Deep Neural Network Model for Predicting Gene Activity Using Three-dimensional Structures of Chemical Compounds

Md. Mohaiminul Islam[1,2], Kevin Jeffers [3], Andrew M. Hogan [4]
Qian Liu[2], Rebecca Davis[3], Silvia Cardona[4,5], Pingzhao Hu[1,2]

[1]Department of Computer Science, E2-445 EITC, University of Manitoba, Winnipeg, MB, R3T 2N2, Canada
[2]Department of Biochemistry and Medical Genetics, 745 Bannatyne Avenue, University of Manitoba, Winnipeg, MB, R3E 0J9, Canada
[3]Department of Chemistry, 360 Parker Building, University of Manitoba, Winnipeg, MB, R3T 2N2, Canada
[4]Department of Microbiology, 213 Buller Building, University of Manitoba, Winnipeg, MB, R3T 2N2, Canada
[5]Department of Medical Microbiology & Infectious Disease, University of Manitoba, Winnipeg, Canada

Corresponding: Rebecca Davis (Rebecca.Davis@umanitoba.ca), Silvia Cardona (Silvia.Cardona@umanitoba.ca), Pingzhao Hu (pingzhao.hu@umanitoba.ca)

Abstract
Experimental approaches to drug discovery are time-consuming and expensive. It is well-known that three-dimensional (3D) structures of chemical compounds contain rich information for drug screening. Therefore, it is critical to develop new models to measure compound structure-activity relationships. To solve this issue, we first developed an algorithm to extract compound structure-specific features from atomic coordinates of conformers created on a specific molecular conformation. A denoising stacked autoencoder model was then proposed to generate deep features. The network was built by stacking layers of denoising autoencoders in a convolutional way. Chemogenetic interactions were then predicted using a support vector machine based on the learned high-level feature representations of the 3D structures of the compounds. The models were evaluated using 59 compounds with 6413 conformers and 242 gene products generated by a chemical genomics strategy for mechanism-based profiling of antibacterial compounds. We demonstrated that the proposed model has excellent performance to classify chemogenetic interactions using the structure features extracted from the chemical compounds.

Key Words: Deep neural network, chemical compounds, three-dimensional structure, drug discovery, gene activity

1. Introduction

Deep learning (DL) is a branch of artificial intelligence, which creates machine learning (ML) models based on existing data to predict the outcomes of new data sets [1, 2, 3]. A DL network is an artificial neural network that takes sample data as the input to the network, transforms the data into an abstract level, and provides predictions as the output. DL network can learn non-linear features using a different number of hidden layers. A hidden layer is a structure in a DL network where various types of non-linear operations can be performed. These layers receive weighted input from its previous layer and perform non-linear operations to transform the data, and pass them to the next layer. DL-based methods can create prediction models without any feature selection. The methods have already solved many problems that previously required human intervention, such as image classification [4], autonomous driving cars [5], face recognition based security [6,7], real-time object tracking and detection [8], natural language processing [9], and speech recognition models [10,11,12]. DL-based methods also achieved promising results in the field of biomedical applications, where image segmentation and computer aided diagnosis can directly extract features from the raw data and without any human engineered features [13, 14, 15]. Recently, DL-based methods have been applied to medical imaging processing [16], which were compared with predefined featured-based conventional methods [15, 17]. These comparisons showed that the superiority of deep learning methods to conventional machine learning methods.

An autoencoder is one of the popular neural network based architectures, which can perform unsupervised feature learning. Autoencoder calculates the difference between the observed/input data and the predicted data. This difference is then backpropagated through the network to update the weights of the hidden layers. The network learns to approximate an identity function by limiting fewer numbers of units in the hidden layer than the input layer. This method is better than the PCA (Principal Component Analysis) technique for dimensionality reduction problems. PCA can provide only linear transformation, while activation functions in the hidden layers of an autoencoder introduce "non-linearities" in encoding. Furthermore, we can form a DL architecture by stacking autoencoders on top of another. Cheng et al. used a type of deep learning-based autoencoder called stacked denoising autoencoder (SDAE) for computer-aided diagnosis regarding breast lesions in images and pulmonary nodules in CT scans [18]. This type of autoencoder is good for automatic feature extraction as well as provides convincing noise tolerance. Cheng et al. showed that their proposed model achieved substantial performance improvements over two conventional computer-aided diagnosis methods [19, 20].

Emerging and re-emerging infectious diseases are a critical public health issue. Infectious diseases are usually caused by microorganisms such as viruses, bacteria, and fungi [21]. Moreover, microorganisims can become drug-resistant, which suggest the need of screening for new antimicrobial drugs. Unfortunately, currently available tools are not able to provide new antimicrobial drugs at the rate required to solve the issues of emerging antimicrobial resistance [22]. Bueso-Bordils et al. [23] introduced a mathematical model which uses molecular topology to classify chemical compounds used to treat bacterial infections (antibiotics) and predict antibacterial activity in virtual compound libraries. Their proposed model used structural descriptors (non- three dimensional (3D) components) and linear discriminant analysis for this classification. While the authors identified 158 compounds as antibacterial candidates from 6,375 compounds, they did not use genome-wide chemical-genetic interaction data, which is a richer descriptor of antimicrobial activity and mechanism of action. In this study, we have proposed a general DL-based framework to predict chemical-genetic interactions in bacteria exposed to antibiotics, as a preliminary step in predicting antibiotic activity. Using previously published chemogenetic profiling of antibiotics in the Gram-positive bacterium *Staphylococcus aureus* (*S. aureus*) and 3D structural descriptors of the compounds, we built a chemogenetic interaction prediction model. We propose that this framework can be used for prediction of chemical-genetic interactions in other bacteria.

2. Methods

2.1 Dataset

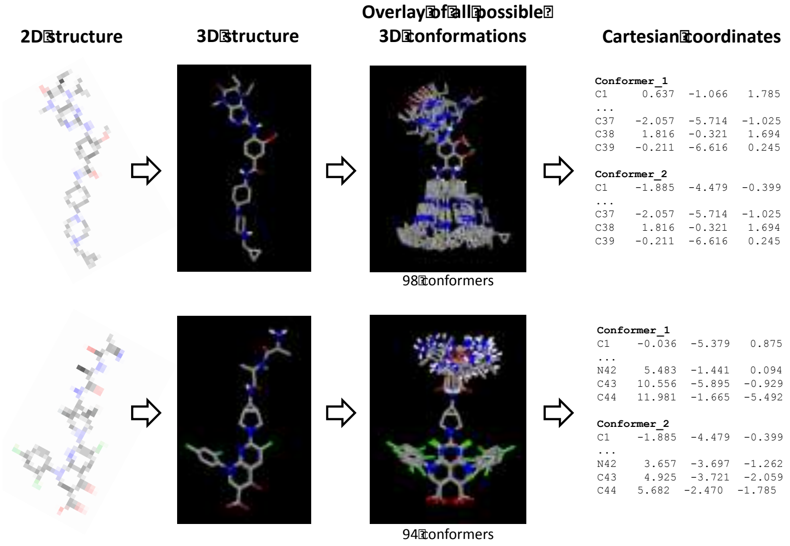
We have collected published chemogenomic profiles of *Staphylococcus aureus (S. aureus)* knockdowns exposed to 59 antibacterial compounds [24]. The study proposed a strategy called antisense induced strain sensitivity (AISS) to diminish the expression of essential genes (genes responsible for bacterial growth). The specific sensitivity of certain strains to the compounds reports chemical-genetic interactions. The study collected 245 antisense strains of *S. aureus* from 628 strains in which essential genes were identified by the conditional-growth phenotype of the knockdown mutation [25]. Donald et al. used 59 antibacterial compounds to test the AISS strategy to determine their mechanism of action [24]. In our experiment, we used 242 genes corresponding to these 245 antisense strains exposed to 59 antibacterial compounds. Donald et al. defined the chemogenetic interactions between genes and compounds as active (interaction) or inactive (no interaction) [24]. For example, the alanine racemase (*alr*) and D-alanine-D-alanine ligase (*ddlA*) genes are inactive for the compound called D-cycloserine. We used these genes and compounds and their chemogenetic interactions to build a machine learning model to predict the activity of these 59 antibacterial compounds against the 242 genes.

2.2 Generation of machine learning samples

It is well-known that conformational isomerism means that two isomers can be converted into each other by single bond rotation. These conformational isomers are called conformers. In our experiment, we examined the 3D structures of the 59 profiled antibacterial compounds generating all possible rotational isomers (conformers) of each molecule. This fully scanned the chemical space available for each compound, providing a robust set of data for comparison with other compounds and ensuring that shape similarities between molecules are accurately described. We generated different number of conformers from these compounds by list of possible dihedral angles using OMEGA 3.0.1.2.: OpenEye Scientific Software, which forms a 3D structure for a given compound (**Figure 1**). We have generated a total of 6,413 conformers from these 59 compounds. Therefore, we had 6,413 conformers as our samples for building machine learning models for predicting chemogenomic profiles.

2.3 Feature extraction for each of the samples

To build our machine learning models, we extracted features from these conformer-specific samples. This required us to transform the conformer-specific raw data into an understandable data structure which can be used for predicting chemogenomic interactions. However, we did not have any specific format to represent the samples since they were represented as a set of Cartesian coordinates (**Figure 1**). Hence, we developed an approach to build a data structure



```
Conformer_1
C1      0.637  -1.066   1.785
...
C37    -2.057  -5.714  -1.025
C38     1.816  -0.321   1.694
C39    -0.211  -6.616   0.245

Conformer_2
C1     -1.885  -4.479  -0.399
...
C37    -2.057  -5.714  -1.025
C38     1.816  -0.321   1.694
C39    -0.211  -6.616   0.245
```

```
Conformer_1
C1     -0.036  -5.379   0.875
...
N42     5.483  -1.441   0.094
C43    10.556  -5.895  -0.929
C44    11.981  -1.665  -5.492

Conformer_2
C1     -1.885  -4.479  -0.399
...
N42     3.657  -3.697  -1.262
C43     4.925  -3.721  -2.059
C44    5.682  -2.470  -1.785
```

Figure 1. Generation of 3D data for molecules in training and tests sets. Conformers are generated through a torsion search process which examines the molecular scaffold and identifies the bonds that may freely rotate. A list of possible dihedral angles is then assigned to each rotatable bond and an exhaustive search is performed. This search is done using OMEGA 3.0.1.2: OpenEye Scientific Software. Each conformer is described by a set of Cartesian coordinates. These coordinates are used in the machine learning process.

to represent the samples. This data structure was used to generate the feature vectors (FVs) for all of the conformer-specific samples. The approach is shown in **Figure 2** and summarized as follows. Briefly, we first found out the unique information from all conformers of the compounds (coordinates with associated atoms). These coordinates were stored in a one-dimensional master vector (MV) of size 75,080. Therefore, each of the cells of the MV stored one of the unique coordinates with associated atoms. We then matched each conformer's coordinates as well as associated atoms with the cells of the MV. If there was an exact match of the coordinates, we assigned "1" in the corresponding cell of the MV, otherwise we assigned "0". This MV was the feature vector for this conformer. We repeated the above steps for all the 6,413 conformers to extract their 75,080-size feature vector from their associated raw Cartesian coordinates. Consequently, we had a 6,413 by 75,080 high dimensional matrix for all the samples.

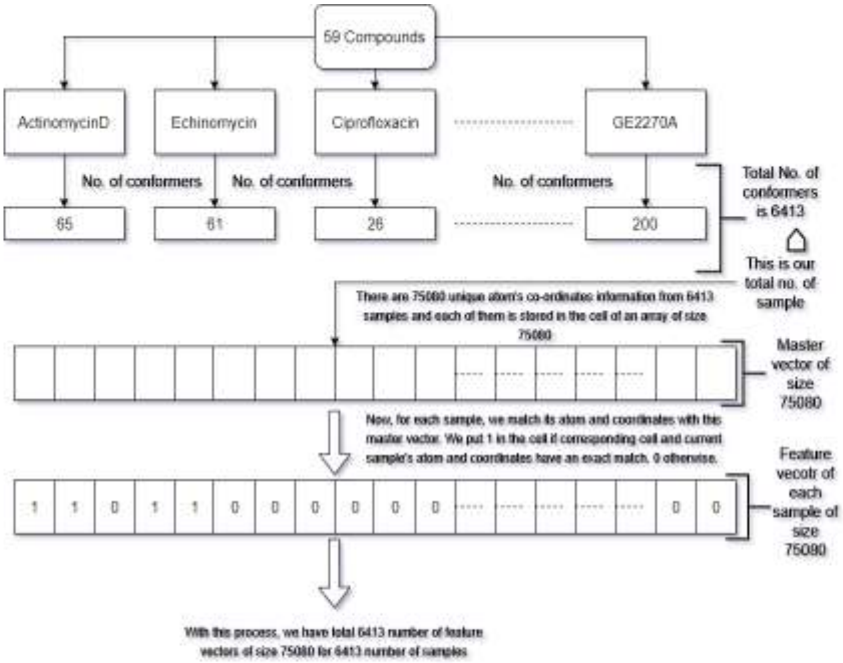

59 Compounds
ActinomycinD
Echinomycin
Ciprofloxacin
GE2270A
No. of conformers
No. of conformers
No. of conformers
65
61
26
200
Total No. of conformers is 6413
This is our total no. of sample
There are 75080 unique atom's co-ordinates information from 6413 samples and each of them is stored in the cell of an array of size 75080
Master vector of size 75080
Now, for each sample, we match its atom and coordinates with this master vector. We put 1 in the cell if corresponding cell and current sample's atom and coordinates have an exact match, 0 otherwise.
1 1 0 1 1 0 0 0 0 0 0 0 0
Feature vecotr of each sample of size 75080
With this process, we have total 6413 number of feature vectors of size 75080 for 6413 number of samples

Figure 2: Feature extraction for each of the conformers.

2.3 Deep features generation for each of the samples

Our extracted feature vectors were high dimensional. Some of the features may be redundant and these high dimensional feature vectors may introduce "overfitting" problem into our prediction model because of too many parameters to be estimated. Hence, we mapped our feature vectors into lower dimension to build an efficient machine learning model to predict the chemogenomic interactions. For this purpose, we used a DL based approach (SDCAE - Stacked Denoising Convolutional Autoencoder) to create low dimensional representation of our feature vectors. Denoising autoencoder can be used to reconstruct corrupted data, which may be common in our high dimensional data. Convolutional autoencoders replaces "fully-connected layer" by "convolutional layer". This operation reduces the total number of hidden units which means fewer parameters to be learned by our network. This helps extract correlated features as well as provides the network less prone to the problem of "overfitting".

The proposed SDCAE (**Figure 3**) is briefly summarized as follows. Firstly, we introduced noise into a proportion of the features, which were then input into SDCAE. Secondly, we performed convolutional operation to capture the correlation among the different coordinates of a conformer as well as to capture their local patterns. This operation reduced the total number of parameters to be learned by the network. Following this convolutional operation, we performed several fully connected operations via different hidden layers with different number of units which eventually encoded information into a lower dimensional vector. After these encoding operations, we performed a set of fully connected operations with different number of hidden units which eventually decoded the information from the encoder layers. Thirdly, we performed a deconvolutional operation to predict our noisy input data. Finally, we calculated the error between the output from our last decoder layer (i.e. the output from deconvolution operation) and the original data (i.e. noiseless uncorrupted data). We trained this network in a backpropagation style.

After completing training the SDCAE network, we took the output from the last encoder layer and this gave us a low dimensional matrix of all the samples of size 6,413 × 500. Therefore, we had a robust deep feature vector (DFV) with size 500 for each of the samples and we called them as deep features. We built our SDCAE model using a deep learning library known as CAFFE [22].

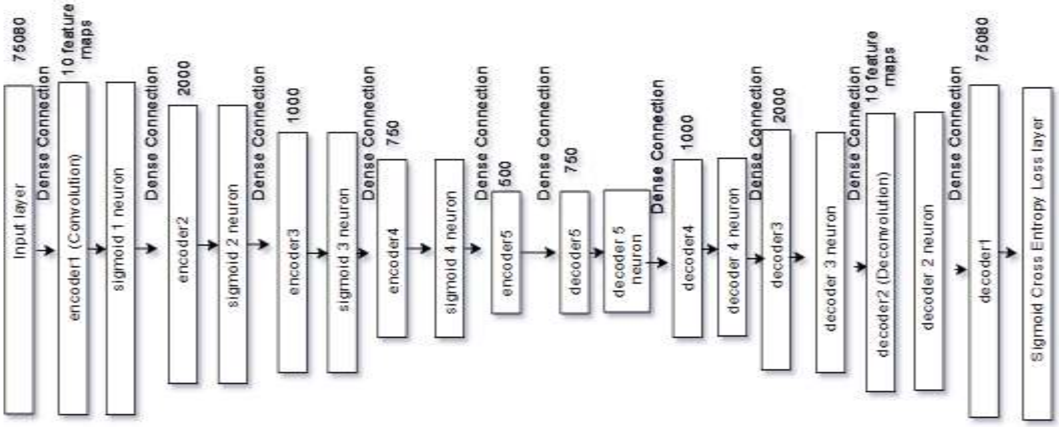


Figure 3: Stacked Denoising Convolutional Autoencoder

2.4 Building SVM models to predict chemogenomic interactions using deep features

Data mining is a process which involves extracting information from raw data as well as using this information to reach into a decision. Classification is a data mining process where objects whose labels are unknown are assigned with one of the known labels. This supervised process requires building a machine learning model to classify data instances into different categories. In this experiment, we have randomly selected 5,000 conformers as training set to build a machine learning model to classify a test set of the remaining 1,413 conformers into two classes: active or inactive. The depletion status (active vs inactive) of each gene-specific mutant was decided based on their chemogenomic profiles. Our goal is to build a machine learning model to predict whether a gene is active or not in a given conformer. This is a standard binary classification problem. We have 101 genes with the depletion status, which means we have to build 101 models to predict all the chemogenetic interactions. In order to build the models, we used the feature matrix of 5,000 (conformer samples) by 500 (deep features) as our training set (TN) and the rest of the 1,413 conformers were used as our test set (TE).

Although we can use a DL-based architecture to build a classification model using these deep features as DL achieved great success in solving the classification problems [4, 5], DL based models are expensive in terms of time and space requirements. A DL-based method takes almost an hour to build a prediction model using our dataset. So, we used another data mining tool known as support vector machine (SVM) [27,28] which takes much lesser time (i.e., ~3-4 mins) than DL-based methods to build a model with our dataset. This classification technique separates the data instances in a surface with a goal to maximize the margin of a decision boundary between the classes. SVM uses the knowledge from the training set to classify the test set. van de Wolfshaar et al. [29] used deep features to train SVM models to solve a classification problem of gender recognition. They achieved 94% to 96% accurate rate for the classification. Therefore, we built our SVM models using deep features as input to solve our binary classification problem. Deep features are robust against the noisiness in the data and encoded with the correlations among the raw features. We used a R package called e1071 [30] to build our SVM models.

3. Results

We had 59 chemical compounds and each of these compounds had different number of conformers. In total, we had 6,413 conformers from these 59 compounds. In our experiment, we treated these 6,413 conformers as our sample set. We also had information about 242 genes with depletion status against antibacterials. These genes were divided into two sets: 141 genes with inactive status in all compounds and 101 genes with either active or inactive in a compound. If a gene "X" is active in a compound "Y", it means the gene "X" is active in all the conformers of the compound "Y".

In our dataset, a conformer was represented by a set of Cartesian coordinates (**Figure 1**). Using the approach described in **Figure 2,** we represented each of all the 6,413 conformers with a 75,080-size one-dimensional vector with unique coordinates, where 1 represented that the given coordinate was found in the conformer and 0 represented that the given coordinate was not found in the conformer. Furthermore, we extracted the 500-size deep feature vector for each of the conformers using the unsupervised deep learning network architecture SDCAE (**Figure 3**). These deep features are robust against the unwanted noisiness and encoded with the correlation among all the 75,080 features. To build the SVM-based machine learning model using the selected deep features to predict whether a gene is active or inactive in a conformer, we randomly selected 5,000 conformers from our sample set (i.e., 6,413 conformers) to train the model for each gene and the rest of the 1,413 conformers to test the model. We used the same label for all the conformers of a compound. For example, the compound called "**Echinomycin**" has a label "**active**" for the gene "***metS***" which means all the available 61 conformers of "**Echinomycin**" have "**active**" label for the gene "***metS***".

To simplify the analysis, we focused on only the 101 genes with either active or inactive status in the conformers. Overall, more than 80% of the genes had prediction accuracy 80% or greater (**Figure 4).** We can see that the prediction performances measured by accuracy for majority of the gene were in the range 98.8% to 100%. We also observed that our proposed framework provides very impressing prediction performances in terms of sensitivity and specificity (**Figure 4.C2**). More than 80% of the genes had specificity close to 1.

4. Discussion

Today human health care is facing the threat of a "Post-Antibiotic Era" by infectious diseases because of the emergence of drug-resistant bacteria. Hence, antibacterial drug discovery is an important task in current health research. However, experimental approaches for drug discovery are time-consuming and expensive. It is necessary to explore how to use highly rich 3D chemical compounds for screening new antibacterial drugs. Here, we proposed a new framework which predicted gene activity in chemical compounds. Our model used a deep learning-based architecture called SDCAE to extract deep features from the 3D structural descriptors of antibacterial compounds for classification purposes. This SDCAE overcomes the curse of high dimensionality and incorporates the correlations among the different coordinates required to form a 3D chemical compound. We used SVM as the ML method to build a classification model using these robust deep features as input. Our experimental results showed that the framework has excellent performance for gene activity prediction. In addition, the proposed DL-based chemogenetic interaction prediction framework is not only limited to predict gene activity in *S. aureus* but also can be used for other bacteria.

In this experiment, to extract features from the 3D chemical compounds, we used a small sample size with 6,413 conformers from only 59 compounds. This may not cover all the Cartesian coordinates in the conformers of all available compounds, which suggest that some Cartesian coordinates from new conformers may not be included in our training set. Hence, we need to explore other approaches to extract conformer-specific features and increase the sample size. Furthermore, we did not use DL-based architecture to build our binary classifiers but used time efficient SVM method to build the classifiers. In our future work, we will explore to build the classifiers using a DL-based architecture.

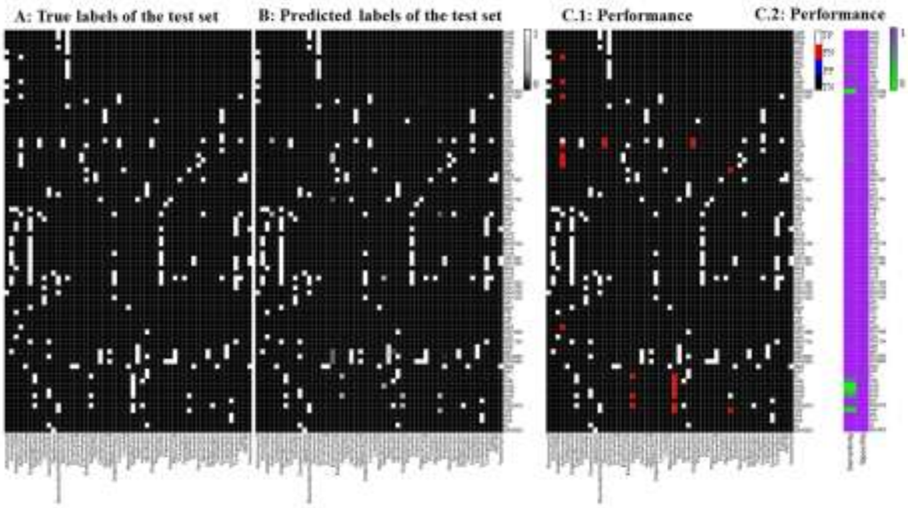


Figure 4: Machine learning-based prediction of chemogenomic profiles. Column names are 53 compounds, rows are 82 genes.
A) Heat map of true labels for the test set for 53 compounds and 82 genes with known depletion (white dots) and non-depletion (black dots) representing averaged test outputs.
B) Heat map of the prediction for the test set representing the averaged predicted outputs. The predicted results for the conformers were averaged into each compound (rows).
C.1) Prediction performance, TP, true positives; (TP), TN, true negatives; FN, false negatives; FP, false positives.
C.2) Consistency of test outputs and predict outputs
The raw data was extracted from Donald, R. G. K. *et al.* A *Staphylococcus aureus* Fitness Test Platform for Mechanism-Based Profiling of Antibacterial Compounds. *Chem. Biol.* 16, 826–836 (2009).

Acknowledgements

This work was supported in part by a Discovery Grant and a CREATE Graduate Training Program in Visual and Automated Disease Analytics (VADA) of the Natural Sciences and Engineering Research Council of Canada to PH, a Research Manitoba Mid Career Award to STC and a University of Manitoba Collaborative Research Grant to STC, PH and RD.

References

1. LeCun Y, Bengio Y & Hinton G. Deep learning. Nature 521, 436–444 (2015).
2. Miotto R, Wang F, Wang S, Jiang X & Dudley JT. Deep learning for healthcare: review, opportunities and challenges. Brief. Bioinform. https://doi. org/10.1093/bib/bbx044 (2017).
3. Zhou S, Greenspan H & Shen D. Deep Learning for Medical Image Analysis. (Academic Press, 2017).
4. Krizhevsky A, Sutskever I & Hinton G. ImageNet classification with deep convolutional neural networks. In Proc. Advances in Neural Information Processing Systems 25 1090–1098 (2012).
5. Eraqi HM, Moustafa MN, Honer J. End-to-end deep learning for steering autonomous vehicles considering temporal dependencies. arXiv preprint arXiv:1710.03804. 2017 Oct 10.
6. Sun X, Wu P, Hoi SC. Face detection using deep learning: An improved faster RCNN approach. Neurocomputing. 2018 Jul 19;299:42-50.
7. Wang Y, Bao T, Ding C, Zhu M. Face recognition in real-world surveillance videos with deep learning method. InImage, Vision and Computing (ICIVC), 2017 2nd International Conference on 2017 Jun 2 (pp. 239-243). IEEE.
8. Zhao ZQ, Zheng P, Xu ST, Wu X. Object detection with deep learning: A review. arXiv preprint arXiv:1807.05511. 2018 Jul 15.
9. Young T, Hazarika D, Poria S, Cambria E. Recent trends in deep learning based natural language processing. ieee Computational intelligenCe magazine. 2018 Aug;13(3):55-75.
10. Sainath T, Mohamed AR, Kingsbury B & Ramabhadran B. Deep convolutional neural networks for LVCSR. In Proc. Acoustics, Speech and Signal Processing 8614–8618 (2013).
11. Hinton G, Deng L, Yu D, Dahl GE, Mohamed AR, Jaitly N, Senior A, Vanhoucke V, Nguyen P, Sainath TN, Kingsbury B. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. IEEE Signal processing magazine. 2012 Nov;29(6):82-97.
12. Zhang Z, Geiger J, Pohjalainen J, Mousa AE, Jin W, Schuller B. Deep learning for environmentally robust speech recognition: An overview of recent developments. ACM Transactions on Intelligent Systems and Technology (TIST). 2018 Apr 24;9(5):49.
13. Veeraraghavan H. MO-A-207B-01: Radiomics: Segmentation & Feature Extraction Techniques. Medical physics. 2016 Jun;43(6Part29):3694-.
14. Ghafoorian M, Karssemeijer N, Heskes T, Uden IW, Sanchez CI, Litjens G, Leeuw FE, Ginneken B, Marchiori E, Platel B. Location sensitive deep convolutional neural networks for segmentation of white matter hyperintensities. Scientific Reports. 2017 Jul 11;7(1):5110.
15. Cheng JZ, Ni D, Chou YH, Qin J, Tiu CM, Chang YC, Huang CS, Shen D, Chen CM. Computer-aided diagnosis with deep learning architecture: applications to breast lesions in US images and pulmonary nodules in CT scans. Scientific reports. 2016 Apr 15;6:24454.
16. Litjens G, Kooi T, Bejnordi BE, Setio AA, Ciompi F, Ghafoorian M, van der Laak JA, Van Ginneken B, Sánchez CI. A survey on deep learning in medical image analysis. Medical image analysis. 2017 Dec 1;42:60-88.
17. Paul R, Hawkins SH, Balagurunathan Y, Schabath MB, Gillies RJ, Hall LO, Goldgof DB. Deep feature transfer learning in combination with traditional features predicts

survival among patients with lung adenocarcinoma. Tomography: a journal for imaging research. 2016 Dec;2(4):388.

18. Larochelle H, Bengio Y, Louradour J, Lamblin P. Exploring strategies for training deep neural networks. Journal of machine learning research. 2009;10(Jan):1-40.
19. Yang MC, Moon WK, Wang YC, Bae MS, Huang CS, Chen JH, Chang RF. Robust texture analysis using multi-resolution gray-scale invariant features for breast sonographic tumor diagnosis. IEEE transactions on medical imaging. 2013 Dec;32(12):2262-73.
20. Sun T, Zhang R, Wang J, Li X, Guo X. Computer-aided diagnosis for early-stage lung cancer based on longitudinal and balanced data. PloS one. 2013 May 15;8(5):e63559.
21. WHO AR. Global Report on Surveillance. Antimicrobial Resistance, Global Report on Surveillance. 2014.
22. Jackson N, Czaplewski L, Piddock LJ. Discovery and development of new antibacterial drugs: learning from experience?. Journal of Antimicrobial Chemotherapy. 2018 Feb 9;73(6):1452-9.
23. Bueso-Bordils JI, Alemán PA, Lahuerta Zamora L, Martin-Algarra R, J Duart M, M Antón-Fos G. Topological model for the search of new antibacterial drugs. 158 theoretical candidates. Current computer-aided drug design. 2015 Dec 1;11(4):336-45.
24. Donald RG, Skwish S, Forsyth RA, Anderson JW, Zhong T, Burns C, Lee S, Meng X, LoCastro L, Jarantow LW, Martin J. A Staphylococcus aureus fitness test platform for mechanism-based profiling of antibacterial compounds. Chemistry & biology. 2009 Aug 28;16(8):826-36.
25. Forsyth RA, Haselbeck RJ, Ohlsen KL, Yamamoto RT, Xu H, Trawick JD, Wall D, Wang L, Brown-Driver V, Froelich JM, King P. A genome-wide strategy for the identification of essential genes in *Staphylococcus aureus*. Molecular microbiology. 2002 Mar;43(6):1387-400.
26. Jia Y, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, Guadarrama S, Darrell T. Caffe: Convolutional architecture for fast feature embedding. InProceedings of the 22nd ACM international conference on Multimedia 2014 Nov 3 (pp. 675-678). ACM.
27. Furey TS, Cristianini N, Duffy N, Bednarski DW, Schummer M, Haussler D. Support vector machine classification and validation of cancer tissue samples using microarray expression data. Bioinformatics. 2000 Oct 1;16(10):906-14.
28. König C, Cruz-Barbosa R, Alquézar R, Vellido A. SVM-based classification of class C GPCRs from alignment-free physicochemical transformations of their sequences. InInternational Conference on Image Analysis and Processing 2013 Sep 9 (pp. 336-343). Springer, Berlin, Heidelberg.
29. van de Wolfshaar J, Karaaba MF, Wiering MA. Deep convolutional neural networks and support vector machines for gender recognition. InComputational Intelligence, 2015 IEEE Symposium Series on 2015 Dec 7 (pp. 188-195). IEEE.
30. Meyer D, Dimitriadou E, Hornik K, Weingessel A, Leisch F. e1071: misc functions of the department of statistics, probability theory group (formerly: E1071), TU Wien. R package version 1.6-7.