Deep Belief Networks and Bayesian Networks for Prognosis of Acute Lymphoblastic Leukemia

Fakhirah D. Ghaisani
Faculty of Computer Science,
Universitas Indonesia
Depok, Indonesia 16424
fakhirah.dianah61@ui.ac.id

Ito Wasito*
Faculty of Computer Science,
Universitas Indonesia
Depok, Indonesia 16424
ito.wasito@cs.ui.ac.id

Moh. Faturrahman
Faculty of Computer Science,
Universitas Indonesia
Depok, Indonesia 16424
moh.faturrahman@ui.ac.id

Ratna Mufidah
Faculty of Computer Science,
Universitas Indonesia
Depok, Indonesia 16424
ratna.mufidah51@ui.ac.id

** Corresponding Author*

ABSTRACT
Cancer is one of main non-communicable diseases. Acute Lymphoblastic Leukemia (ALL), a type of white blood cancer, is one of the most common pediatric cancers. Analysis of cancer prognosis is necessary to determine the proper treatment for each patient. However, cancer data analysis is challenging because multiple risk factors may influence the prognosis of cancer, including gene and clinical condition of patient. This study aims to develop prediction model for cancer prognosis using clinical and gene expression (microarray) data. In this research, manifold learning is applied to microarray data to reduce its dimension, then two Deep Belief Network (DBN) models for both clinical and microarray data are trained separately. Probabilities obtained from Clinical DBN model and Microarray DBN model are integrated using softmax nodes on Bayesian Network structure. Based on various experiments, the best integration model obtained is DBN+BN 32 with prediction accuracy 84.2% for 2-years survival, 70.2% for 3-years, 68.4% for 4-years, and 73.7% for 5-years. This prediction model can be used in cancer analysis and help doctor to decide proper treatment for patient.

CCS Concepts
• Applied Computing → Life and medical sciences → Bioinformatics

Keywords
Cancer; leukemia; acute lymphoblastic leukemia; manifold learning; dimensionality reduction; deep belief network; bayesian network; data integration; microarray.

1. INTRODUCTION
Cancer is one of main non-communicable diseases (NCD) together with cardiovascular disease, chronic respiratory disease, and diabetes. This disease causes approximately 8.2 million human deaths in the world each year [1]. Acute Lymphoblastic Leukemia (ALL) is a type of leukemia (blood cancer) which is one of the most common pediatric cancers, contributing approximately 25% of cancer diagnoses among children younger than 15 years old. [2].

ICACS '17, August 10–13, 2017, Jeju Island, Republic of Korea

ACM ISBN 978-1-4503-5284-0/17/08…$15.00
https://doi.org/10.1145/3127942.3127947

Analysis of cancer prognosis is necessary to determine the proper treatment for each patient. However, cancer data analysis is challenging because multiple risk factors may influence cancer prognosis, including gene, clinical condition of patient, and cancer stage. Previous cancer studies have successfully collected an enormous amount of cancer patient data [4]. Applying machine learning techniques, these data can be used to develop prediction model for cancer prognosis. This model can be used to predict cancer progression on patient, such as recurrence, distant metastasis, survival, and benefit of certain treatment given to the patient [3]. Therefore, it can help doctor to diagnose and decide proper treatment for cancer patient.

Most previous works in disease diagnosis have used only patient clinical data [4-6]. Meanwhile for cancer diagnosis, some studies use microarray data [4] or both clinical and microarray data [3, 8], considering cancer is a genetical disease [3]. Challenge in clinical and microarray data analysis is high-dimensional data (particularly microarray data) compared to number of samples. Thus, the number of variables are much larger compared to the number of equations. Besides, data integration method is needed to combine information from clinical and microarray data which have different characteristics.

Bayesian Networks (BN) has been used for data integration in some previous researches. Gevaert et al. (2006) [8] integrate clinical and microarray data with experiment three types of integration method (full/early, partial/intermediate, and decision/late integration) using BN for prognosis of cancer. Experiment results in [8] show that late and intermediate integration outperforms early integration technique. In other work, Khademi and Nedialkov (2015) [3] use late integration method for prognosis of cancer. In [3], clinical and microarray data are trained separately, then the two models obtained (clinical model and microarray model) are integrated using softmax nodes on BN structure. In that study, clinical model is constructed using BN, meanwhile microarray model is constructed using Deep Belief Networks (DBN). Experiment results in [3] using NKI dataset show that DBN and BN integration model outperforms Clinical BN model. However, structure and parameter learning in BN is

challenging, especially when using more complex data. This problem can be addressed by reducing complexity of the data. Gevaert et al. (2006) [8] applied feature selection method to microarray data to select some genes. But, this gene selection technique may lead to loss of important genetic information for cancer prognosis.

In this study, Deep Learning method, specifically DBN is utilized to develop prediction model using clinical data and microarray data. These two DBN models for both clinical and microarray data are trained separately. Then, probabilities obtained from Clinical DBN model and Microarray DBN model are integrated using softmax nodes on BN structure [3]. Deep Learning method is widely used for classification and clustering tasks, especially when complex and large-scale data is used. Some previous works also have been used DBN to predict disease prognosis [3, 5-7].

The aim of this research is to construct classification model using DBN and BN to integrate clinical and gene expression data of cancer patient. Thus, this model can be used for prognosis of cancer, particularly ALL.

2. METHODOLOGY

2.1 Manifold Learning

The most common dimensionality reduction method is Principal Component Analysis (PCA). PCA is a linear dimensionality reduction method, assuming high-dimensionality data can be represented as a linear field. This property of PCA can cause loss of useful information on the reduced data.

To overcome this problem, a non-linear dimensionality reduction method named manifold learning is used. This method is based on assumption that most of data are only artificially high-dimensional. In other words, data with thousands of features (high-dimension) can be represented as a function of some parameters (with lower dimension). In manifold learning, all data points (samples) form low-dimensional manifold in high-dimensional space [9]. One of common algorithms in manifold learning is Isometric Feature Mapping (Isomap). Further and detailed explanation about manifold learning and Isomap can be found in [9]. In this paper, we use manifold learning method to reduce microarray data dimension.

2.2 Deep Belief Networks

Deep Belief Networks (DBN) is one of deep learning model. DBN is a graphical generative model which consists of Restricted Boltzmann Machine (RBM) on the top two layers and Sigmoid Belief Networks (SBN) on other layers below. The top two layers are connected indirectly and symmetrically forming associative memory, other layers below are connected top-down (directed), while the bottom layer represents data vector [10].

Learning process in DBN starts with unsupervised pre-training on RBM until equilibrium sample is reached. There is a fast RBM training algorithm named contrastive divergence. The result of pre-training forwarded to the next layer, until states of each layers obtained. This process is a generative model of DBN. Further and detailed explanation about DBN can be found in [10]. In this paper, DBN is utilized to construct prediction model with clinical and microarray data.

2.3 Bayesian Networks

Bayesian Networks is a probabilistic graphical model which represents random variables and conditional dependencies. Its structure is Directed Acyclic Graph (DAG). Each node Xi (random variable) has Conditional Probability Distribution (CPD) P(Xi|parent(Xi)), which defines probability of certain node given its parent. The most common CPD used in BN are Table CPD, Gaussian CPD, and Softmax CPD. Table CPD is used for discrete node and discrete parent, Gaussian CPD is used for continuous node and discrete and continuous parent, and Softmax CPD can be used to represent discrete node with continuous parent [3]. In this paper, we use BN with Softmax CPD to integrate probability obtained from Clinical DBN and Microarray DBN.

3. PROPOSED METHOD

In this paper, we construct two prediction models using clinical data and gene expression data of cancer patient separately utilizing DBN. Each model will produce probabilities for classification. Then, these two models will be integrated using softmax nodes on Bayesian Networks structure [3], thus probabilities for classification from two different source of information obtained (from clinical and gene expression data). In general, framework of algorithm used in this research is given in Figure 1. This framework is proposed by Khademi et al. Further and more detailed explanation can be found in [3]. In previous work, Khademi et al. construct clinical prediction model using BN, meanwhile in this paper we use DBN.

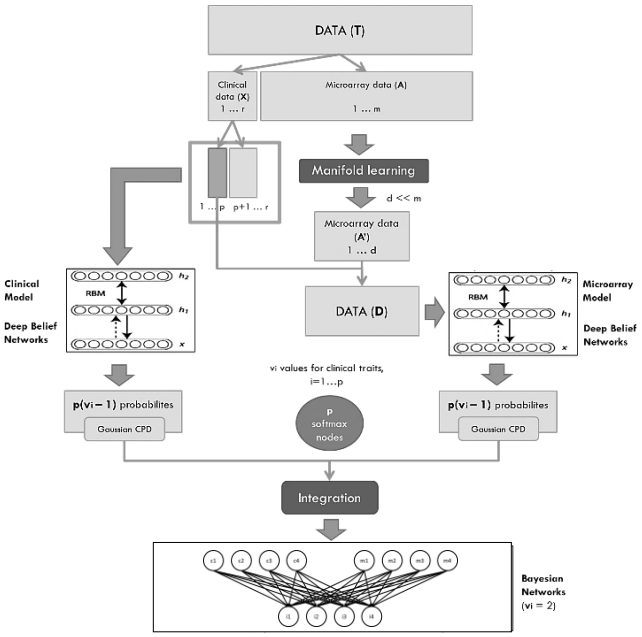

Figure 1. DBN and BN model construction.

3.1 Dimensionality Reduction with Manifold Learning

Microarray data is a high-dimensional data (54676 features in TARGET ALL Phase II Dataset) thus applying dimensionality reduction technique is required. We use Isometric Feature Mapping algorithm, utilize Subspace Learning Library on Matlab developed by Cai et al. [13]. The result of dimensionality reduction is microarray data with dimension d << 54676.

3.2 Deep Belief Networks (DBN) Model Construction

Two DBN models are constructed, Clinical DBN and Microarray DBN. 41 (out of 44) clinical attributes from ALL dataset are used as the input to Clinical DBN model. Meanwhile, d dimension

microarray data (result of dimensionality reduction) as the input to Microarray DBN model. Then, each DBN model is trained separately. Many experiments are conducted with variation in hyper-parameters. In the last layer of each DBN, there are 4 nodes as many as the number of clinical attributes we want to predict. Performance of each DBN model will be evaluated to get the best hyper-parameters setting for both clinical and microarray data.

DBN used consists of two RBMs with 40 nodes in each hidden layer, as in [3]. The first RBM is a probability-probability RBM, whereas the second RBM is a probability-gaussian RBM. First, data is normalized using Minmax method (subtracted by minimum value, then divided by the difference of maximum and minimum values). We use learning rate 0.001 and RBM sampling method Free Energy Persistent Contrastive Divergence (FEPCD) [14] which trained with 2000 epochs. Two experiment scenarios to find the best DBN model are conducted. The first scenario is mini-batch size experiment on DBN training. The mini-batch sizes used are 10, 50, and 100. The second scenario is number of epochs experiment on DBN fine-tuning. The method used in DBN fine-tuning process is backpropagation on Multi-Layer Perceptron (MLP), while the number of epochs used are 40, 200, 800.

3.3 Model Integration with Bayesian Networks (BN)

After the best parameter settings and models for Clinical DBN and Microarray DBN have been obtained, probability for each clinical outcome can be obtained. In this paper, 8 probabilities are obtained. These probabilities can be represented as continuous nodes with Gaussian CPD on BN. Then, these nodes are integrated using 4 softmax nodes. As results of these softmax nodes, probabilities can be used to classify patient.

BN structure used in this research is a BN with 32 edges (BN-32) which connect all continuous nodes (result of DBN models) with all softmax nodes. Illustration of BN-32 is given in Figure 2. Node cj shows probability of attribute j has the value 1 based on Clinical DBN model, node mj shows the probability of attribute j has the value 1 based on Microarray DBN model, while node ij represents probability of attribute j has the value 1 based on BN-32 integration model. The edges in this BN structure show that the value of attribute j depends on probabilities of all attributes obtained from Clinical DBN and Microarray DBN.

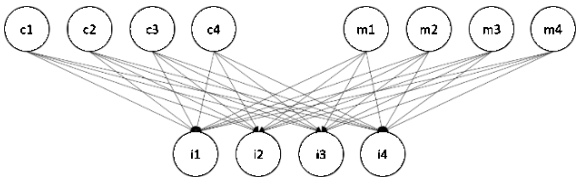

Figure 2. Structure of BN-32.

4. EXPERIMENT

We conduct experiments to find the best DBN and BN (DBN+BN) models for prognosis of ALL. Furthermore, we compare classification performance of the best DBN+BN model with other classification methods, such as k-Nearest Neighbor (k-NN) and Support Vector Machine (SVM). Further explanation of the research process is described below.

4.1 Data

In this study, we use childhood Acute Lymphoblastic Leukemia (ALL) patient data. We use TARGET ALL Phase II dataset, which is the part of TARGET (Therapeutically Applicable Research to Generate Effective Treatments) project by Office of Cancer Genomics (OCG), National Cancer Institute, National Institute of Health [11]. There are 301 patients, 44 clinical attributes, and 54676 gene expression attributes in this dataset. This dataset is available online and can be accessed freely on https://ocg.cancer.gov/programs/target/data-matrix [12]. Proportion of class member for all attributes/traits we want to predict given in Table 1.

Table 1. Proportion of class member in ALL data

Class	Data			
	2-years	3-years	4-years	5-years
0 (Dead)	90 (30%)	141 (47%)	178 (59.33%)	198 (66%)
1 (Alive)	210 (70%)	159 (53%)	122 (40.67%)	102 (34%)

4.2 Evaluation

Evaluation is done by applying 10-fold cross validation, then quantitative performance is measured by calculating classification accuracy of each prediction model. Classification accuracy is defined as

$$Accuracy = \frac{\#correctly\_classified\_instances}{size\_of\_data}.$$

5. RESULT

5.1 Mini-Batch Size Experiment on DBN Training

Table 2 presents average and standard deviation of prediction accuracy for Clinical and Microarray DBNs with mini-batch size 10, 50, and 100. Table 2 shows that range of prediction accuracy value for each mini-batch size is overlapping, thus this result is inconclusive. Then, performance of each model is evaluated using head-to-head comparison as summarized in Table 3.

Table 2. Prediction accuracy of Clinical DBN and Microarray DBN with mini-batch size 10, 50, 100

Classifier	Mini-Batch Size	2-years	3-years	4-years	5-years
Clinical DBN	10	**0.824** ± 0.018	**0.760** ± 0.007	**0.753** ± 0.013	**0.775** ± 0.019
	50	**0.841** ± 0.017	**0.783** ± 0.010	**0.791** ± 0.012	**0.804** ± 0.017
	100	**0.844** ± 0.014	**0.775** ± 0.020	**0.780** ± 0.016	**0.798** ± 0.017
Microarray DBN	10	**0.769** ± 0.013	**0.667** ± 0.020	**0.609** ± 0.023	**0.682** ± 0.015
	50	**0.771** ± 0.021	**0.628** ± 0.035	**0.576** ± 0.023	**0.665** ± 0.029
	100	**0.774** ± 0.016	**0.604** ± 0.025	**0.602** ± 0.024	**0.683** ± 0.016

Based on mini-batch size variation used in Clinical DBN, the lowest accuracy is obtained by training with mini-batch size 10. It is because 10 instances of clinical data are unrepresentative of the overall clinical data training which consist of 270 instances/records. Then, the accuracy increased as the mini-batch size is increased to 50, which leads to the highest accuracy. Besides, the accuracy decreased when the mini-batch size is

increased to 100. It is because for clinical data training we used, mini-batch size 100 is too large, thus it may cause overfitting. Meanwhile in Microarray DBN training, the highest accuracy is obtained by mini-batch size 10. Moreover, experiment result shows that the mini-batch size 10 well-performed in predicting value for attributes whose proportion of class member is balanced (3-years and 4-years survival).

The different result for Clinical DBN and Microarray DBN is caused by different type and characteristics of clinical and microarray data that are used. Clinical data consists of binary values, whereas the microarray (gene expression) data is a normalized Gaussian data whose values are more varied than clinical data. Besides, the different number of attributes in clinical data (36 attributes) and reduced microarray data (299 attributes) can take effect on mini-batch size setting as well.

Table 3. Summary of head-to head comparison applied. 10 experiments are conducted for each mini-batch size setting. Then, accuracy in each experiment are compared and proportion of winning (has higher accuracy) frequency are measured. In this table, average of winning proportion to predict 2-years survival with mini-batch size 10 (config-1) values 0.1 (10%). This means among 10 experiments, config-1 outperforms config-1 0 time (0%), outperforms config-2 (mini-batch size 50) 2 times (20%), and outperforms config-3 (mini-batch size 100) once or 1 time (10%). Therefore, those percentage are averaged, obtains 10% (0.1)

Classifier	Mini-Batch Size	2-years	3-years	4-years	5-years	Avg
Clinical DBN	**10**	0.10	0.10	0.00	0.07	**0.07**
	50	0.37	0.53	0.63	0.53	**0.52**
	100	0.43	0.37	0.37	0.37	**0.38**
Microarray DBN	**10**	0.17	0.63	0.43	0.33	**0.39**
	50	0.40	0.23	0.13	0.23	**0.25**
	100	0.33	0.33	0.43	0.40	**0.32**

5.2 DBN Fine-tuning Epoch Experiment

Table 4 presents average and standard deviation of prediction accuracy for Clinical and Microarray DBNs with fine-tuning epoch 40, 200, and 800. However, result in Table 4 is also inconclusive. Table 5 gives summary of head-to-head comparison for each fine-tuning epoch variation.

The result shows that fine-tuning epoch which results the highest accuracy is 200 for Clinical DBN and 800 for Microarray DBN. As the number of epoch increased, training error is decreased. However, too large number of epoch may cause overfitting [15] which thus increases the testing error.

In Clinical DBN training, the highest accuracy is obtained by 200 fine-tuning epochs. Meanwhile, number of epoch 40 results the lowest accuracy because the model has not learned the data well. Fine-tuning with 800 epochs also performed well (slightly lower than accuracy of 200 epochs based on experiments). However, we choose 200 as the best-performed fine-tuning epoch for Clinical DBN instead of 800 for the sake of efficiency.

Table 4. Prediction accuracy of Clinical DBN and Microarray DBN with number of fine-tuning epoch 40, 200, 800

Classifier	#Fine-tuning Epoch	2-years	3-years	4-years	5-years
Clinical DBN	**40**	**0.825** ±0.010	**0.751** ±0.014	**0.746** ±0.015	**0.765** ±0.023
	200	**0.841** ±0.017	**0.783** ±0.010	**0.791** ±0.012	**0.804** ±0.017
	800	**0.836** ±0.010	**0.777** ±0.014	**0.790** ±0.018	**0.801** ±0.016
Microarray DBN	**40**	**0.731** ±0.016	**0.653** ±0.030	**0.502** ±0.027	**0.600** ±0.022
	200	**0.769** ±0.013	**0.667** ±0.020	**0.609** ±0.023	**0.682** ±0.015
	800	**0.776** ±0.013	**0.686** ±0.033	**0.630** ±0.033	**0.727** ±0.031

Table 5. Summary of head-to head comparison applied. See caption on Table 3.

Classifier	#Fine-tuning Epoch	2-years	3-years	4-years	5-years	Avg
Clinical DBN	**40**	0.10	0.00	0.03	0.03	**0.04**
	200	0.47	0.50	0.47	0.47	**0.48**
	800	0.40	0.47	0.47	0.43	**0.44**
Microarray DBN	**40**	0.00	0.10	0.00	0.00	**0.03**
	200	0.43	0.37	0.47	0.33	**0.40**
	800	0.53	0.53	0.53	0.67	**0.57**

5.3 Method Comparison

Table 6 gives average and standard deviation of prediction accuracies for 2-years, 3-years, 4-years, and 5-years survival.

Table 6. Classification accuracy

Survival	Clinical DBN	Microarray DBN	DBN + BN-32	SVM	k-NN
2-years	**0.836** ±0.010	**0.776** ±0.013	**0.842** ±0.012	**0.733** ±0.009	**0.700** ±0.028
3-years	**0.777** ±0.014	**0.686** ±0.033	**0.702** ±0.017	**0.759** ±0.007	**0.690** ±0.023
4-years	**0.790** ±0.018	**0.630** ±0.033	**0.684** ±0.021	**0.708** ±0.007	**0.617** ±0.033
5-years	**0.801** ±0.016	**0.727** ±0.031	**0.737** ±0.021	**0.756** ±0.008	**0.689** ±0.031

In Clinical DBN, the highest accuracy obtained is to predict 2-years, 5-years, 4-years, and 3-years survival, respectively. It is because of the proportion of class member in each attribute/clinical trait. Proportion of dead patient (class 0) and alive (class 1) can be seen in Table 1. As showed that Clinical DBN is best in predicting 2-years survival (unbalanced data) while worst in predicting 3-years survival (balanced data). Then,

sensitivity and specificity score are measured to evaluate the ability of each model in predicting each class (0 and 1). In 2-years survival prediction, the sensitivity is 0.909738, while its specificity is 0.67082. This means Clinical DBN model constructed can predict positive class (1) well, but still not good in predicting negative class (0). For other attributes/traits (3, 4, 5-years survival), the difference of sensitivity and specificity score is not large. The DBN Microarray also good at predicting majority class for each attribute/trait.

Classification performance of DBN+BN 32 depends on performance of Clinical DBN and Microarray DBN. The best performance obtained is in 2-years survival prediction, because Clinical DBN and Microarray DBN also perform well. In 2-years survival prediction, the accuracy of DBN + BN 32 integration model higher than Clinical DBN and Microarray DBN. Meanwhile for predictions of other traits, the performance of DBN + BN 32 model is worse than Clinical DBN and Microarray DBN. It may be happened when Clinical DBN and Microarray DBN give different prediction result.

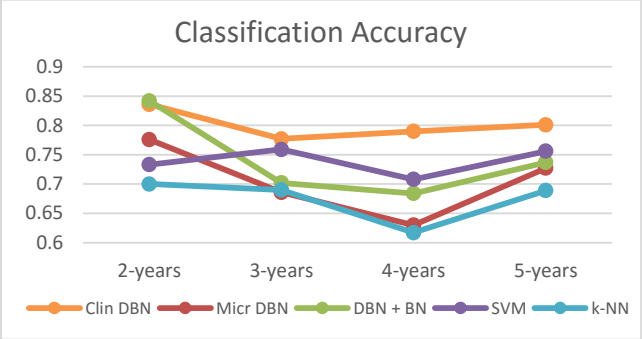

Figure 3. Classification Accuracy.

Based on experiment results, overall the best method to predict survival of ALL patient on dataset used is Clinical DBN, which is constructed as the result of clinical data training.

6. CONCLUSION AND FUTURE WORK

This research shows that Deep Belief Networks (DBN) and Bayesian Networks (BN) can be used in survival prediction and classification of Acute Lymphoblastic Leukemia (ALL) patient. Besides, BN can be used to integrate two prediction models obtained from two different data. In this paper, clinical and gene expression (microarray) data are used.

Based on experiments with various hyperparameter settings on Clinical DBN, the best performance obtained by 36-40-40-2 DBN trained with mini-batch size 50, and fine-tuning epoch 200. Meanwhile for Microarray DBN, the best-performed model obtained is a 299-40-40-2 DBN trained with mini-batch size 10 and fine-tuning epoch 800. The best-performed integration model obtained is DBN+BN 32 with prediction accuracy 84.2% for 2-years survival, 70.2% for 3-years survival, 68.4% for 4-years survival, and 73.7% for 5-years survival.

In this paper, the best prediction model obtained is Clinical DBN, followed by SVM, DBN+BN 32, Microarray DBN, and k-NN. Basically, hyperparameter settings on DBN (and neural network in general) to obtain the best prediction model is data-dependent. Thus, the best prediction model can be obtained based on experiments. For future research, more experiments on hyperparameter settings and architectures of DBN and BN are needed to be done. Other research direction is to conduct experiments on various microarray data pre-processing methods. Besides, more patient records and clinical attributes may increase the ability of DBN and BN to predict ALL patient survival.

7. REFERENCES

[1] World Health Organization (WHO). Fact Sheet: Noncommunicable Diseases. 2015. http://www.who.int/mediacentre/factsheets/fs355/en/, , accessed: October 17 '16.

[2] National Cancer Institute. Childhood Acute Lymphoblastic Leukemia Treatment (PDQ) – Health Professional Version. 2016. https://www.cancer.gov/types/leukemia/hp/child-all-treatment-pdq, accessed: October 17 '16.

[3] Khademi, M. and Nedialkov, N. S. 2015. Probabilistic graphical models and deep belief networks for prognosis of breast cancer. *IEEE 14th International Conference on Machine Learning and Applications*.

[4] Fakoor R., Huber, M., Ladhak, F., and Nazi, A. 2013. Using deep learning to enhance cancer diagnosis and classification. *Proceedings of the 30th International Conference on Machine Learning*, Atlanta, Georgia, USA. JMLR: W&CP Vol.28.

[5] Abdel-Zaher, A. M. and Eldeib, A. M. 2015. Breast cancer classification using deep belief networks. *Expert System with Applications* 46, 139-144.

[6] Boxtel, M. V., Donoghue, J. O., and Roantree, M. 2015. A configurable deep network for high-dimensional clinical trial data. *IEEE International Joint Conference on Neural Networks (IJCNN)*. Doi: 10.1109/IJCNN.2015.7280841.

[7] Li, H., Li, X., Ramanathan, M., and Zhang, A. 2014. Identifying informative risk factors and predicting bone disease progression via deep belief networks. *Methods* 69, 257-265.

[8] Gevaert, O., et al. 2006. Predicting the prognosis of breast cancer by integrating clinical and microarray data with Bayesian networks. *Bioinformatics Oxford Journals* 22, 14, e184-e190. Oxford University Press. doi:10.1093/bioinformatics/btl230.

[9] Cayton, L. 2005. *Algorithms for Manifold Learning*. Technical Report, pp.1-17. University of California at San Diego.

[10] Hinton, G. E, Osindero, S., and Teh, Y. W. 2006. A fast learning algorithm for deep belief nets. *Neural Computation* 18, 1527-1554.

[11] Ma, X., et al. 2015. Rise and fall of subclones from diagnosis to relapse in pediatric B-acute lymphoblastic leukaemia. *Nature Communication* 6, 6604. doi:10.1038/ncomms7604.

[12] Office of Cancer Genomics (OCG), National Cancer Institute, National Institute of Health. Target Data Matrix. 2016. https://ocg.cancer.gov/programs/target/data-matrix, accessed: October 17 '16.

[13] Cai, D., He, X., and Han, J. 2007. Isometric Projection. *Proc. 2007 AAAI Conference on Artificial Intelligence*. AAAI'07.

[14] Keyvanrad, M. A. and Homayounpour, M. M. 2014. *A Brief Survey on Deep Belief Networks and Introducing A New Object Oriented Toolbox (DeeBNet)*. Technical Report. Cornell University. arXiv:1408.3264v7

[15] Bengio, Y. 2012. Practical Recommendations for Gradient-Based Training of Deep Architectures. *Neural Networks: Tricks of the Trade 2012* 7700, 437-478.