Prediction of Human Population Responses to Toxic Compounds by a Collaborative Competition

ISSN: 2311-8636 (Print)
ISSN: 2312-2021 (Online)

DOI prefix: 10.18034

Source of Support: Nil

No Conflict of Interest: Declared

Email for correspondence:
bynagari.gs@gmail.com

Naresh Babu Bynagari

Andriod Developer, Keypixel Software Solutions, 777 Washington rd Parlin NJ 08859, Middlesex, **USA**

ABSTRACT

When it becomes completely possible for one to computationally forecast the impacts of harmful substances on humans, it would be easier to attempt addressing the shortcomings of existing safety testing for chemicals. In this paper, we relay the outcomes of a community-facing DREAM contest to prognosticate the harmful nature of environment-based compounds, considering their likelihood to have disadvantageous health-related effects on the human populace. Our research quantified the cytotoxicity levels in 156 compounds across 884 lymphoblastic lines of cells. For the cell lines, the transcriptional data and genotype are obtainable as components of the initiative known as the Tox21 1000 Genomes Project. In order to accurately determine the interpersonal variations between toxic responses and genomic profiles, algorithms were created by participants in the DREAM challenge. They also used this means to predict the inter individual disparities of cytotoxicity-related data at the populace level from the organizational characteristics of the considered environmental compounds. A sum of 179 predictions were submitted and then evaluated at odds with experiment-derived data set to the blinded lot of participants. The cytotoxicity forecasts performed better than random, showcasing modest interrelations and consistency with a complexity of trait genomic prognostics. Contrastingly, the response of population-level predictions to a variety of compounds proved higher. The outcomes spotlight the likeliness of forecasting health-associated risks with regards to unidentified compounds, despite the reality that one's risk with estimation accuracy persists as suboptimal. Most of the computational means through which chemical toxicity can be predicted are more often than not based on non-mechanistic cheminformatics-inspired solutions. They are typically also reliant on descriptions in QSAR arsenals and usually related with chemical structures rather vaguely. Majority of these computational methods for determining toxicness also employ black box math algorithms. Be that as it may, while such machine learning models might possess much lower capacities for generalization and interpretability, they often achieve high accuracy levels when it comes to predicting a variety of toxicity results. And this is reflected unambiguously by the outcomes of the Tox21 competition. There is a huge capitalization on the ability of present-day Artificial Intelligence (AI) to determine the benchmark data of Tox21 with the aid of a series of 2D-rendered chemical drawings, using no chemical descriptors whatsoever. In particular, we processed some unimportant 2D-based molecules sketches within a controlled convolutional neural 2D network—also represented as 2DConvNet). We also demonstrated that today's image recognition tech culminates in prediction correctness which can be compared to cutting-edge cheminformatics contraptions. Moreover, the 2DConvNet's image-based model was evaluated comparatively dwelling on a set of external compounds from the stables of the Prestwick chemical library. They led to an experimental recognition of substantial and initially undocumented antiandrogen tendencies for diverse drugs in the generic and well-established categories.

Key Words: Toxic Compounds, Population, Collaborative Competition

INTRODUCTION

When we are able to determine the level of toxic response present in a given population, we would also be able to assist in the establishment of safe exposure levels to the alien compounds. We would also be able to identify the section of the populace that are at a significantly high risk due to the adverse effects from the contamination. The present assessment of risks cannot be held responsible for the individual disparities that exist in the response from the chemical exposure. In addition, we perform safety testing at standard levels on a little fraction of the environmental compounds that are already in existence (Judson, R. et al., 2009). This also employs costly animal archetypes, which also consume a lot of time and isn't consistent when it comes to reflecting the safety profiles of individuals (Jacobs, A.C. & Hatfield, 2013). Algorithms with the capacity to supply correct in silico prognostication of human safety risks can make for an error-free and cost-efficient tool that will help in the identification of the possible risks to the given populaces (Bynagari, 2014). Nonetheless, predictions from the past have been hampered by the scarcity of data regarding the variability of populations as well as the drawbacks associated with extrapolation from prototype organisms (Zeise, L. et al., 2013; Dorne, J.L.C.M, 2010).

When there is a development of high throughput in studies concerning vitro toxicity, the utilization of human-based prototypes (Abdo, N. et al., 2015) and meteorically reducing sequencing costs have helped sizable and genetically differentiated populations to undergo characterization. In vitro systems, significant throughput has been used to successfully gain access to the changes existing in the transcriptional (Burczynski, M.E. et al., 2000) and phenotypical (Uehara, T. et al., 2011) traits as it has to do with response to the exposure of the considered compounds (Kleinstreuer, N.C. et al., 2014).

Continually, genomically (Caliskan, M., Cusanovich, D.A., Ober, C. & Gilad, Y.. 2011) attributed lines of cells with ever decreasing non-genetic variation sources have been employed in the identification of genetic variants and transcripts in relation to vitro and clinical drug responses alike (Ref 11 and 12). With these technologies, it is possible to enable the automated toxicity tests for an extensive array of compounds in lines of human cells (Ganapathy, 2015). This can be done to assess the level of responses among the population and also examine the variation between the associated risk profiles across the numbers (Ref 13). This project is a salient aspect of a community-based open challenge with the Dialogue on Reverse Engineering Assessment and Methods, also known as DREAM (Margolin, A.A. et al., 2013; Costello, J.C. et al., 2014; Ganapathy, 2016).

The researchers who participated in the DREAM project were asked to predict interpersonal variability in response to cytotoxicity. The researchers predicted based on genomic and transcriptional profiles. The challenge also had them predict the parameters of cytotoxicity at the population level and across chemicals that are based on the structural characteristics of the compounds involved in the process. Toxicity on the cellular level was examined for 156 compounds across the cell lines of lymphoblastoid as derived from 884 individuals from clear-cut regional subpopulations sourced from Asia, Europe, Africa and the Americas (1000 Genomes Project Consortium. et al, 2012). The transcriptional and genetic information obtained from these cell lines were available under the auspices of the 1000 Genomes Project (1000 Genomes Project Consortium. et al., 2010).

The set of data possesses double the count of cell lines and thrice the amount of compounds in comparison with the study which previously stood as the largest (Brown, C.C. et al., 2014). The paper carried out an evaluation on the submitted cutting-edge

approaches to modeling in order to put a benchmark on the existing most reliable practices in the predictive modelling universe. Additionally, the community challenge was able to pinpoint the algorithms capable of prediction with more improved random correctness, personal and group-level responses to various compounds factoring in only genomic information. Despite the reality that these outcomes are a representation of the betterment occurring over previous efforts to forecast the response in cytotoxicity, significant improvements for predictive correctness never cease to be critical.

In the DREAM contest, the cytotoxic data employed is made up of the EC10 information, which was generated from the lines of cells in response to the common 156 compounds based in the considered environment. The participants were equipped with a set of training cytotoxicity information. This data was provided for 620 lines of cells as well as 106 compounds. Then, it was coupled with genotype-related data for the entire cell lines, RNA-seq information for 337 lines of cells and chemical characteristics for all the compounds involved. DREAM was split into two non-dependent sub contests.

In the first sub challenge, those who partook were requested to forecast EC10 values for a different test group of 264 lines of cells in reaction to the 106 compounds in the environment. On a mindful note, just 91 toxic compounds were utilized for the eventual scoring procedure. Whereas, in the second sub challenge, the participants were asked to accurately determine the parameters of the concerned population in terms of middle-stage EC10 values as well as 5th interquartiles distance. The predictions were made for a different test set comprising 50 separate compounds.

PARTICIPATION IN THE DREAM CHALLENGE

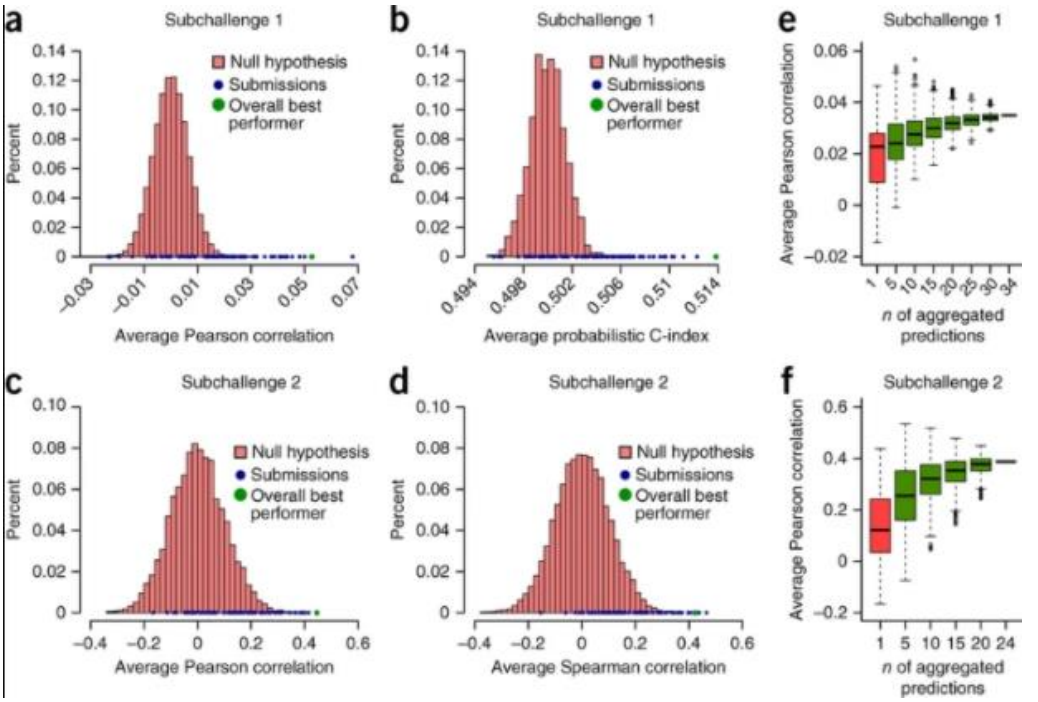
Fanning from over 30 countries, a total of 213 people registered to enter the NIEHS-NCATS-UNC DREAM Toxicogenetics Challenge. NIEHS-NCATS-UNC is the acronym used to describe the collaboration between the National Institute of Environmental Health Sciences, the National Center for Advancing Translational Sciences and the University of North Carolina. In this seemingly social experiment, people who partook were supplied with data subsets to learn models for a period of 3 months. Furthermore, the models were evaluated on an additional subset of test data that participants were not allowed to see. The learning information involved, firstly, a quantification of the cytotoxicity proneness per cell line. Considering EC10 values, the rate at which a decrease of 10 percent in viability transpired. 106 compounds were tested for a span of 337 lines of cells.

The training data, secondly, included genotypes for all the 884 lines of cells. It, thirdly, comprised RNA-seq-hinged measurements of genetic transcripts for 337 lines of cells. Fourthly, the training data consisted of the structural characteristics of the whole 156 separate compounds. A sum of 34 research groups tendered 99 forecasts or interpersonal variability in response to the first sub challenge. Another 23 research teams came up with 80 different predictions based on the toxicity parameters at the population level in reaction to the second sub challenge. DREAM Toxicogenetics Challenge offered the unmatched avenue to compare the predictive performance across an extensive array of ultra-modern approaches to the prediction of cytotoxic reaction to compounds existing in the environment.

The First Sub Challenge: Determining Interpersonal Variability

In this sub challenge, evaluations were carried out on the given models based on their capacity to forecast EC10 values in a blind experiment that consists of EC10 values that have been quantified in 264 cell lines that were not a part of the learning set. The accuracy of the

predictions was graded with the aid of a pair of metrics. The first is the Pearson correlation ®, a metric that functions as an evaluator for the linear reliance that exists between predicted and technically quantified EC10 values. The second is the rank-based metric, wherein the underministic C-index 15 takes the probabilistic nature of the gold standard into consideration as a result of the technical provenance of the noise in the related measures through the evaluation of the concordance between ranks in cell line cytotoxicity.



Scoring-concerned analyses were restricted to 91 compounds, not including 15 that did not experience any cytotoxicity across the whole headcount. That was done in order to sidestep the emergence of noise among the ranking. For all the metrics, the research teams' overall rankings were arrived at by ranking the teams differently for every compound involved and then conducting an averaging across all the compounds. For this paper, we initially conducted assessments to ascertain where the forecasts were substantially more improved compared to the random (Neogy, & Paruchuri, 2014). This we did by comparing the average ® with the average pCi. We computed across the compounds to obtain each submission with the concording null model of aimlessly, empirically sourced EC10 values.

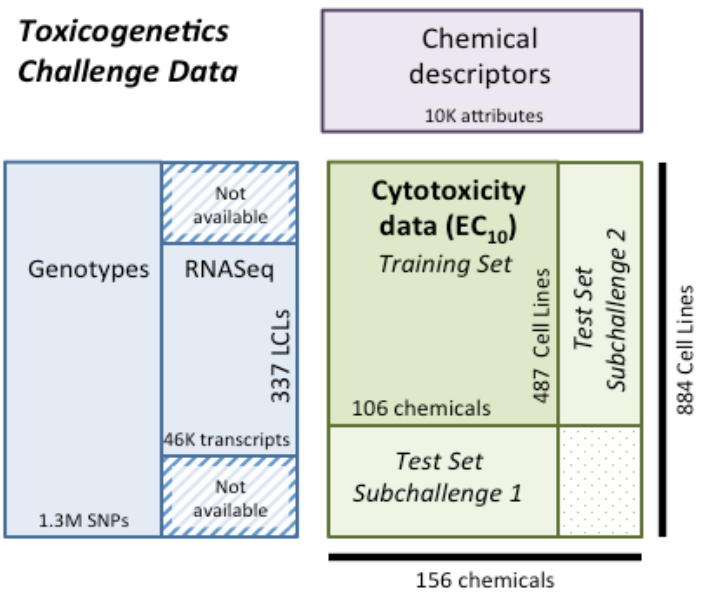
Of the total 99 submissions, the void postulation of indefinitely formulated forecasts can be bounced back based on the false rate discovery rate (FDR) principle (< 0.05), which auto-corrects for a multiplicity of conjectures. For 46 presentations using ®, 47 using pCi and 42 that blend both training metrics. Across all the considered compounds, the average values for ® and pCi were quite modest, with maximum and minimum values of 0.07 and 0.51, respectively. This suggests that cytotoxic reactions to exposure to chemical compounds is not generally well predicted based on a single nucleotide polymorphism data. In spite of the average predictive capacity turning out low, there was no uniformity in the performance among all the compounds. Variation levels in prognostic performance across the compounds start from −0.21 to 0.28 when it comes to ® values.

For pCi values, the performance ranged from 0.45 to 0.56. We carried out tests to determine whether the cytotoxicity present in each of the compounds can be predicted in a manner better than ordinary chance. For each of them, the forecast EC10 values for all the teams were juxtaposed with the void, unselective model. With this analysis, we could verify that the forecasts were, to a large extent, better than random guesses, for the majority of the involved compounds; 55 out of 91. This is despite the reality that some of the performances are quite poor. Rankings for the teams with the best performance proved wholesome duly considering the compounds employed in the scoring process.

Sub Challenge One: the Method with the Best Performance

The method with the most impressive performance for the forecasting of interpersonal variability in cytotoxin-related reactions as well had the capacity to determine using maximum ® that is equal to (average pCi = 0.51). Like the analysis for scoring, this method as well cancelled the 15 compounds that had no success inducing cytotoxic response. One set of 0.15 SNPs was chosen for inclusivity purposes for this analysis, using a pair of approaches (Vadlamudi, 2016).

The first approach involving SNPs are not synonymous existing within any gene, including SNPs close to genes defined by the 2kb upstream region and the 500bp. The second approach involves leftover SNPs if they are situated within close proximity with 41 KEGG20 gene member sets (Kanehisa, M. et al., 2014) They will be documented in the MsigDB database (Subramanian, A. et al., 2005) to symbolize the cycle or finality of cells, cancer or cancer biology.



If they were able to demonstrate how it correlates (P>0.06) with the representation of no less than one native gene leamning of the RNA-seq type of information, where comes in the eQTL analysis. The information that is stored inside this set of SNP was later archived into ten clusters believed to be generic. Doing so, k-means clustering was put into use based on the initial trio of basic components derived by scaling analysis in various dimensions (Purcell, S. et al., 2007; Ganapathy, 2016).

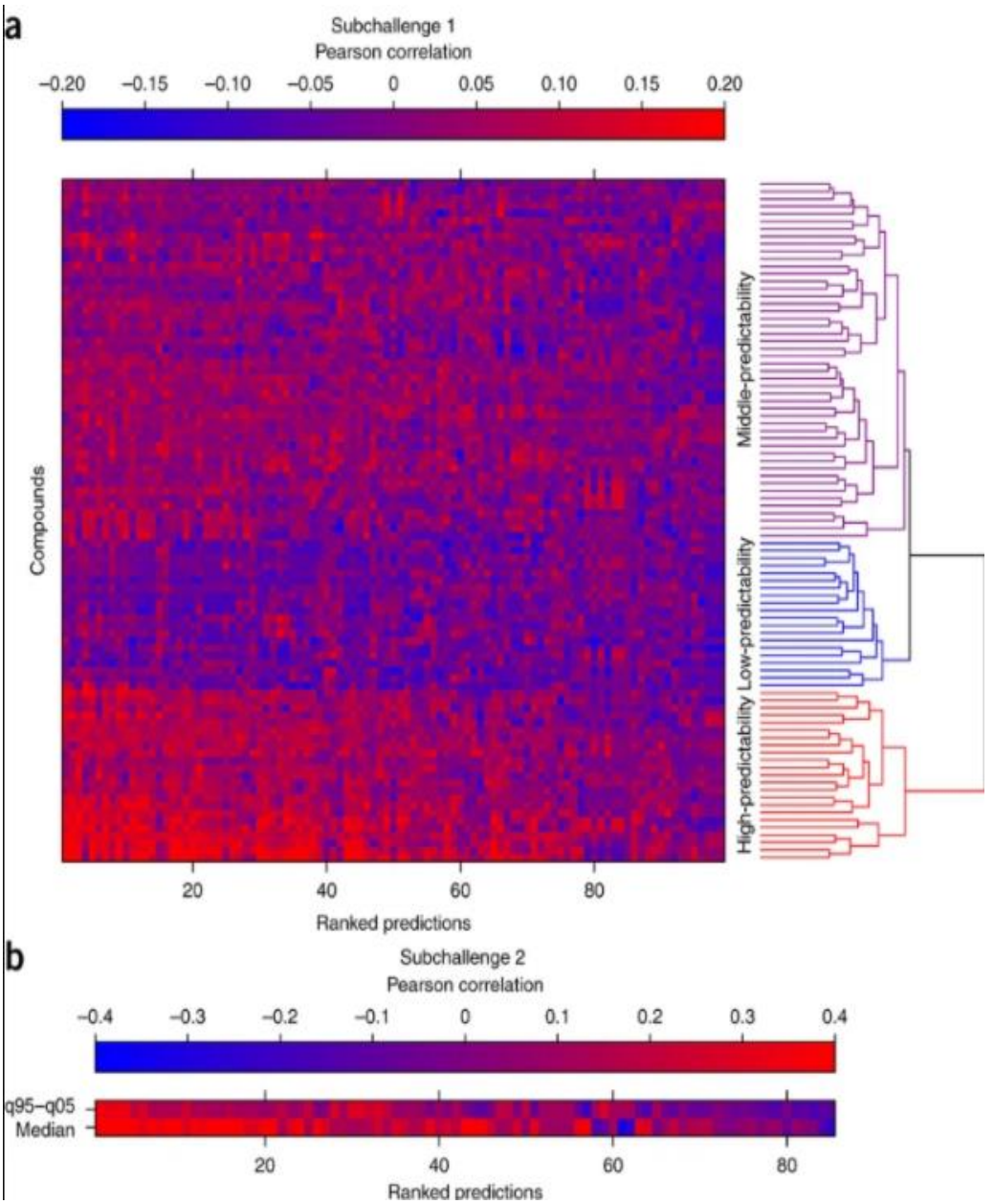
The variable genetic cluster that resulted from the experiment was high symbolic of comprehended geographic sub-populaces. However, it also comprised extra data that was not necessarily represented by each sub demography. With this variable, a cytotoxicity model was built for each of the compounds using the algorithm called Random Forest. Random Forest was combined with sex, geographical locale and experimental groups. In order to choose the parameters needed for clustering and to decide on the approaches for SNP filtering, cross-validation was observed. In the last scoring stage, the forecast method garnered the strongest performance in the midst of dozens of considered models, all of which are judged by experiment-derived true-response information. The entailments of this particular modelling method are further addressed in supplementary predictive models and online techniques.

The Second Sub Challenge: Predicting Parameters at the Population Level

Predicting parameters at the population level was graded with the aid of both the Pearson correlation ® and the Spearman correlation (rs) alongside an approach that is in semblance to the previous contest. For both stats, the worldwide performance of each team's submission was assessed through the averaging of the correlations that have been separately computed for median EC10 values. That is a vivid representation of the archetypal response to cytotoxicity. The assessment also factored in the disparity between the 5th and 9th percentiles for EC10 values, which, in turn, represents a quantification of the dispersion rate of the populace. The juxtaposition with the void model of indefinite forecasts was performed in order to conduct an assessment on the statistical symbolism of compound forecasts. Out of the 80 predictions that were tendered, the null postulation of unevenly formulated predictions was not accepted (FDR < 0.05) for 13 predictions using average ® and for 17 that were using average (rs). For a 13 prediction count, the void proposition was not accepted when both metrics were applied. The same kind of result was derived when the Fisher's method was employed in the assessment of the significance of each submission.

To emphasize, rankings for the teams with the highest levels of performance proved to be robust in relation with the compounds used for the scoring process. The medium cytotoxicity (which is also referred to as the average EC10) of compounds seemed easier to predict compared to the variability in the response of the population. That is, ® renegade from −0.31 to 0.66 while (rs) ranged from −0.29 to 0.72). In response to interquartile ranges, ® ranged from −0.22 to 0.37 while (rs) started from −0.14 to 0.48.

The general criterion for evaluating the second sub challenge teamed up the prediction of median with the interquartile range. The method that showed the most impressive performance when it comes to forecasting the median and interquartile distance with ® and (rs) is equivalent to 0.52 (0.45) and 0.37 (0.40). The flow of work used by this specific approach comprises four major steps: feature choosing, group selection, the development of the model and the trial compound forecast. Attributes were chosen from structural sources based on the attributes of the chemicals derived in three comparable ways. They are CDK23 (Steinbeck, C. et al., 2003 and SiRMS (Kuz'min, V.E., Artemenko, A.G. & Muratov, E.N., 2008) descriptors, the pair of which exist under the auspices of the facilitators of the Challenge, and the Dragon descriptors (Todeschini, R., Consonni, V., Mauri, A. & Pavan., 2004). Normalized separately, chemical descriptors that were in correlation with the core toxicity were used for learning the considered models. The models that use the Dragon descriptors had the highest performance level in both cross-validation and the eventual grading (Vadlamudi, 2015). In the steps, compounds were distributed into four distinct categories based on hierarchical jam-packing of their EC10 profiles in all of the 487 lines of cells. Built in separation for each compound category.



Best Performing Approach in Sub Challenge Two

Random Forest lexicons were used to choose the attributes that appear specific to forecasting in that group, of course, using smaller compounds as learning guidelines. For every new compound, toxicity was estimated with the use of a weighted median of forecasts from all four category-specific models. This is where the weights are determined according to their semblance with individual compound clusters. The similarity approach considered the distance from the cluster in the category-specific spaces containing

descriptors. The aforementioned learning methods were applied to determine the average EC10 values and the interquartile ranges alike.

When it came to median EC10, cell-line-facing forecasts were formulated with distinct models, after which they were averaged. Particularly for interquartile distance, a multi-set model was created to necessarily match the quantified interquartile distance for every compound involved. Additional information for the approach to modelling have been talked about in Supplementary Predictive Models and Online Methods. Regardless of the fact that this blueprint emerged with the best performance generally, other approaches like KSPA offered up a more ideal forecast of the average cytotoxicity with ® and (rs) respectively equivalent to 0.65 and 0.72.

COMPOUND-BASED PREDICTABILITY

Clearly, the compounds were ungrouped into three separate clusters, leaning on the correctness of the cytotoxic forecasts. A compound cluster for which determinations are high across the entire model chain of 14 compounds, a compound cluster for which forecasts are at a minimum across all models of 17 compounds, and a compound cluster for which predictions varied throughout all models. This observed separation proved consistent between the pair of metrics applied in the evaluation of team performances. After that, we tested for the characteristics that had the capacity to acutely differentiate between the compounds in the high-against-low clusters of predictability. A good number of chemical descriptors proved able when it comes to distinguishing between the high and low predictability of the compounds in the environment.

Noteworthy, a method known as the Lipinski rule (Lipinski, C.A., Lombardo, F., Dominy, B.W. & Feeney, P.J., 2001) was among those features that set it apart from the others. Similar to anticipations, the compounds existing in the high forecastable cluster possessed lower pooled variations, which makes them a lot less noisy compared to those in the cluster born of poor forecasts. Contradicting expectations, the chemical compounds found in the high-versus low clusters of predictability didn't show distinctive attributes considering the distribution of cytotoxic reaction throughout the populace in terms of average and interquartile radius nor to the approximated heritability of compound-based cytotoxicity (Bynagari, 2016). Be that as it may, we noticed that when a principal component analysis is in progress, on the cytotoxicity-related information, we were able to discern between the chemical compounds with increased and decreased forecast ability. That is an indication that the predictability was, at least, in part, as a result of the cytotoxic pedigree of the compounds throughout the surveyed populace (Ganapathy, 2016).

ANALYSIS OF SURVEYED DATA & CONCLUSION

For this paper, we received a total of 75 out of 99 response submissions for the first sub challenge. For the second sub challenge, we received 51 submissions out of 80. Correspondingly, this makes for 21 out of 34 teams for the first sub challenge and 14 out of 23 for the second tranche. Overviewing the information obtained from the survey, we noticed that there is an impact of used data and models on the performances of the submission (Bynagari, 2015). In order to manage the reality that every team submitted up to five predictions that could not be independent of each other, the forecasts made use of similar data and approaches—based on the survey-derived data—were averaged and taken as a single prediction. With this model, we were able to obtain 49 non-dependent submissions for the first and second sub challenge (Paruchuri, 2015).

That brings us to the data input for the forecasts. To create a solution for the first sub challenge, 89 percent of the participants who submitted predictions for the survey favored SNPs data as made available by the organizers of the exercise. The SNPs data were provided either alone or alongside other information using extra sources like pathway info and GO terms to sift through them. The RNA-seq data were used for nearly half (about 47 percent) of the submissions, thus showcasing the provision of general enhancement performances. Only a minor count of the participants (about 16 percent) as well included their preferred predictive models and details regarding the chemical descriptors.

For the second sub challenge, the majority of submissions, reaching 78 percent, did not consider any genetic information to forecast the cytotoxicity of the new compounds in the test environment. As regards the chemical properties, around 76 percent favored at least one chemical descriptor provided by the organizers of the challenge (CDK and SiRMS). Nonetheless, since as much as 45 percent of the teams partaking added exclusive information from external sources such as the ubChem37 and ChEMBL36 public databases or a variety of chemical descriptors such as PubChem37 (Wang, Y. et al., 2012) and ChEMBL36 (Gaulton, A. et al., 2012) open-to-all databases or various chemical descriptors in the lineage of ECFP38 and Dragon25.

REFERENCES

1000 Genomes Project Consortium. et al. A map of human genome variation from population-scale sequencing. Nature 467, 1061–1073 (2010).

1000 Genomes Project Consortium. et al. An integrated map of genetic variation from 1,092 human genomes. Nature 491, 56–65 (2012).

Abdo, N. et al. Population-based in vitro hazard and concentration-response assessment of chemicals: the 1000 Genomes high-throughput screening Study. Environ. Health Perspect. 123, 458–466 (2015).

Burczynski, M.E. et al. Toxicogenomics-based discrimination of toxic mechanism in HepG2 human hepatoma cells. Toxicol. Sci. 58, 399–415 (2000).

Bynagari, N. B. (2014). Integrated Reasoning Engine for Code Clone Detection. *ABC Journal of Advanced Research*, *3*(2), 143-152. https://doi.org/10.18034/abcjar.v3i2.575

Bynagari, N. B. (2015). Machine Learning and Artificial Intelligence in Online Fake Transaction Alerting. *Engineering International*, *3*(2), 115-126. https://doi.org/10.18034/ei.v3i2.566

Bynagari, N. B. (2016). Industrial Application of Internet of Things. *Asia Pacific Journal of Energy and Environment*, *3*(2), 75-82. https://doi.org/10.18034/apjee.v3i2.576

Caliskan, M., Cusanovich, D.A., Ober, C. & Gilad, Y. The effects of EBV transformation on gene expression levels and methylation profiles. Hum. Mol. Genet. 20, 1643–1652 (2011).

Costello, J.C. et al. A community effort to assess and improve drug sensitivity prediction algorithms. Nat. Biotechnol. 32, 1202–1212 (2014).

Dorne, J.L.C.M. Metabolism, variability and risk assessment. Toxicology 268, 156–164 (2010).

Ganapathy, A. (2015). AI Fitness Checks, Maintenance and Monitoring on Systems Managing Content & Data: A Study on CMS World. *Malaysian Journal of Medical and Biological Research*, *2*(2), 113-118. https://doi.org/10.18034/mjmbr.v2i2.553

Ganapathy, A. (2016). Blockchain Technology Use on Transactions of Crypto Currency with Machinery & Electronic Goods. *American Journal of Trade and Policy*, *3*(3), 115-120. https://doi.org/10.18034/ajtp.v3i3.552

Ganapathy, A. (2016). Speech Emotion Recognition Using Deep Learning Techniques. *ABC Journal of Advanced Research*, *5*(2), 113-122. https://doi.org/10.18034/abcjar.v5i2.550

Ganapathy, A. (2016). Virtual Reality and Augmented Reality Driven Real Estate World to Buy Properties. *Asian Journal of Humanity, Art and Literature*, *3*(2), 137-146. https://doi.org/10.18034/ajhal.v3i2.567

Gaulton, A. et al. ChEMBL: a large-scale bioactivity database for drug discovery. Nucleic Acids Res. 40, D1100–D1107 (2012).

Jacobs, A.C. & Hatfield, K.P. History of chronic toxicity and animal carcinogenicity studies for pharmaceuticals. Vet. Pathol. 50, 324–333 (2013).

Judson, R. et al. The toxicity data landscape for environmental chemicals. Environ. Health Perspect. 117, 685–695 (2009).

Kleinstreuer, N.C. et al. Phenotypic screening of the ToxCast chemical library to classify toxic and therapeutic mechanisms. Nat. Biotechnol. 32, 583–591 (2014).

Lipinski, C.A., Lombardo, F., Dominy, B.W. & Feeney, P.J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv. Drug Deliv. Rev. 46, 3–26 (2001).

Margolin, A.A. et al. Systematic analysis of challenge-driven improvements in molecular prognostic models for breast cancer. Sci. Transl. Med. 5, 181re1 (2013).

Neogy, T. K., & Paruchuri, H. (2014). Machine Learning as a New Search Engine Interface: An Overview. *Engineering International*, *2*(2), 103-112. https://doi.org/10.18034/ei.v2i2.539

Paruchuri, H. (2015). Application of Artificial Neural Network to ANPR: An Overview. *ABC Journal of Advanced Research*, *4*(2), 143-152. https://doi.org/10.18034/abcjar.v4i2.549

Uehara, T. et al. Prediction model of potential hepatocarcinogenicity of rat hepatocarcinogens using a large-scale toxicogenomics database. Toxicol. Appl. Pharmacol. 255, 297–306 (2011).

Vadlamudi, S. (2015). Enabling Trustworthiness in Artificial Intelligence - A Detailed Discussion. *Engineering International*, *3*(2), 105-114. https://doi.org/10.18034/ei.v3i2.519

Vadlamudi, S. (2016). What Impact does Internet of Things have on Project Management in Project based Firms?. *Asian Business Review*, *6*(3), 179-186. https://doi.org/10.18034/abr.v6i3.520

Wang, Y. et al. PubChem's BioAssay Database. Nucleic Acids Res. 40, D400–D412 (2012).

Zeise, L. et al. Addressing human variability in next-generation human health risk assessments of environmental chemicals. Environ. Health Perspect. 121, 23–31 (2013).

--0--