignment of class labels to instances of a data set. The decision trees are represented by nodes that specify a particular attribute of the data and branches that represent a test the value of each attribute, and leaves that correspond to the terminal decision of class assignment for an instance in the data set. [15]

## RESULTS AND DISCUSSION

All the models developed herein have identified eighteen molecular descriptors that can be attributed to renal toxicity with fatal consequences. Some molecular descriptors are common in all models and some are different. This difference can be attributed to the fact that the compounds in our dataset are not a set of homologous series. Table 1 lists the eighteen molecular descriptors found by all three models along with their definition.

**Table 1: List of important descriptors found along with their category and definition**

| Descriptor | Definition | Category |
|---|---|---|
| Mp | mean atomic polarizability | Constitutional indices |
| nR04 | number of 4-membered rings | Ring descriptors |
| nR09 | number of 9-membered rings | Ring descriptors |
| nR11 | number of 11-membered rings | Ring descriptors |
| MAXDP | maximal electrotopological positive variation | Topological indices |
| Lop | Lopping centric index | Topological indices |
| T(N..P) | sum of topological distances between N..P | 2D Atom Pairs |
| X4Av | average valence connectivity index of order 4 | Connectivity indices |
| AAC | mean information index on atomic composition | Information indices |
| S0K | Kier symmetry index | Topological indices |
| MSD | Mean square distance index (Balaban) | Topological indices |
| Mv | mean atomic van der Waals volume | Constitutional indices |
| nCL | number of Chlorine atoms | Constitutional indices |
| T(F..F) | sum of topological distances between F..F | 2D Atom Pairs |
| X1Av | average valence connectivity index of order 1 | Connectivity indices |
| PCR | ratio of multiple path count over path count | Walk and path counts |
| ISIZ | Information index on molecular size | Information indices |
| IAC | Total information index of atomic composition | Information indices |

***3.1. Simple K-means Clustering Model***

This model resulted in two clusters 0 and 1, showing the centroid of each cluster as well as statistics on the number and percentage of instances assigned to different clusters shown in Table 2. Cluster centroids are the mean vectors for each cluster (so, each dimension value in the centroid represents the mean value for that dimension in the cluster). Thus, centroids can be used to characterize the clusters.

**Table 2: Cluster centroids**

| Attribute | Full Data | Cluster #0 | Cluster #1 |
|---|---|---|---|
| No of Molecules | (57) | (47) | (10) |
| S0K | 139.5594 | 149.0958 | 94.7381 |
| Lop | 1.1011 | 1.1216 | 1.0051 |
| Mp | 0.6523 | 0.6464 | 0.68 |
| nR04 | 0.0351 | 0.0426 | 0 |
| nR09 | 0.2982 | 0.2553 | 0.5 |
| nR11 | 0.1053 | 0.0638 | 0.3 |
| MAXDP | 5.3026 | 5.544 | 4.1682 |
| T(N..P) | 0.7895 | 0.1702 | 3.7 |
| X5Av | 0.0323 | 0.0334 | 0.0267 |
| AAC | 1.6326 | 1.6131 | 1.7243 |
| Death | 0.1754 | 0 | 1 |

The predicted clustered instances have accurately detected not only the number of drugs in the dataset with death as a consequence but also the exact drugs reported to cause death as patient consequence shown in Table 3 below.

**Table 3: Cluster assignment of the fifty seven drugs in the dataset**

| Drug | Death reported | Cluster Assigned | Drug | Death reported | Cluster Assigned |
|---|---|---|---|---|---|
| AMPICILLIN | 0 | cluster0 | LIPITOR | 0 | cluster0 |
| METHOTREXATE | 0 | cluster0 | LOPID | 0 | cluster0 |
| AREDIA | 0 | cluster0 | LOSEC | 0 | cluster0 |
| ATACAND | 0 | cluster0 | LOVENOX | 0 | cluster0 |
| AVANDIA | 0 | cluster0 | LYRICA | 0 | cluster0 |
| AVAPRO | 0 | cluster0 | METFORMIN | 1 | cluster1 |
| AZATHIOPRINE | 1 | cluster1 | MICARDIS | 1 | cluster1 |
| CEFTAZIDIME | 0 | cluster0 | MUTAMYCIN | 0 | cluster0 |
| CELEBREX | 0 | cluster0 | NALFON | 0 | cluster0 |
| CIPRO | 0 | cluster0 | PHENACETIN | 0 | cluster0 |
| CLINDAMYCIN | 0 | cluster0 | PREDNISONE | 0 | cluster0 |
| CLOZARIL | 1 | cluster1 | PREPULSID | 0 | cluster0 |
| CRIXIVAN | 0 | cluster0 | PREXIGE | 0 | cluster0 |
| CYCLOMEN | 0 | cluster0 | PRINIVIL | 0 | cluster0 |
| DEXAMETHASONE | 0 | cluster0 | PROZAC | 0 | cluster0 |
| DICLOFENAC | 1 | cluster1 | RAMIPRIL | 0 | cluster0 |
| DIDANOSINE | 0 | cluster0 | RAPAMUNE | 0 | cluster0 |
| DIFLUCAN | 1 | cluster1 | RITONAVIR | 0 | cluster0 |
| DIGOXIN | 0 | cluster0 | SANDOSTATIN | 0 | cluster0 |
| DIOVAN | 0 | cluster0 | SAQUINAVIR | 0 | cluster0 |
| FELDENE | 0 | cluster0 | SEROQUEL | 0 | cluster0 |
| FLUOROURACIL | 0 | cluster0 | TACROLIMUS | 0 | cluster0 |
| FUROSEMIDE | 1 | cluster1 | TETRACYCLINE | 0 | cluster0 |
| GENTAMICIN | 0 | cluster0 | VALIUM | 0 | cluster0 |
| HYDROCHLOROTHIAZIDE | 0 | cluster0 | VASOTEC | 0 | cluster0 |
| IBUPROFEN | 0 | cluster0 | VIREAD | 1 | cluster1 |
| INDOMETHACIN | 1 | cluster1 | WARFARIN | 0 | cluster0 |
| ISOPTIN | 0 | cluster0 | ZYPREXA | 1 | cluster1 |
| LAMIVUDINE | 0 | cluster0 | | | |

The most significant descriptors found from this analysis are SOK Kier symmetry index, MAXDP Maximal electrotopological positive variation, nR09 number of 9- membered rings and T (N..P) Sum of topological distances between N..P 2D Atom Pairs. The prediction accuracy and its interpretation are shown in confusion matrix in Table 4.

***3.2. Linear Regression Model***

The model developed is as follows:

Death = -2.9788 * MSD + -3.2052 * Mv + 0.4048 * nCL + 0.1173 * nR09 + -0.1674 * MAXDP + 0.017 * T(N..P) + 0.2521 * T(F..F) + 0.7898 * PCR + -2.3532 * X1Av + 3.3352 …….[1]

Correlation coefficient 0.7866

Afterwards we checked if descriptors appearing in regression model i.e. MSD, Mv, nCL, nR09, MAXDP, T(N..P), T(F..F), PCR, X1Av are cross correlated. Two pairs that showed bit high correlation were MSD-MAXDP and Mv-PCR. So after dropping systematically each of the above four descriptor from regression equation $R^2$ was calculated again from remaining descriptors.

DROPPING ATTRIBUTES Mv AND PCR FROM ORIGINAL EQUATION [1] FOR TRAINING SET:

Death = -1.9901 * MSD + 0.349 * nCl + 0.1356 * nR09 + -0.1426 * MAXDP + 0.0208 * T(N..P) + 0.2291 * T(F..F) + -2.274 * X1Av + 2.0059 ……..[2]

Correlation coefficient 0.7159 for training set

The signs of the regression coefficients suggest the direction of influence of explanatory variables in the models. The descriptors MSD, Mv, MAXDP and X1Av contribute negatively to death, whereas nCl, nR09, T(N..P), T(F..F), PCR have shown positive contribution to death. Thus a lower value of descriptors nCl, nR09, T(N..P), T(F..F), PCR and a higher value of descriptors MSD, Mv, MAXDP and X1Av would be favorable to preventing death.

***3.3 Decision Tree Model***

The decision tree model developed by us shown below displays a series of yes/no (Y/N) rules to classify drugs into fatal (y) and non-fatal (n) categories based on most relevant descriptors. Four descriptors PCR, ISIZ, IAC and MAXDP were found relevant for the set of drugs in explaining renal toxicity. The molecular descriptors along with transition rules are shown below in Figure 2.

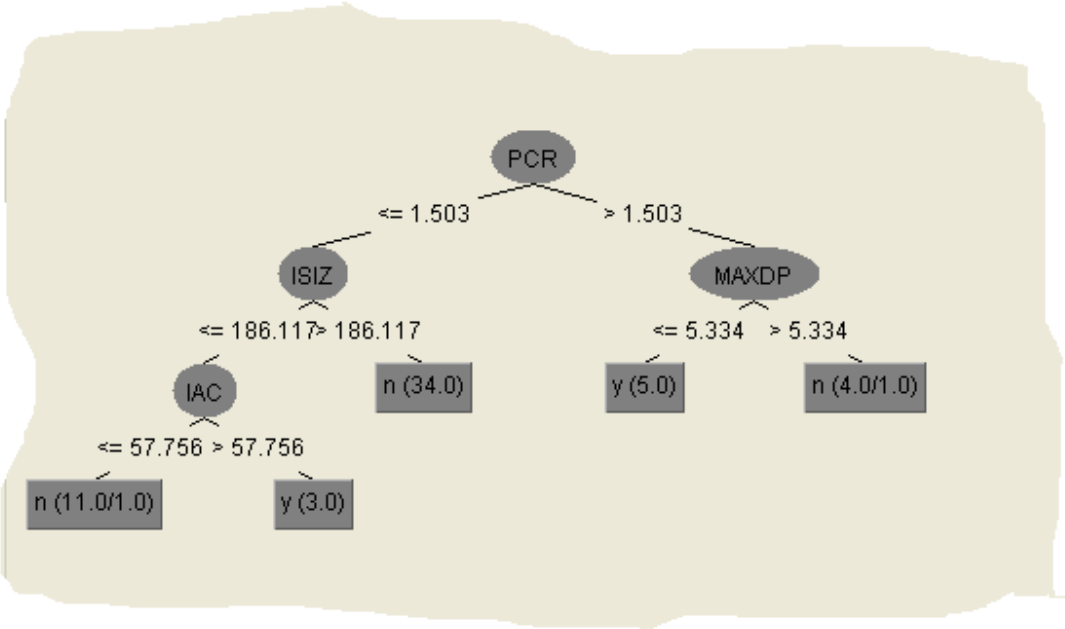


**Figure 2: The decision tree model: The model displays a series of yes/no (Y/N) rules to classify drugs into fatal (Y) and non-fatal (N) categories based on four descriptors: PCR, ISIZ, IAC and MAXDP. The molecular descriptors are denoted by circle. The transition rules are depicted on the branches. And final category assigned is depicted by rectangle with number of instances in dataset represented by the category i.e. death (Y/N)**

Wrongly classified drugs depicted by a slash in the model above are drugs Metformin and Indomethacin classified as no instead of yes. The classification error for the drug metformin can be attributed to it being the only compound

without aromatic ring structure. The important descriptors detected from this model are PCR Ratio of multiple path count over path count, ISIZ Information index on molecular size, IAC Total information index of atomic composition and MAXDP Maximal electrotopological positive variation. The prediction accuracy of the decision tree model and its interpretation is shown in confusion matrix Table 4.

Prediction accuracy of all models was calculated using confusion matrix along with true positive rate, true negative rate, false positive rate and false negative rate and tabulated below.

**Table 4: Comparison of prediction accuracy of all models with their interpretation**

| **Confusion Matrix for Simple Kmeans Clustering model** | | | |
|---|---|---|---|
| | Death Predicted No | | Death Predicted Yes |
| Death Actual No | A = 47 | | B = 0 |
| Death Actual Yes | C = 0 | | D = 10 |
| **Confusion matrix for linear regression equation [2]** | | | |
| | Death Predicted No | | Death Predicted Yes |
| Death Actual No | A = 8 | | B = 0 |
| Death Actual Yes | C = 0 | | D = 2 |
| **Confusion Matrix for the decision tree model** | | | |
| | Death Predicted No | | Death Predicted Yes |
| Death Actual No | A = 47 | | B = 0 |
| Death Actual Yes | C = 2 | | D = 8 |
| **Parameters used to interpret confusion matrix** | | | |
| Parameter | Kmeans Model | Regression Model | Decision Tree Model |
| Accuracy (AC) | $AC^1 = 1$ | AC = 1 | AC = 0.965 |
| True positive rate /sensitivity (TP) | $TP^2 = 1$ | TP = 1 | TP = 0.8 |
| False positive rate (FP) | $FP^3 = 0$ | FP = 0 | FP = 0 |
| True negative rate /specificity (TN) | $TN^4 = 1$ | TN = 1 | TN = 1 |
| False negative rate (FN) | $FN^5 = 0$ | FN = 0 | FN = 0.2 |

*Formula/ Significance: [1]AC = [(A +D)/(A + B + C +D)] indicates the proportion total number of correct predictions, [2] TP = [D/(C +D)] indicates the proportion of correctly identified molecules, [3]FP = [B/(A + B)] indicates the proportion of incorrectly identified molecules, [4] TN = [A/(A + B)] indicates the proportion of correctly identified molecules, [5] FN = [C/(C +D)] indicates the proportion of incorrectly identified molecules.*

The descriptors from constitutional, topological, 2D autocorrelation, informational, connectivity and ring descriptors have together formed models that explain the activity. Ring descriptors and topological indices are the predominant classes for the three models created.

The descriptor Maximal electro-topological positive variation (MAXDP) is shared by all the three models. The greater value of the descriptor MAXDP is favorable for preventing death has been illustrated by all three models.

The descriptor, nR09, number of 9-membered rings is shared by Kmeans & linear regression model. The 9-membered ring of the benzimidazole substructure can be inferred from Pubchem database to contribute positively towards death. Avoiding this substructure in a drug molecule or its metabolites will be helpful in preventing death due to kidney failure.

Similarly the descriptor, T(N..P), sum of topological distances between N..P 2D atom pairs, is shared by both Kmeans & linear regression models would prevent death if its value is lowered.

The descriptor, PCR ratio of multiple path count over path count, is common descriptor for linear regression model and decision tree model. Sign of the regression coefficient associated with PCR is positive therefore a decrease in value of this predictor will result in a decrease in probability of death as patient consequence.

Earlier research on drug induced renal toxicity has revealed many facts relating to causes of kidney failure. Some of the drugs identified by them have also shown up in our research. The drug Micardis/Telmisartan detected from our analysis has been known to inhibit angiotensin-converting enzyme that could injure the kidney by its hemodynamic

effect. The drugs indinavir/Crixivan and methotrexate can cause tubular injury as a result of precipitation/crystallization thus causing obstruction. Immune-mediated interstitial damage in the form of acute interstitial nephritis (AIN) is commonly seen with antibiotics. The antibiotics ampicillin, tetracycline, gentamicin, clindamycin etc. have been detected in our study. The Non-steroidal anti-inflammatory drugs (NSAIDs) Diclofenac & indomethacin that have been detected with fatal consequence have been associated with kidney injury by other researchers as well. [16].

One of the most important functions of the kidney is the filtration and excretion of nitrogenous waste products from the blood [17]. All the drugs causing death due to kidney failure detected from this study have nitrogen atom. The topological descriptor T(N..P) predicted contributes positively to death has been identified from all models that resulted from this study.

Since the various drugs included in our dataset are not structurally homologous, single well-defined SAR cannot be detected. Therefore three different models have been developed and analyzed [18]

## CONCLUSION

The greater value of the descriptor, MAXDP - Maximal electrotopological positive variation, is favorable for preventing death has been underlined by all three models. Avoiding benzimidazole substructure in a drug molecule or its metabolite will be helpful in preventing death due to kidney failure. Similarly the descriptor, T(N..P) - sum of topological distances between N..P 2D atom pairs, would prevent death if its value is decreased. A decrease in value of the descriptor, PCR - ratio of multiple path count over path count will result in decrease in probability of fatal consequences.

The QSAR approach based models used in this study provides interesting insight into drug design with concomitant reduction in toxicity. The models reported herein and inferences are expected to help designers of drugs and ligands for avoiding fatal consequences due to kidney failure.

## CONFLICT OF INTEREST

The authors declare that there is no conflict of interests regarding the publication of this paper. Also, they declare that this paper or part of it has not been published elsewhere. Ms. Vasudha Satalkar initiated the collaborative project as a part of her Ph.D. studies, designed data collection tools, carried out data collection, cleaned and analysed the data, and wrote the initial draft of the paper, under the guidance of Dr. Dattatraya Joshi. Ms. Vasudha Satalkar in consultation with Dr. Sudhir Kulkarni wrote the statistical analysis plan, and refined the model parameters to improve prediction accuracy. Initial draft of the paper prepared by Ms. Vasudha Satalkar, was revised Dr. Sudhir Kulkarni and Dr. Dattatraya Joshi. All authors read and approved the final version.

**Acknowledgements**
Authors wish to thank Dr. Igor Tetko, Institute of Structural Biology, Virtual Computational Chemistry Laboratory, 85764 Neuherberg (bei Munich) for his valuable help and guidance with DRAGON descriptors. Authors also wish to thank Dr. Prasanna Joeg, Head, Computer Engineering Dept. and Mr. Abhijeet Khurpe **of** MIT College of Engineering Pune for allowing the use of their Research Lab & help with software installation.

## REFERENCES

[1] N Anderson & J Borlak. *PloS One*, **2011,** *6*(10), e25221.
[2] CW Hsu & JM Symons. *Pediatric Nephrology (Berlin, Germany)*, **2010**, *25*(12), 2401–12.
[3] LM Stevens. *JAMA*, **2013**, *Vol 301*(No. 6), 686.
[4] BK Sharma; P Singh; P Pilania; M Shekhawat; & YS Prabhakar. *Journal of Enzyme Inhibition and Medicinal Chemistry*, **2012,** *27*(2), 249–60.
[5] S Ajmani; A Agrawal & SA Kulkarni. *Journal of Molecular Graphics & Modelling*, **2010,** *28*(7), 683–94.
[6] EL Piparo & A Worth. *Review of QSAR Models and Software Tools for predicting Developmental and Reproductive Toxicity*. Luxembourg: European Union. **2010** doi:10.2788/9628
[7] AF Fliri; WT Loging; PF Thadeio & RA Volkmann. *Proceedings of the National Academy of Sciences of the United States of America,* **2005** *102*(2), 261–6. doi:10.1073/pnas.0407790101

[8] AF Fliri; WT Loging; PF Thadeio & RA Volkmann. *Nature Chemical Biology*, **2005** *1*(7), 389–397. doi:10.1038/nchembio747
[9] H Hong; W Tong; H Fang; L Shi; Q Xie; J Wu & DM Sheehan. *Environmental Health Perspectives*, **2002,** *110*(1), 29–36.
[10] AD Rodgers; H Zhu; D Fourches & I Rusyn. *Chem Res Toxicol*. **2010,** *23*(4), 724–732. doi:10.1021/tx900451r.Modeling
[11] Weka [software]: Waikato Environment for Knowledge Analysis, ver. 3.6.4, **1999**c, Univ. of Waikato, New Zealand. http://www.cs.waikato.ac.nz/ml/weka/
[12] Canada Vigilance Adverse Reaction Online Database: http://www.hc-sc.gc.ca/dhp mps/medeff/databasdon/index-eng.php#cont (Accessed Nov. 16, **2013**)
[13] EDRAGON: http://www.vcclab.org/lab/edragon/ (Accessed Jan 24, **2014**)
[14] FA Pasha; M Ansari & SK Mishra. *Chemical Biology & Drug Design*, **2009**. *73*, 537–544.
[15] MD Twa; S Parthasarathy & C Roberts. *Optom Vis Sci.*, **2005**. *82*(12), 1038–1046.
[16] N Leung; A Eirin; MV Irazabal; DE Maddox; HD Gunderson; FC Fervenza,; & VD Garovic. *Renal Failure*, **2009** *31*(8), 749–52.
[17] DP Basile, MD Anderson & TA Sutton. *Compr Physiol*, **2012**, *2*(2), 1303–1353
[18] R Guha. *J Comput Aided Mol Des*, **2008,** *22*, 857–871.