On the ChEMBL Platform, a Large-scale Evaluation of Machine Learning Algorithms for Drug Target Prediction

Naresh Babu Bynagari

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

ABSTRACT

Deep learning is currently the most successful machine learning technology in a wide range of application fields, and it has recently been used to forecast possible therapeutic targets and screen for active compounds in drug discovery research. However, it is unclear whether deep learning can outperform existing computational methods in drug discovery tasks due to the lack of large-scale studies, the compound series bias that is common in drug discovery datasets, and the hyperparameter selection bias that comes with the large number of potential deep learning architectures. As a result, we compared the outcomes of different deep learning methods to those of other machine learning and target prediction methods on a large-scale drug development dataset. We employed a stacked cluster-cross-validation technique to avoid any biases from hyperparameter selection or compound series. We discovered that (i) deep learning methods beat all competing methods, and (ii) deep learning's prediction performance is often comparable to that of tests conducted in wet labs (i.e., in vitro assays).

Source of Support: ChEMBL Platform, Machine Learning Algorithms, Drug Target Prediction

INTRODUCTION

The drug development procedure often entails a vast number of biological experiments and tests, referred to as "assays," that are used to assess the biological effects of chemical compounds. These effects, which include toxicity (Molina et al., 2013) and the inhibition or activation of proteins or entire biological processes, decide whether a chemical molecule will succeed or fail in its quest to become a commercial medicine.

It takes a long time and money to carry out these experiments. To obtain a single data point, a cell line must normally be grown. Even the multibillion-dollar Tox21 project (Huang et al., 2016), for example, could only screen a few thousand molecules for as few as twelve hazardous effects. As a result, precise computational target prediction methods are extremely beneficial in assisting and improving the drug discovery process.

Deep learning, a new computer technique that has made an impact in a variety of fields, has lately been successfully used not only to target prediction (Ma et al., 2015; Mayr et al., 2016) but also to a variety of other chemistry problems. For example, autonomous molecule generation (G´omez-Bombarelli et al., 2016; Segler et al., 2018; Olivecrona et al., 2017; Yang et al., 2017; Preuer et al., 2018; Paruchuri, 2017), chemical synthesis planning, (Segler et al., 2018) drug synergy prediction (Preuer et al., 2017), or modeling quantum interactions (Sch¨utt et al., 2017) and speeding up quantum mechanical computations (Smith et al., 2017), all of which could aid in the development of novel efficient molecular organic light-emitting diodes (G´omez-Bombarelli et al., 2016; Bynagari, 2014).

The primary purpose of this research was to compare the performance of deep learning with that of other approaches for predicting pharmacological targets.

LITERATURE REVIEW

Because deep learning architectures allow for multitask learning (Caruana, 1997; Deng et al., 2013; Bengio et al., 2013) and automatically create complex features, they appear to be well suited to target prediction (Bengio et al., 2013). First, multitask learning has the advantage of allowing for multilabel information and thus allowing for the use of target relationships. Hidden unit representations can be exchanged between predictions tasks when using multitask learning. Because few measurements are available for some goals, single task prediction may fail to generate an appropriate representation. Deep learning, on the other hand, can make use of representations acquired across tasks and with a few training examples, you can improve task performance. Figure 1 shows that many chemicals were measured assays (le), and – as a result of this finding – that there are assays with a high degree of correlation are available (right). Second, take a big breath. A compound's hierarchical representation is provided via networks. Higher levels denote more complicated properties (Bengio, 2013; Ganapathy, 2015). Single atoms are grouped together as functional groups and reactive centers, which in turn dene pharmacophores, resulting in a hierarchy of characteristics. The state-of-the-art approach in which chemists and medication designers think about the qualities of each chemical component is one of these features (Kazius et al., 2005).

When comparing drug target prediction algorithms, there are various problems to avoid, including choosing a comparison dataset, compound series bias in chemical datasets, and hyperparameter selection.

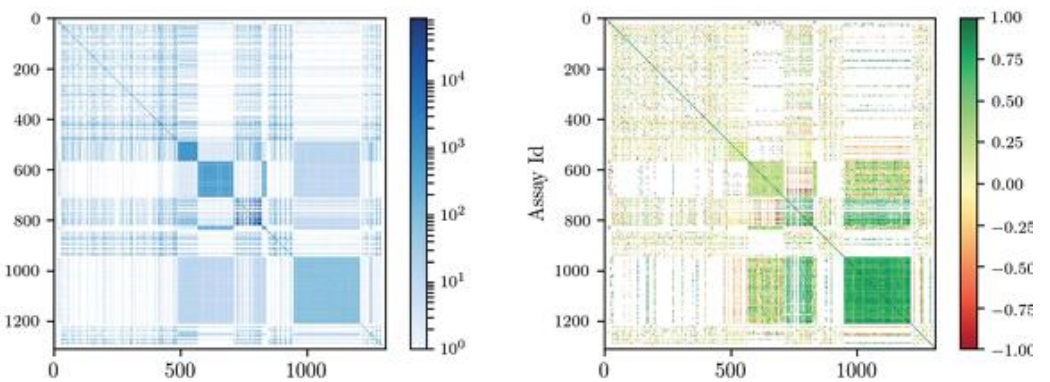

Figure 1: Assay correlation (left: number of compounds (log-scaled) measured on both assays, right: Pearson correlation on commonly measured compounds).

To begin with, many technique comparison studies only include a single or a small number of assays or targets (Ma et al., 2015; Ramsundar et al., 2015; Kearnes et al., 2016; Koutsoukas et al., 2017), whereas compound databases, such as ChEMBL (Bento et al., 2014; Bynagari, 2015), have many more tests.

As a result, despite the enormous amount of publicly available data, these research limit the conclusions of method comparisons to a small group of tests and underestimate the multitask learning effect. Some target prediction algorithms can use information from related assays to improve the predictive performance of a specific experiment. Multitask learning algorithms are the name given to such algorithms. Information from similar tests can help assays with few measurements in particular. Other potential benefits of multitask settings are overlooked, like the ability to generate predictions for a large number of experiments at once, which might aid chemists and biologists in conceptualizing how specific chemicals would operate at the cellular level. As a result, including a large number of assays in a technique comparison research is highly desired in order to assess the benefits of multitask learning in terms of prediction performance and to provide more general, useful information comparative declarations on target extrapolation approaches.

Second, most comparison studies suffer from compound series bias (Sheridan, 2013), which leads to an overestimation of certain approaches' performance. Chemical compounds are often formed as chemical scaffolds rather than individual compounds, and new substances are derived from these scaffolds by adding various functional groups (Ganapathy, 2016b). Predicting target activity for a compound from a new compound series is more challenging than predicting target activity for compounds from a series that is already in the training set (Bynagari, 2016). As a result, if the projected prediction performance suffers from compound series bias, it is overoptimistic in comparison to how the prediction method is employed in practice to forecast compounds from fresh compound series.

Third, hyperparameter selection biases performance estimates (hyperparameter selection bias). This is particularly true in deep learning, which allows for a wide range of architectures, activation functions, learning rates, and regularization parameters. If the adjustment of hyperparameters for creating predictive models is influenced by label information from the test set, bias may appear (Vadlamudi, 2016). In practice, however, no test set labels exist to alter hyperparameters.

As a result, the estimation of prediction performance is frequently overoptimistic. Because different learning algorithms have varying amounts of hyperparameters and varied adjustment capabilities for the hyperparameters, different learning algorithms have different overt tendencies. As a result, a technique comparison influenced by hyperparameter selection bias is usually unjust.

METHODS

To circumvent the first issue, we used the ChEMBL database to extract a large benchmark dataset that enables for accurate evaluation of machine learning approaches for compound target prediction. There are over 500, 000 chemicals and over 1, 000 assays in the dataset. These assays are different in size and correspond to a range of target classes (e.g. enzymes, ion channels, and receptors) (Ganapathy, 2017). Many of the assays in the dataset have only a few measurements (a few hundred to several hundreds), however there are also numerous assays with a huge number of observed chemicals (tens of thousands).

Cluster-cross-validation4 is used to solve the second problem. The set of data points is randomly partitioned into many folds in traditional cross validation. Each fold serves as a test set once during processing, with the remaining folds forming the training set. The training set is accessible to develop a new predictive model in each iteration, while the model's prediction performance is estimated on the test set. Cluster-cross-validation distributes entire clusters of compounds across folds rather than distributing data points to folds randomly (Bynagari, 2017). As a result, chemical molecules from the same cluster can be found in either the training or test sets. Cluster-cross-validation, in particular, prevents some data points from a given cluster from being included in the training set while others from the same cluster are included in the test set (Ganapathy & Neogy, 2017).

A machine learning system must therefore accurately predict the activity of chemicals from fresh scaffolds in a large number of examples in a cluster-cross-validation benchmark. Cluster-cross-validation analyzes the performance of techniques for predicting compounds based on novel chemical scaffolds, taking into account the manner chemical compounds are formed. Applying a stacked cross-validation scheme to the third problem solves it (Baumann and Baumann, 2014; Hochreiter and Obermayer, 2004). The algorithms' prediction performance is measured in the outer loop, while the hyperparameters of the various techniques are adjusted in the inner loop, allowing the methods to choose their optimum settings for generating predictive models in the outer loop. In our layered cluster-cross-validation configuration, we used a total of three distinct folds.

The inner loop employs one of our benchmark dataset's three folds for training and one fold for validating the hyperparameter combinations found in each iteration, while the outer loop uses the final fold as a test fold. For training a model, the outer loop employs the inner loop's training and test folds. The outer loop hyperparameters are chosen based on an inner loop cross-validation prediction performance criterion. As a result, hyperparameter selection does not skew the performance estimates produced by layered cross-validation.

We conducted an experiment that compares the accuracy of in silico predictions to the accuracy of in vitro measurements, in addition to an in silico prediction performance comparison study. We explicitly discuss the issue where two assays are dissimilar but must evaluate the same biological effect of a chemical. We compared whether a virtual assay or a surrogate in vitro assay is more accurate at predicting the outcome of an assay of interest because our in silico prediction method might be considered a virtual assay.

RESULTS AND DISCUSSION

In order to create a benchmark dataset, we treated target prediction as a binary classification issue. The goal is to predict a binary assay result, which reveals whether a molecule binds to a certain receptor, inhibits a pathway, or causes hazardous effects. Even if the assays under consideration share the same biomolecular target, each ChEMBL experiment is treated as a separate classification problem.

As a result, we avoid combining results from incompatible types of assays (Kalliokoski et al., 2013) (for example, binding assays, antagonist assays, and agonist assays cannot be compared because an antagonist is negative in an agonist assay and vice versa). We created a methodology for applying binary labels to the assay results because the raw assay measurement signal is only a real number, and binary labels are not provided. As a result, we were able to generate a large-scale benchmark dataset from ChEMBL.

We compared the predictions of several deep learning architectures with a variety of methods, including support vector machines (Cortes and Vapnik, 1995) (SVMs) and K-nearestneighbors (KNNs) as examples of similarity-based classification methods and random forests (Breiman, 2001) (RFs) as an example of feature-based classification methods. In addition, we compared naïve bayes (NB) and SEA (Keiser et al., 2007; Keiser et al., 2009; Keiser et al., 2009), which we deemed to be examples of target prediction algorithms created specifically for drug discovery. They normally get a whole 2D image as input, and one of the most distinguishing features of this network type is that parameters are shared among neurons. CNNs have multiple convolution and pooling layers, with the convolution layer outputs often computed using a parametrized kernel and the pooling layer outputs typically computed using a simple aggregation function. In this paper, we look at graph convolutional networks that utilise neighbourhoods defined by a molecular graph topology rather than pixel neighbourhoods as in 2D images. We specifically examined two implementations.

We employed the area under the receiver operating characteristic curve (Hanley and McNeil, 1982) (abbreviated as ROC-AUC or, as it is our default metric, simply AUC) as a performance assessment criterion for comparing target prediction algorithms. The AUC criterion is a popular metric for evaluating computational target prediction systems (Dahi et al., 2014; Huang et al., 2016).

We identified assay pairs in ChEMBL that evaluate the same biological effect in order to compare in silico predictions to in vitro measurements. The assay with fewer measured compounds was used as the ground truth, while the assay with a larger number of measured compounds was used as the surrogate assay. The surrogate's in vitro prediction accuracy was then compared to the in silico prediction accuracy.

There are 456, 331 compounds in the ChEMBL benchmark dataset, which we constructed and utilized to assess several target prediction algorithms. The molecular graphs of chemical compounds are used to describe them. Only graph convolutional networks, on the other hand, can handle graphs directly. We developed a sequence or a vectorial representation of the compounds for the other compared machine learning methods. We created the SMILES representation, which is used as an LSTM input. We used conventional software to construct a number of chemical descriptors for procedures that require numerical vectors (Cao et al., 2013; Hinselmann et al., 2011). Static features, semisparse features, toxicophore features, and dynamic features were all lumped together into four categories. Experts usually identify static features that indicate specific molecular properties. Their number is usually axed, while dynamic properties are retrieved in a prespecified manner on the y from a compound's chemical structure. Dynamic features typically have sparse binary or count distributions, implying that only a tiny percentage of compounds have the characteristic. Static features, on the other hand, are more likely to have continuous or sparse distributions.

The amount of semisparse features is predetermined, just like static features, however the construction concept is comparable to dynamic features. Toxicophore characteristics are the absence or presence of a set of predetermined structural alarms, known as toxicophores, in a chemical (Vadlamudi, 2017). We compared the prediction performances for the following feature categories or combinations of feature categories individually: common static features (Cao et al., 2013) (StaticF), common semisparse features (SemiF), including MACCS descriptors, as well as extended connectivity ngerprint features (Rogers and Hahn, 2010) (ECFP) and depth rst search features (Swamidass et al., 2005) (DFS).

Large-scale comparison

We obtained a performance estimate for each technique, feature category, and assay using our nested cluster-cross-validation procedure, which we refer to as "assay-AUC" (mean of ROCAUC values over the folds). Compound series and the hyperparameter selection technique have no effect on this estimate. For ECFP6 characteristics, the distribution of the assay-AUC values is also displayed in Figure 2. We used Wilcoxon signed rank tests between all pairs of algorithms to see if one method significantly outperformed another. The p-values for ECFP6 characteristics as well as the combination of ECFP6 and ToxF (ECFP6 + ToxF).

Note that we couldn't utilize the static features for the NB statistics since the approach required binary features; instead, we used a binarized version of the features that mapped all count values over zero to one for the other feature categories in NB. We also calculated solely SEA results for ECFP6 characteristics. We skipped the computation of the other feature categories due to the low performance compared to other approaches and the high computing demand.

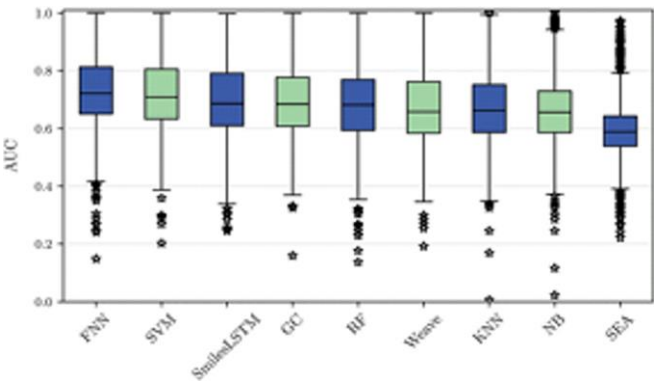

Figure 2. Performance comparison of drug target prediction methods.

For each feature category, we found that FNNs significantly outperform (a 14, 0.01) the other investigated approaches. Furthermore, for practically all feature categories except StaticF features, FNNs outperform graph convolutions (GC, Weave) or SmilesLSTM. SVMs are the second best approach. If ECFP6, ECFP6 + ToxF, or SemiF features are employed, they are significantly better than graph convolution networks or SmilesLSTM. The SmilesLSTM has a higher average AUC than the two graph-based techniques, which is surprising. It may also be seen that traditional machine learning approaches, such as SVMs or RFs, outperform traditional chemoinformatics methods. Many algorithms benefit from ECFP6 + ToxF features in general, however FNNs based on the feature category "SemiF" obtain the best results. FNNs have the best overall performance across all prediction tasks.

Machine learning models as virtual assays

We tested whether FNNs can predict assay outcomes as precisely as another (surrogate) in vitro assay measuring the same target, because we identified deep learning as the best method for compound target prediction. Our technique for establishing the benchmark dataset, which included the use of weakly active and weakly inactive drugs, determined the activity of both considered in vitro assays (Vadlamudi, 2015). In the case of in silico tests, the anticipated activity is determined by the computer model, while in the case of surrogate in vitro assays, it is determined by the data. We can compare the performance of an in silico and an in vitro assay in this way.

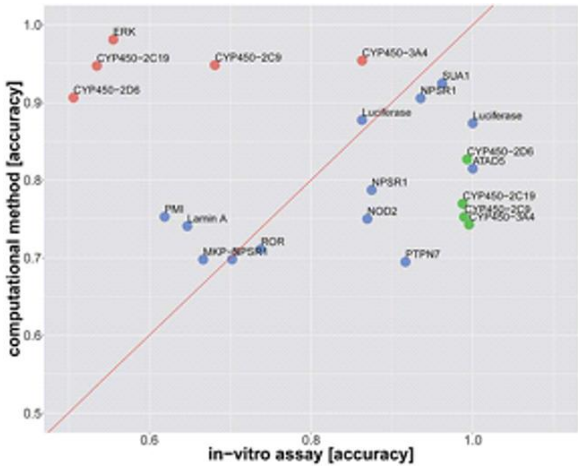

Figure 3. Comparison of prediction accuracy for an in vitro assay.

The average accuracy for predicting a selected assay using a surrogate in vitro assay that measured the same target was 0.81 - 0.17, while the average accuracy for predicting the selected assay using DNN models was 0.82 0.10. There was no significant difference in accuracy between the surrogate in vitro assay and the computational technique for 13 of the 22 assay pairings (see Fig. 3). In five of the 22 cases, the computational technique was more accurate than the surrogate in vitro assay. The surrogate in vitro assay outperformed deep learning in four of the 22 assays. Overall, deep learning's predictive performance for an assay with a specific target is comparable to surrogate assays assessing the same target.

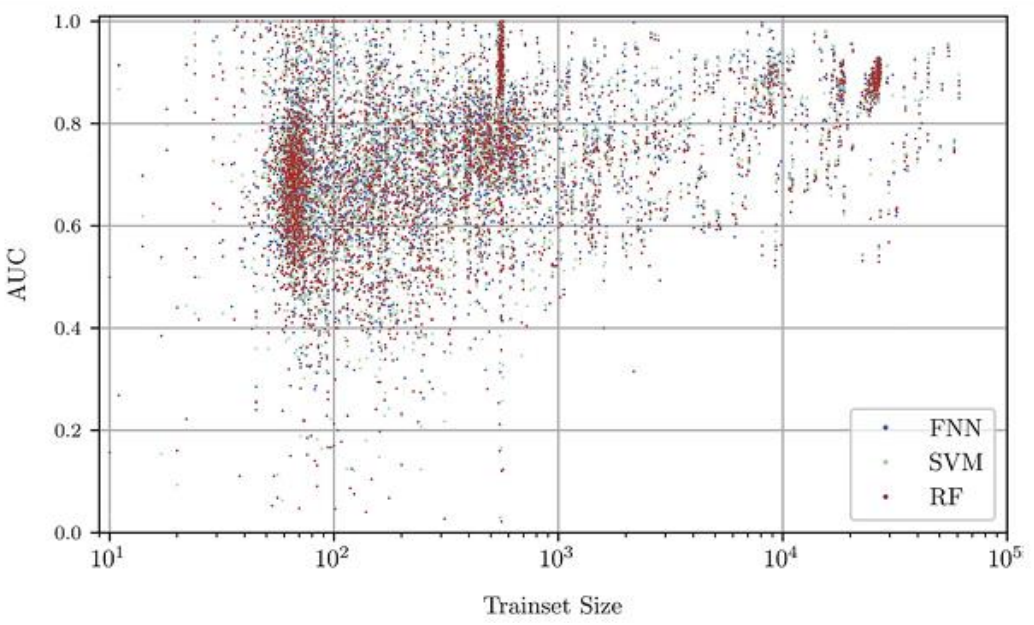

Figure 4. Scatterplot of predictive performance (“AUC”, y-axis) and size of the training set (“trainset size”, x-axis). Colors indicate three different predictive methods, namely FNNs, SVMs, and RFs.

Dataset size and prediction performance

We looked at the association between the AUC values of the test set and the size of the training set to see if there was a link between dataset size and performance (see Figure 4). In principle, it can be seen that higher training set sizes lead to better predictions (Neogy & Paruchuri, 2014). Even if there are only a few examples available, the AUC reveals that the performance is almost always superior to random classification (Paruchuri, 2015).

Prediction performance for different ChEMBL target classes and assay types.

We also looked into if there were any changes in performance between different types of tests. To this goal, we looked at the primary ChEMBL target classes allocated to the assays, as well as the assay types to which an assay belongs (Ganapathy, 2016a). Figure 5 depicts a boxplot of prediction performance for each of the key ChEMBL target classes, while Figure 6 depicts a boxplot for the various assay types. On the one hand, the assays' principal target classes assigned by ChEMBL, and on the other side, the assay types to which an assay belongs. Figure 5 depicts a boxplot of prediction performance for each of the key ChEMBL target classes, while Figure 6 depicts a boxplot for the various assay types. The number of assays used to create the boxplots is listed next to the class or type name. It's worth noting that Figure 5 is based on only a subset of assays (those with annotations) and that assays can belong to multiple classes. Figure 5 indicates that the prediction performance of DNNs is clearly superior to random across all classes, implying that the application of DNNs may be quite broad and not limited to a few well-known targets. Deep learning works well for functional and binding experiments, as seen in Figure 6.

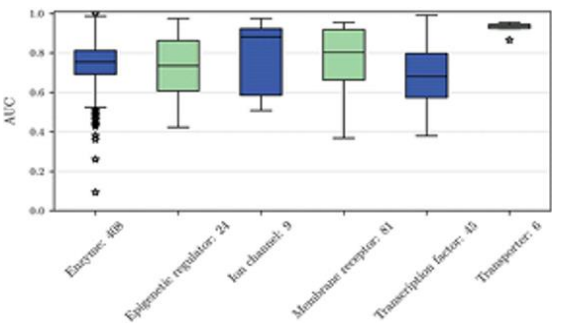

Figure 5. Boxplot of assay-AUC values for various assay classes when using a DNN on a combination of ECFP6 and ToxF features.

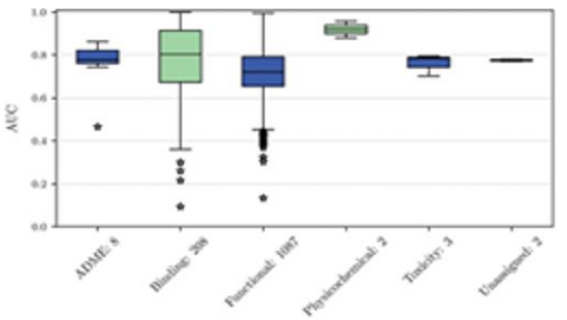

Figure 6. Boxplot of assay-AUC values for various assay types when using a DNN on a combination of ECFP6 and ToxF features

CONCLUSION

We evaluated deep learning's predicted performance to a number of other drug target prediction approaches, avoiding the normal biases in compound target prediction method comparison research. FNNs outperform other approaches for drug target prediction, according to our findings. This finding isn't limited to a single form of molecular descriptor, but rather applies to all sorts of molecular descriptors. Furthermore, we discovered that deep learning enables the creation of models with great prediction performance for a wide range of goals. As the training dataset grows, so does the performance. We also demonstrated that deep learning is comparable to – and sometimes even better than – surrogate in vitro assays for predicting a specific target. Large compound-assay databases, like as ChEMBL, provide enough data for very accurate deep learning models to be built. We believe that employing in-house databases maintained by pharmaceutical corporations as high-quality, large-scale training sets could lead to even more performance increases.

REFERENCES

Bahdanau, D., Cho K. and Bengio, Y. 2014. arXiv preprint arXiv: 1409.0473.

Baumann, D. and Baumann, K. 2014. *J. Cheminf.*, 6, 1.

Bengio, Y. 2013. *Proceedings of the First International Conference on Statistical Language and Speech Processing*, pp. 1–37.

Bengio, Y., Courville, A. and Vincent, P. 2013. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35, 1798–1828.

Bento, A. P., Gaulton, A., Hersey, A., Bellis, L. J., Chambers, J., Davies, M., Kr¨uger, F. A., Light, Y., Mak, L., McGlinchey, S. 2014. *Nucleic Acids Res.*, 42, D1083–D1090.

Breiman, L. 2001. *Mach. Learn.*, 45, 5–32.

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

Bynagari, N. B. (2017). Prediction of Human Population Responses to Toxic Compounds by a Collaborative Competition. *Asian Journal of Humanity, Art and Literature*, *4*(2), 147-156. https://doi.org/10.18034/ajhal.v4i2.577

Cao, D.S., Xu, Q.S., Hu Q.N. and Liang, Y.Z. 2013. Bioinformatics, 29, 1092–1094.

Caruana, R. 1997. *Mach. Learn.*, 1997, 28, 41–75.

Cho, K., Van B., Merri¨enboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk and Bengio, Y. 2014. arXiv preprint arXiv: 1406.1078.

Cortes C. and Vapnik, V. 1995. *Mach. Learn.*, 20, 273–297.

Dahl, G. E., Jaitly N. and Salakhutdinov, R. 2014. arXiv preprint arXiv: 1406.1231.

Democratizing Deep-Learning for Drug Discovery, Quantum Chemistry, Materials Science and Biology, https://github.com/deepchem/deepchem, 2016.

Deng, L., Li, J., Huang, J.T., Yao, K., Yu, D., Seide, F., Seltzer, M., Zweig, G., He, X., Williams, J., Gong, Y. and Acero, A. 2013. *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference*, pp. 8604–8608.

Duvenaud, D. K., Maclaurin, D., Iparraguirre, J., Bombarell, R., Hirzel, T., Aspuru-Guzik A. and Adams, R. P. 2015. *Advances in Neural Information Processing Systems* 28, pp. 2224–2232.

G´omez-Bombarelli, R., Aguilera-Iparraguirre, J., Hirzel, T. D., Duvenaud, D., Maclaurin, D., Blood-Forsythe, M. A., Chae, H. S., Einzinger, M., Ha, D.G., Wu, T. 2016. *Nat. Mater.*, 15, 1120.

G´omez-Bombarelli, R., Wei, J. N., Duvenaud, D., Hern´andez-Lobato, J. M., S´anchez-Lengeling, B., Sheberla, D., Aguilera-Iparraguirre, J., Hirzel, T. D., Adams, R. P. and Aspuru-Guzik, A. 2016. *ACS Cent. Sci.*, 4, 268–276.

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. (2016a). 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. (2016b). 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

Ganapathy, A. (2017). Friendly URLs in the CMS and Power of Global Ranking with Crawlers with Added Security. *Engineering International*, *5*(2), 87-96. https://doi.org/10.18034/ei.v5i2.541

Ganapathy, A., & Neogy, T. K. (2017). Artificial Intelligence Price Emulator: A Study on Cryptocurrency. *Global Disclosure of Economics and Business*, *6*(2), 115-122. https://doi.org/10.18034/gdeb.v6i2.558

Graves A. and Jaitly, N. 2014. Proceedings of the 31st International Conference on Machine Learning, 2014, pp. II-1764–II-1772.

Hanley J. A. and McNeil, B. J. 1982. Radiology, 143, 29–36.

Hinselmann, G., Rosenbaum, L., Jahn, A., Fechner N. and A. Zell, J. Cheminf., 2011, 3, 1–14.

Hochreiter S. and Obermayer, K. 2004. Kernel Methods in Computational Biology, MIT Press, pp. 319–355.

Hochreiter S. and Schmidhuber, J. 1997. Neural Comput., 9, 1735–1780.

Hochreiter, S. 1991. MSc thesis, Institut f¨ur Informatik, Lehrstuhl Prof. Dr. Dr. h.c. Brauer, Technische Universit¨at M¨unchen.

Hochreiter, S., Bengio, Y., Frasconi P. and Schmidhuber, J. 2000. A Field Guide to Dynamical Recurrent Networks, IEEE, pp. 237–244.

Huang, R., Xia, M., Nguyen, D.T., Zhao, T., Sakamuru, S., Zhao, J., Shahane, S. A., Rossoshek, A. and Simeonov, A. 2016. *Front. Environ. Sci. Eng.*, 3, 85.

Kalliokoski, T., Kramer, C., Vulpetti A. and Gedeck, P. 2013. PLoS One, 8, 1–12.

Kazius, J., McGuire, R. and Bursi, R. 2005. *J. Med. Chem.*, 48, 312–320.

Kearnes, S., Goldman, B. and Pande, V. 2016. arXiv preprint arXiv: 1606.08793, 2016.

Kearnes, S., McCloskey, K., Berndl, M., Pande V. and Riley, P. 2016. *J. Comput. Aided Mol. Des*., 30, 595–608.

Keiser M. J. and Hert, J. 2009. Chemogenomics, Humana Press, pp. 195–205.

Keiser, M. J., Roth, B. L., Armbruster, B. N., Ernsberger, P, Irwin J. J. and Shoichet, B. K. 2007. Nat. Biotechnol., 25, 197–206.

Keiser, M. J., Setola, V., Irwin, J. J., Laggner, C., Abbas, A. I, Hufeisen, S. J., Jensen, N. H.,. Kuijer, M. B., Matos, R. C., Tran, T. B., Whaley, R., Glennon, R. A., Hert, J.,

Koutsoukas, A., Monaghan, K. J., Li X. and Huan, J. 2017. *J. Cheminf*., 9, 42.

Krizhevsky, Sutskever A., I. and Hinton, G. E. 2012. Advances in Neural Information Processing Systems 25, 2012, pp. 1097– 1105

LeCun, Y., Bottou, L., Bengio Y. and Haffner, P. 1998. *Proc. IEEE*, 86, 2278–2324.

Ma, J., Sheridan, R. P., Liaw, A., Dahl, G. E. and Svetnik, V. 2015. *J. Chem. Inf. Model*., 55, 263–274.

Mayr, A., Klambauer, G., Unterthiner, T. and Hochreiter, S. 2016. *Front. Environ. Sci. Eng*., 3, 80.

Molina, D. M., Jafari, R., Ignatushchenko, M., Seki, T., Larsson, E. A., Dan, C., Sreekumar, L., Cao, Y. and Nordlund, P. 2013. *Science*, 341, 84–87.

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

Olivecrona, M., Blaschke, T., Engkvist, O. and Chen, H. 2017. *J. Cheminf*., 9, 48.

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

Paruchuri, H. (2017). Credit Card Fraud Detection using Machine Learning: A Systematic Literature Review. *ABC Journal of Advanced Research*, *6*(2), 113-120. https://doi.org/10.18034/abcjar.v6i2.547

Preuer, K., Lewis, R. P. I., Hochreiter, S., Bender, A., Bulusu, K. C. and Klambauer, G. 2017. *Bioinformatics*, 34, 1538–1546.

Preuer, K., Renz, P., Unterthiner, T., Hochreiter, S. and Klambauer, G. 2018. arXiv preprint arXiv: 1803.09518.

Ramsundar, B., Kearnes, S., Riley, P., Webster, D. Konerding, D. and Pande, V. 2015. arXiv preprint arXiv: 1502.02072.

Rogers D. and Hahn, M. 2010. *J. Chem. Inf. Model*., 50, 742– 754.

Russakovsky, O., J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg and Fei-Fei, L. 2015. *Int. J. Comput. Vis*., 115, 211–252.

Sch¨utt, K., Kindermans, P.J., Sauceda Felix, H. E., Chmiela, S., Tkatchenko, A. and M¨uller, K.R. 2017. *Advances in Neural Information Processing Systems*, 30, pp. 991–1001.

Segler, M. H. S., Kogej, T., Tyrchan, C. and Waller, M. P. 2018. *ACS Cent. Sci*., 4, 120–131.

Segler, M. H., Preuss, M. and Waller, M. P. 2018. *Nature*, 555, 604.

Sheridan, R. P., 2013. *J. Chem. Inf. Model.*, 53, 783–790.

Simonyan K. and Zisserman, A. 2014. arXiv preprint arXiv: 1409.1556.

Smith, J. S., Isayev, O. and Roitberg, A. E. 2017. *Chem. Sci*., 8, 3192–3203.

Sutskever, I., Vinyals O. and Le, Q. V. 2014. Advances in Neural Information Processing Systems 27, pp. 3104–3112.

Swamidass, S. J., Chen, J., Bruand, J., Phung, P., Ralaivola, L. and Baldi, P. 2005. Bioinformatics, 21, i359–i368.

Szegedy, C., Liu, W. Jia, Y., Sermanet, P., Reed, S., Anguelov, D. Erhan, D., Vanhoucke V. and Rabinovich, A. 2015. *The IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), pp. 1–9.

Thomas, K. L., Edwards, D. D., Shoichet B. K. and Roth, B. L. 2009. Nature, 462, 175–181.

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

Vadlamudi, S. (2017). Stock Market Prediction using Machine Learning: A Systematic Literature Review. *American Journal of Trade and Policy*, *4*(3), 123-128. https://doi.org/10.18034/ajtp.v4i3.521

Weininger, D. 1988. *J. Chem. Inf. Comput. Sci.,* 28, 31–36.

Wu, Z., B. Ramsundar, E. N. Feinberg, J. Gomes, C. Geniesse, A. S. Pappu, K. Leswing and Pande, V. 2018. Chem. Sci., 9, 513–530.

Yang, X., Zhang, J., Yoshizoe, K., Terayama, K. and Tsuda, K. 2017. *Sci. Technol. Adv. Mater*., 18, 972–976.

--0--