

# Towards Accurate Graph Neural Network Reconstruction incorporating Node, Edge and Adjacency matrix: Efficient feature extraction for Toxicity Classification

Tanjeem Azwad Zaman - 1805006  
MD Toki Tahmid - 1805030



# Base Paper (1)

Title:

DeepTox: Toxicity Prediction using Deep Learning

Published at - Frontiers in Environmental Science (IF: 4.6)

Publication Date - February, 2016

Citations - 607

Link:

<https://www.frontiersin.org/articles/10.3389/fenvs.2015.00080/full?fbclid=IwAR1mOno6AMbbQnYlrFWbKkumipHGF5p7qLmmOGfPjzfIcEWKUcm3RzupTM>

Why this paper:

Reputable Journal and large number of citations, hence results are reliable

Easily Reproducible

Open to Modification, Addition, Implementation

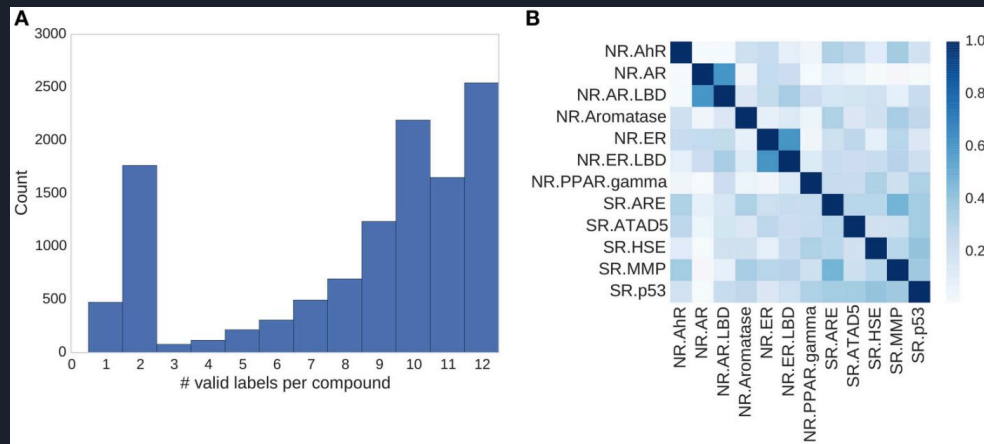
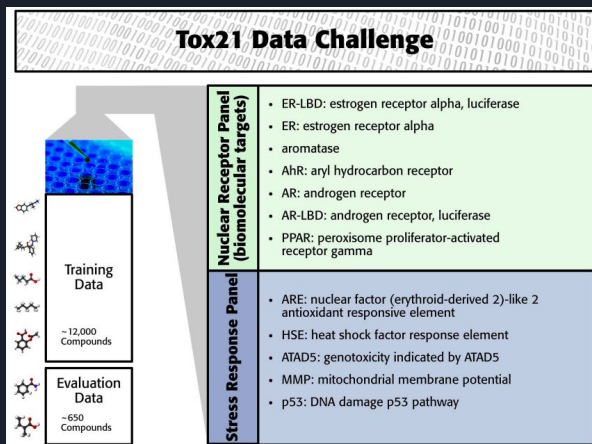
Works solely on Tox-21, a dataset essential for our augmentation

# Base Paper (2)

Problem Definition:

Multi-label binary classification of the Tox-21 Dataset

Dataset:



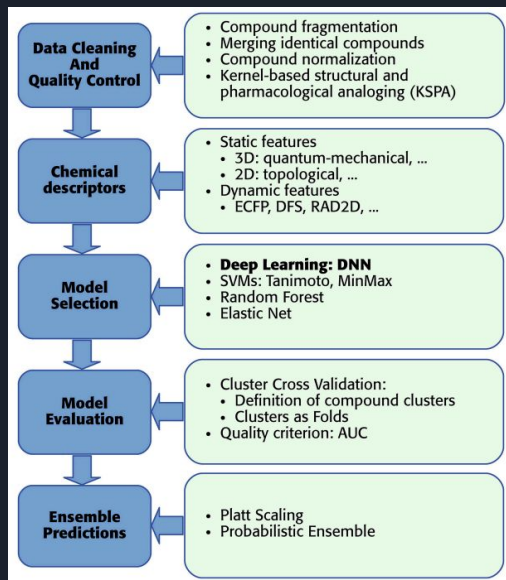
# Base Paper (3)

## Key Ideas:

Extensive Feature extraction for each molecule

Train a DNN and use an ensemble of other well-performing models for final prediction.

## Solution overview:



# Base Paper (4)

## Key Results:

	AVG	NR	SR	AhR	AR	AR-LBD	ARE	Aromatase	ATAD5	ER	ER-LBD	HSE	MMP	p53	PPAR-g
<i>our method</i>	<b>0.846</b>	<b>0.826</b>	<b>0.858</b>	<b>0.928</b>	0.807	<b>0.879</b>	<b>0.840</b>	0.834	0.793	<b>0.810</b>	0.814	<b>0.865</b>	0.942	0.862	<b>0.861</b>
AMAZIZ	0.838	0.816	0.854	0.913	0.770	0.846	0.805	0.819	<b>0.828</b>	0.806	0.806	0.842	<b>0.950</b>	0.843	0.830
dmlab	0.824	0.811	0.850	0.781	<b>0.828</b>	0.819	0.768	<b>0.838</b>	0.800	0.766	0.772	0.855	0.946	<b>0.880</b>	0.831
T	0.823	0.798	0.842	0.913	0.676	0.848	0.801	0.825	0.814	0.784	0.805	0.811	0.937	0.847	0.822
microsomes	0.810	0.785	0.814	0.901	-	-	0.804	-	0.812	0.785	<b>0.827</b>	-	-	0.826	0.717

	AVG	NR	SR	AhR	AR	AR-LBD	ARE	Aromatase	ATAD5	ER	ER-LBD	HSE	MMP	p53	PPAR-g
DNN	0.837	0.827	0.851	0.923	0.778	0.825	0.829	0.804	0.775	0.791	0.811	0.863	0.930	0.860	0.856
SVM	0.832	0.819	0.849	0.919	0.822	0.748	0.818	0.819	0.781	0.799	0.798	0.848	0.946	0.854	0.827
RF	0.820	0.805	0.840	0.917	0.776	0.812	0.810	0.806	0.786	0.770	0.746	0.826	0.945	0.835	0.805
ElNet	0.803	0.787	0.826	0.897	0.788	0.692	0.778	0.763	0.768	0.765	0.805	0.844	0.924	0.818	0.799



# Code and Data Availability

- Code: Unavailable
- Data:
  - [https://moleculenet.org/datasets-1?fbclid=IwAR2Ojx55X5cL7zNzy0tmkrADOX\\_aLNv37JCPHBDGLWnKKgD4\\_bEpwDI9AHs](https://moleculenet.org/datasets-1?fbclid=IwAR2Ojx55X5cL7zNzy0tmkrADOX_aLNv37JCPHBDGLWnKKgD4_bEpwDI9AHs) -> Physiology-> Tox21
  - OR, <https://paperswithcode.com/dataset/tox21-1>
- Other resources:
  - [https://github.com/nierja/tox/tree/master?fbclid=IwAR29P\\_BS5WC3BfTPKKtSsnu1CIDymDghdZz8KDYN3T-iBRtJz5uE1-lUeZw](https://github.com/nierja/tox/tree/master?fbclid=IwAR29P_BS5WC3BfTPKKtSsnu1CIDymDghdZz8KDYN3T-iBRtJz5uE1-lUeZw)

# Proposed Experiment

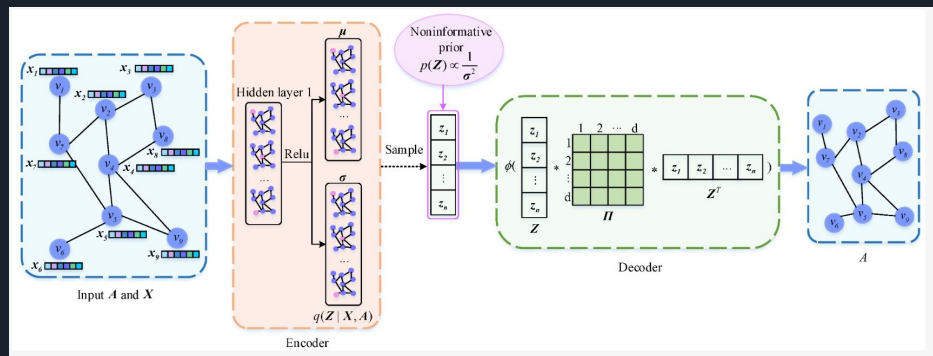
## Feature extraction using a novel multi-modal graph autoencoder

### General Approach:

1. Node feature reconstruction with GCNN and message passing

### Our Approach:

1. Convert main graph to line graph and reconstruct edge features too
2. Incorporate adjacency reconstruction loss in the loss function too
3. Total loss : Node loss + edge loss + adjacency matrix loss



# Classification Task

1. Extract features from the hidden layer of the autoencoder. Thus we use the autoencoder as a feature extractor
2. For classification instead of using simple CNN/ DNN layers, we use attention layers too to understand the intra-feature dependencies between the features of hidden dimension

