Deep Learning Causal Attributions of Breast Cancer

Daqing Chen[1], Laureta Hajderanj[1], Sarah Mallet[2], Pierre Camenen[2],

Bo Li[3], Hao Ren[3] and Erlong Zhao[4]

[1] School of Engineering, London South Bank University, London SE1 0AA, UK

[2] Electronics and Digital Technologies Department, Polytech Nantes, 44300 Nantes, France

[3] School of Informatics and Electronics, Northwestern Polytechnical University, Xi'an 710072, China

[4] School of Information Science and Technology, Northwest University, Xian 710127, China

Abstract. In this paper, a deep learning-based approach is applied to high dimensional, high-volume, and high-sparsity medical data to identify critical casual attributions that might affect the survival of a breast cancer patient. The Surveillance Epidemiology and End Results (SEER) breast cancer data is explored in this study. The SEER data set contains accumulated patient-level and treatment-level information, such as cancer site, cancer stage, treatment received, and cause of death. Restricted Boltzmann machines (RBMs) are proposed for dimensionality reduction in the analysis. RBM is a popular paradigm of deep learning networks and can be used to extract features from a given data set and transform data in a non-linear manner into a lower dimensional space for further modelling. In this study, a group of RBMs has been trained to sequentially transform the original data into a very low dimensional space, and then the *k*-means clustering is conducted in this space. Furthermore, the results obtained about the cluster membership of the data samples are mapped back to the original sample space for interpretation and insight creation. The analysis has demonstrated that essential features relating to breast cancer survival can be effectively extracted and brought forward into a much lower dimensional space formed by RBMs.

Keywords: Restricted Boltzmann Machines, Deep Learning, Survival Analysis, *k*-means Clustering Analysis, Principal Component Analysis.

1 Introduction

Breast cancer is the most diagnosed cancer in women, affecting over 2.1 million women each year globally, and it causes the greatest number of cancer-related deaths among women. In 2018, it is estimated that 627,000 women died from breast cancer – that is approximately 15 of all cancer deaths among women [1].

In order to improve breast cancer survival and life expectancy, it is crucial to learn and understand the factors that might affect breast cancer survival rate and outcomes following certain treatments. In the past years, enormous studies have been undertaken intensively in this area aiming to identify the causal attributions of breast cancer survival from multiple perspectives including biological, diagnostical, and data mining techniques.

From an analytical point of view, analysing breast cancer data has been challenged by a) The volume of the data to be explored tends to be quite big since it has been accumulated for several decades. For example, the SEER (Surveillance Epidemiology and End Results) data set [2] contains some 291,760 breast cancer incidences collected from 1974 up to date; and b) The number of variables (features) of the data usually is considerably high, e.g., over a thousand or even more. The high dimensionality is mainly caused by the categorical variables contained in the data that may have many distinct symbolic values, since each of the distinct values of a categorical variable needs to be transformed into a unit vector using, for example, the one-hot encoding (orthogonal encoding) method to make a categorical variable applicable for an algorithm. Due to these factors, some typical data mining algorithms, such as the *k*-means clustering algorithm, may not perform well with high dimensionality, high volume, and high sparsity data.

The *k*-means clustering algorithm is a widely used unsupervised descriptive modelling approach for grouping (segmenting) a given data set based on similarities among the data samples with respect to the values taken by certain variables involved. The algorithm is simple, effective in general, and can converge within a few iterations. However, the algorithm has several problems when applied to a data set with high dimensionality and high sparsity:

1. The typical Euclidean distance for similarity measure is inefficient when the number of variables is large, and the number of samples is relevantly small [3];
2. The computational complexity of the algorithm increases with the number of dimensions [4]; and
3. It is difficult to determine the cluster centroids if the data values are sparse, i.e., only a small number of data entries having a non-null value. An example of such a data is the resultant data set transformed from a data that contains many categorical variables using the one-hot encoding method. This can also be viewed as an asymmetry data matrix. In addition, sparsity has made the algorithm very sensitive to noise.

To address the high dimensionality problem involved in the *k*-means clustering, a proper dimensionality reduction approach is usually considered. The principal component analysis (PCA) is a popular approach for such a purpose. PCA forms a linear transformation to transform the original data into a new space spanned by a set of principal components. Depending on the significance of each principal component, only a few significant principal components could be selected to form a subspace with a low dimensionality, and then the *k*-means clustering can be performed in the subspace. The PCA-based subspace clustering approach has been applied to medical image segmentation [5].

It should be noted that data samples may become insufficient when dealing with high dimensional data since it may lead a modelling process that involves too many parameters to learn and/or to optimize with relatively a very small number of samples.

In this paper, a deep learning-based approach is applied to identify critical casual attributions that might affect the survival of a breast cancer patient. The SEER breast cancer data is explored in this study. The data set contains rich patient-level and treatment-level information on breast cancer incidences, such as cancer site, cancer stage, treatment received, and cause of death. Restricted Boltzmann machines (RBMs) are proposed for dimensionality reduction in the present analysis. RBM is a popular paradigm of deep learning networks and can be used to extract features from a given data set. RBMs transform data in a non-linear manner into a lower dimensional space for further modelling, for instance, clustering analysis and classification. In this study, a group of RBMs has been trained to sequentially transform the original data into a very low dimensional space, and then the *k*-means clustering is conducted in this space. Furthermore, the results obtained about the cluster membership of the data samples are mapped back to the original space for interpretation and insight creation. The analysis has demonstrated that essential features in the data set can be effectively extracted and brought forward into a much lower dimensionality space formed by RBMs.

The remainder of this paper is organized as follows. Section 2 provides a literature review on the relevant works in relation to diagnosis analytics in breast cancer data. Section 3 discusses in detail the methodology adopted in this study including the entire analytical process and RBMs. The SEER data set is described in Section 4 along with the essential data pre-processing performed on the data. Section 5 gives a detailed account about the analysis experiments, and further in Section 6, the findings from the analysis are interpreted and summarized. Finally, in Section 7, concluding remarks are discussed and the further research is outlined.

2 Relevant Works

Data mining techniques have been widely used in medical research, and especially, in medical diagnosis of diverse types of cancer. Various models have been developed for this purpose including qualitative models [6][7][8][9], quantitative models [10][11], and hybrid models [12]. Very recently, many deep learning-based models have been considered in several case studies [13].

Segmenting breast cancer patients was studied in [6] and it was found that the number and the types of the resultant clusters were similar in terms of symptom occurrence rates or symptom severity ratings. Five clusters were identified using symptom occurrence rates while six clusters were established using symptom severity ratings. The types of clusters were also similar. A bisecting *k*-means algorithm was applied to analyze three diseases: breast cancer, Type 1 diabetes, and fibromyalgia in [7]. Their results showed that, although the clusters established were different from each other, all the clusters had several common features. In [8] the time effect and symptom for patients who received chemotherapy was examined using clustering analysis. An ensemble learning-based algorithm for lung cancer diagnosis was

investigated [14]. The algorithm can achieve a high classification accuracy with a low false no-cancer rate (false negative rate). The model contained two levels. The first level consisted of a group of individual neural networks which can be used to identify if a cell is a cancer cell or not. A cell was considered a cancer cell if any network classified it a cancer cell. The second level employed a group of neural networks to determine which type of cancer a recognized cancer cell belongs to. This algorithm has implemented a certain classification scheme. A Bayesian network structure was proposed for canner diagnosis purpose [10]. The network can be trained using a direct causal learner algorithm and has been applied to both simulated and real data sets.

In addition, algorithms for dimensionality reduction and feature extraction have also been considered. A two-step algorithm was investigated to address the high dimensionality problem in genes analysis [15]. Interestingly in [16] a statistical test and genetic algorithm was utilized for feature selection, and further leave-one-out cross validation was used along with receiver operating characteristic curve to identify which features to be used in order to achieve the best classification performance. Several real-life data sets were chosen to testify and evaluate the effectiveness of the algorithm.

Comparing different data mining algorithms for a better cancer risk or survival rate prediction and classification has received a great research attention. In [12] various Bayesian-based classifiers were explored for the prediction of the survival rate of 6 months after treatment, including naive Bayesian classifier, selective naive Bayesian classifier, semi-naive Bayesian classifier, tree-augmented naive Bayesian classifier, and *k*-dependence Bayesian classifier. The performance of all these classifiers were evaluated and compared. Artificial neural networks (ANNs) were employed to examine mammograms [11]. The work has demonstrated that ANNs with two hidden layers performed better than ANNs with only one hidden layer. In [17] logistic regression, ANNs, and Bayesian networks were compared for accuracy. Their results showed that the Bayesian model outperformed other methods.

Sensible new knowledge can be discovered when applying data mining algorithms in medicine. Yang [18] proposed a vicinal support vector classifier to handle data from different probability distributions. The proposed method had two steps: clustering and training. In the first step, a supervised kernel-based deterministic annealing clustering algorithm was applied to partition the train data into different soft vicinal areas in the feature space. By doing so, they constructed vicinal kernel functions. In the training step, the objective function, called vicinal risk function, was minimized under the constraints of the vicinal areas defined in the clustering step. Kakushadze [19] applied *k*-means to cluster different types of cancer with genome data without using nonnegative matrix factorization. They found that, out of 14 types of cancer, three had no cluster-like structures, two had high within-cluster correlations, and the others had common structure.

In conclusion, many data mining technologies such as neural networks and Bayesian networks are applied to construct cancer diagnosis models. The constructed models can diagnose various cancer and have shown a high accuracy. On the other hand, there is still a clear lack of dealing with high-dimensional and high-sparsity medical data.

3 Methodology

In this paper a deep learning-based approach is proposed for feature extract and dimensionality reduction in order to identify crucial casual attributions with high dimensional, high-volume, and high-sparsity breast cancer data. The entire analysis process is illustrated conceptually in Figure 1 and the key techniques applied are discussed below.

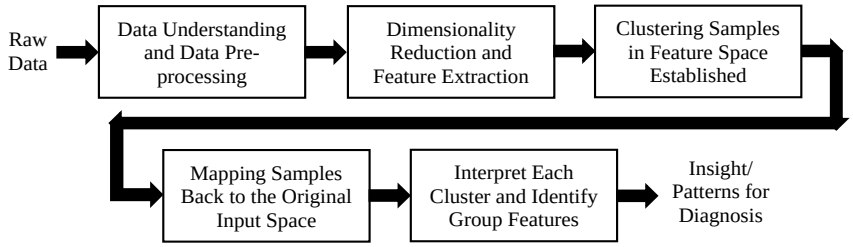

Fig. 1. The analysis process with the key steps.

3.1 Restricted Boltzmann Machines for Dimensionality Reduction and Feature Extraction

RBMs are used in this research for feature extraction and dimensionality reduction. Using an RBM, data from an original high-dimensional space can be transformed into a feature space with a much smaller number of dimensions for analysis, such as the *k*-means clustering. It is our intention in this study to examine if this approach can effectively address the issues relating to cluster analysis in a high-dimensional and highly-sparse space in order to partition the breast cancer patients into various meaningful groups based on their similarities in relation to a set of features and measures.

A typical topology of RBM is shown in Figure 2. An RBM is an energy-based generative statistical model with hidden and visible variables [21]. The energy of the visible node state and hidden node state is defined as

$$E(v,h) = -\sum_{i \in visible} b_i v_i - \sum_{i \in hidden} b_i v_i - \sum_{i,j} v_i h_j w_{ij} \quad (1)$$

where v_i and h_j are the *ith* visible variable and the *jth* hidden variable, i.e., the original input i and the feature j in the feature space; b_i and b_j are the biases to the nodes, and w_{ij} is the connection weight between them. The RBM assigns a probability to every possible pair of visible and hidden variables using the energy function

$$P(v,h) = \frac{1}{Z} e^{-E(v,h)} \quad (2)$$

where Z is the sum of all the possible pairs of the visible and hidden variables expressed as

$$Z = \sum_{v,h} e^{E(v,h)} \tag{3}$$

The probability of the visible v is given as

$$P(v) = \frac{1}{Z}\sum_{h} e^{-E(v,h)} \tag{4}$$

To minimize the energy of that input data, the weights and the biases will be adjusted by

$$\Delta w_{ij} = \eta(\langle v_i h_i \rangle - \langle v_i' h_i' \rangle) \tag{5}$$

where η is the learning rate, and $\langle \bullet \rangle$ denotes the expectation under the distribution of the variables v_i and h_i, and their reconstructed pair v_i' and h_i'.

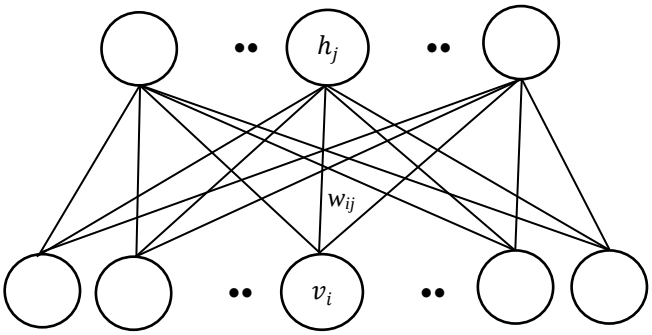

Fig. 2. Typical structure of RBM.

In practice, a single RBM may not be sufficient to extract features with a reduced dimensionality. Often several RBMs are used in a sequential way, forming stacked RBMs layer-by-layer. The outputs of a trained RBM in the stack is used as the inputs to the next adjacent RBM for training. The number of RBMs to be used varies and usually is determined using a trial-and-error approach. For a detailed guidance on the training the readers can refer to, for example, [22].

3.2 *k*-means Cluster Analysis

k-means clustering is one of the most popular algorithms in data mining for grouping samples into a certain number of groups (clusters) based on Euclidean distance measure. Assume $V_1, V_2, \cdots, V_n$ are a set of vectors, and these vectors are to be assigned to k clusters $S_1, S_2, \cdots, S_k$. Then the objective function of the *k*-means clustering can be expressed as

$$f(m_1, m_2, \cdots, m_k) = \sum_{i=1}^{k} \sum_{V_j \epsilon S_i} \left\| V_j - m_i \right\|^2 \quad (6)$$

where m_i represents the centroid of cluster S_i.

4 Data Pre-processing

The SEER breast cancer data explored in this study contains totally 291,760 incidences registered in the US from 1974 to 2017. The data set has two separate collections with incidences from 1974 to 2014 and 2014 to 2017, respectively. Note that the number of variables in the two sets of collections is 134 and 130, respectively. Most of the variables in the data set are categorical type with a varying number of distinct values from 2 to more than 100. The original data sets need to be pre-processed and transformed to a target data set for analysis. The most crucial task in the data pre-processing process is to identify if there are any data quality issues in the data and further to adopt appropriate strategies to address them accordingly.

The SEER data under consideration has typical data quality problems, including inconsistent variables in the two collections of the data, some duplicate variables and therefore directly or indirectly correlated to each other, missing values, and in-comparative value ranges for several numeric variables. As such, the main tasks involved in the data pre-processing are as follows:

1. Select meaningful and common variables that are applicable to both sets of the data. Further, remove any duplicate variables that are identical from statistical perspective and/or from medical diagnostical perspective. As a result, a total of 130 variables have been chosen to use.
2. Remove any incidences that contain missing value. The removal is reasonable and acceptable since the entire data set is big and there were only some 10,000 incidences containing missing values. As such no replacement of missing value is needed.
3. Transform the value range of each numeric variable into a unit interval [0,1] using the min-max normalization.
4. Represent each distinct value of a categorical value as a unit vector using the one-hot encoding. This leads to a significant number of dummy variables to be created.

The data pre-processing process was very time-consuming, and it has eventually led to a resultant target data set with 260000 incidences and 961 variables. The variable **Survival** that represents if a patient survived has been considered the target variable since this analysis is aiming at identifying crucial factors that potentially affect the survival of a breast cancer patient.

It should be noted that no incidences have been removed although they may be considered outliers, and this is because each instance should be analysed.

5 Experimental Settings

Using the target data set created, RBMs have been implemented for dimensionality reduction and feature extraction, and further the *k*-means clustering analysis has been applied to the samples in the feature space formed by the RBMs.

Three RBM models, RBM_1, RBM_2 and RBM_3, have been constructed in a sequential manner as follows:

- RBM_1 has 961 input nodes and 625 (25×25) hidden nodes;
- RBM_2 has 625 input nodes and 169 (13×13) hidden nodes; and
- RBM_3 has 169 input nodes and 81 (9×9) hidden nodes.

RBM_1 needs to be trained using the target data set; RBM_2 needs to be trained using the outputs of the trained RBM_1, and RBM_3 needs to be trained using the outputs of the trained RBM_2. In other words, the original data in a 961-dimensioanl space has been transformed into an 81-dimesional space for analysis.

RBM_1, RBM_2, and RBM_3 have been trained with 10,000, 20,000, and 20,000 iterations, respectively. The initial values of all the connection weights and the biases were randomly selected from a uniform distribution in the interval [-1, 1].

Following the entire analysis process as shown in Fig. 1, the *k*-means clustering analysis has been applied to the outputs of RBM_3. The number of centroids was set to 6 with randomly selected initial centroids to start the clustering process.

6 Pattern Interpretations and Findings

In order to interpret each of the clusters created for diagnosis purpose, the results obtained about the cluster membership of all the samples have been mapped back to the original space (of 961 dimensions) since the variables in each of the lower dimensional spaces is not interpretable.

Note that each instance of a patient's records in each of the spaces, e.g., the original and all the RBM spaces, can be visualized by a "facial" like imagery and this enables each patient's profile can be compared with each other in an initiative and easy way. Examples of such imagery description are provided in Figure 3, where each two rows from the top to the bottom contains 10 patients' profile of the same cluster in the RBM_3 (81 dimensions) space and their counterparts in the original (961 dimensions) spaces, respectively. Only samples from 4 out of the 6 clusters created were selected. The value for each pixel of the images is either 1 or 0 if a component associated is binary data type, or between 0 and 1 if the component associated is numeric data type, indicting a scale grading between black and white, especially in the original space.

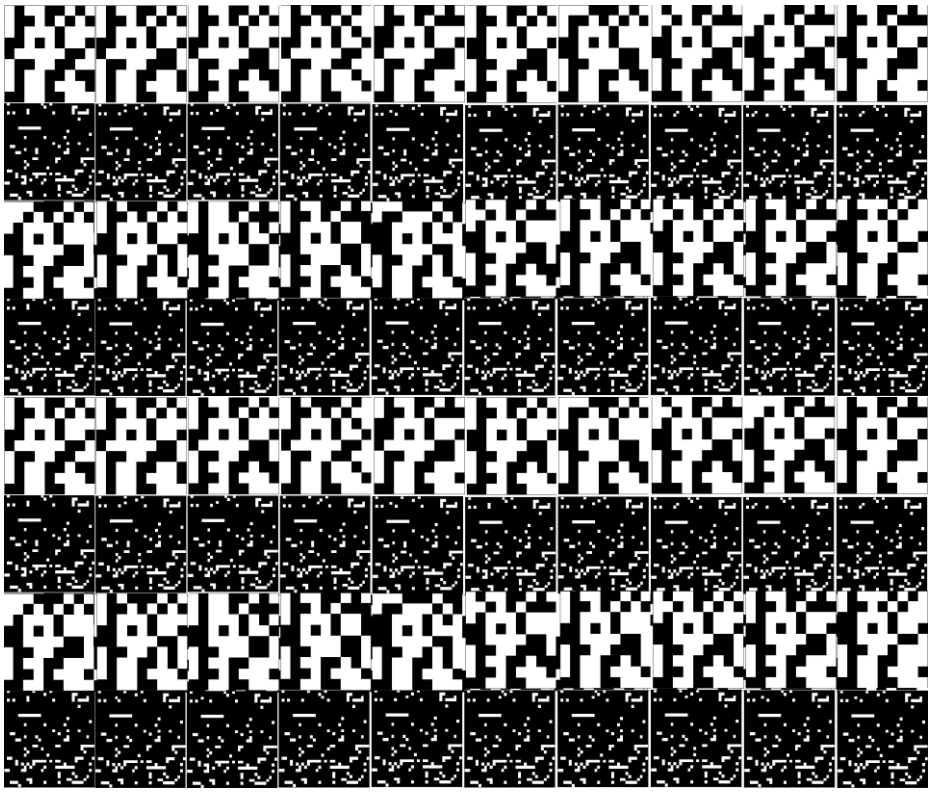

Fig. 3. Samples of a patient's profile in a "facial" like imagery description. Each two rows from the top to the bottom contains 10 patients' profile of the same cluster in the RBM_3 (81 dimensions) space and their counterparts in the original (961 dimensions) spaces, respectively.

The six clusters are labelled as Clusters 0, 1, …, and 5. To examine the clusters, and to compare them with each other, several variables have been used as shown in Table 1. In terms of factors affecting survival rate, there are several factors are crucial as highlighted in yellow in the Table.

It appears that if a patient had a surgery performed and the stage of tumor are two essential casual attributors. The survival rate for those who either didn't have surgery performed or were not recommended for a surgery had a very low survival rate (only 25%). In addition, patients in this group usually had a stage IV or not known tumor. As such accurately and timely detect the stage of a tumor and recommend on having a surgery or not accordingly is crucial.

Table 1. Table captions should be placed above the tables.

Cluster	0	1	2	3	4	5
Survival (%)	25.00	92.00	72.00	89.00	74.00	78.00
Grade (%)						

Grade 1	7.00	23.00	10.00	14.00	3.00	32.00
Grade 2	27.00	45.00	41.00	31.00	8.00	59.00
Grade 3	31.00	24.00	43.00	27.00	80.00	3.00
Grade 4	3.00	0.00	1.00	2.00	3.00	0.00
Cell not Determined	32.00	8.00	5.00	26.00	6.00	6.00
Number of Malignant Tumors						
1	58.00	1.00	68.00	47.00	64.00	57.00
2	33.00	71.00	26.00	39.00	28.00	33.00
3	7.00	23.00	5.00	11.00	7.00	8.00
4	2.00	4.00	1.00	2.00	1.00	1.00
5	0.00	1.00	0.00	1.00	0.00	1.00
6	0.00	0.00	0.00	0.00	0.00	0.00
7	0.00	0.00	0.00	0.00	0.00	0.00
Treatment						
Surgery Performed	38.00	95.00	99.00	95.00	99.00	99.00
Surgery not Recommended	34.00	2.00	1.00	3.00	1.00	1.00
Contraindicated	2.00	0.00	0.00	0.00	0.00	0.00
Died before	0.00	0.00	0.00	0.00	0.00	0.00
Unknown Reason for No Surgery	15.00	1.00	0.00	2.00	0.00	0.00
Refused	4.00	1.00	0.00	0.00	0.00	0.00
Recommended	1.00	1.00	0.00	0.00	0.00	0.00
Unknown if Surgery Performed	6.00	0.00	0.00	0.00	0.00	0.00
Alive or Died due to Cancer	24.00	1.00	65.00	57.00	68.00	67.00
Dead	41.00	0.00	15.00	1.00	7.00	3.00
Not first Tumor	35.00	99.00	20.00	42.00	25.00	30.00
In Situ	0.00	3.00	0.00	99.00	0.00	0.00
Malignant	100.00	97.00	100.00	1.00	100.00	100.00
Primaries						
One primary only	58.00	1.00	67.00	47.00	63.00	57.00
First of Two or More Primaries	6.00	0.00	12.00	12.00	12.00	13.00
Second of Two or More Primaries	30.00	79.00	18.00	34.00	22.00	26.00
Third of Three or More Primaries	4.00	16.00	2.00	6.00	3.00	4.00
Fourth of Four or More Primaries	1.00	3.00	0.00	1.00	0.00	0.00
Fifth of Five or More Primaries	0.00	0.00	0.00	0.00	0.00	0.00
Sixth of Six or More Primaries	0.00	0.00	0.00	0.00	0.00	0.00
Seventh of Seven or More Primaries	0.00	0.00	0.00	0.00	0.00	0.00
Eighth of Eight or More Primaries	0.00	0.00	0.00	0.00	0.00	0.00

Ninth of Nine or More Primaries	0.00	0.00	0.00	0.00	0.00	0.00
Unknown	0.00	0.00	0.00	0.00	0.00	0.00
Stage						
Stage 0	0.00	3.00	0.00	99.00	1.00	0.00
Stage I	2.00	62.00	1.00	0.00	62.00	84.00
Stage IIA	2.00	21.00	36.00	0.00	30.00	12.00
Stage IIB	2.00	6.00	28.00	0.00	3.00	1.00
Stage III NOS	1.00	0.00	0.00	0.00	0.00	0.00
Stage IIIA	2.00	2.00	22.00	0.00	0.00	0.00
Stage IIIB	8.00	1.00	3.00	0.00	1.00	0.00
Stage IIIC	4.00	1.00	9.00	0.00	0.00	0.00
Stage IV	41.00	1.00	0.00	0.00	0.00	0.00
Not Applicable	2.00	0.00	0.00	0.00	0.00	0.00
Stage Unknown	36.00	3.00	2.00	0.00	2.00	2.00

A consistent pattern can be identified if, for example, examining the two clusters, Cluster 0 and Cluster 1, using a radar graph as shown in Figure 3.

It is evident from Figure 3 that high survival rate was closely correlated with if a tumor was localized and if a surgery was performed. On the other hand, low survival was in general related to a tumor was regional and distant, and a surgery was not performed.

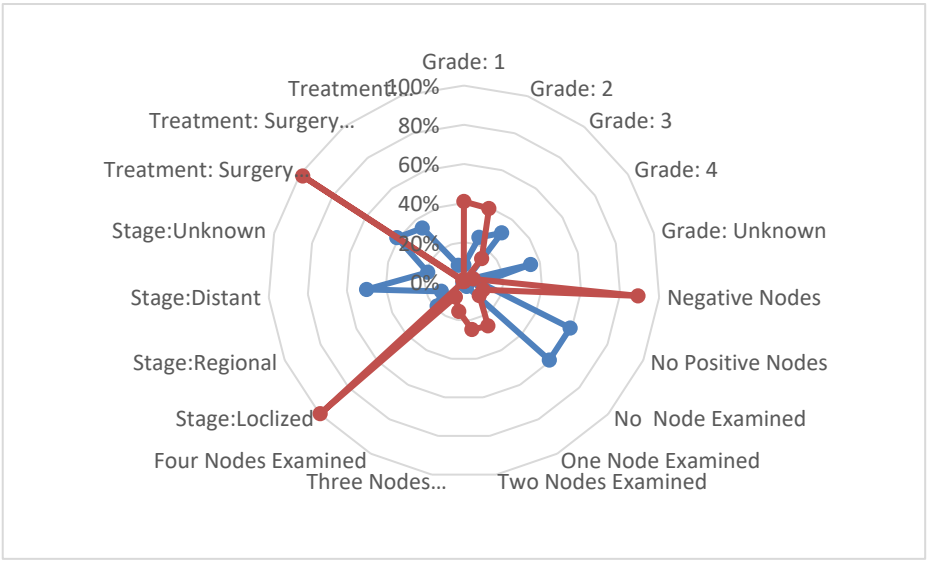

Fig. 3. Low survival rate (in blue) vs. high survival rate (in red) with several factors.

7 Concluding Remark and Future Work

In this paper, a deep learning-based approach is applied to high dimensional, high-volume, and high-sparsity medical data to identify critical casual attributions that

might affect the survival of a breast cancer patient. The analysis has demonstrated that essential features in a high-dimensional sample space can be effectively extracted and brought forward into a much lower dimensionality space formed by RBMs. This has provided a novel approach to understand high-dimensional data.

The analysis has also identified several crucial casual factors that significantly affect a patient's survival rate among all the variables.

Further research will focus on exploring what features RBM extracts, how to interpret the feature space established by an RBM and design an ideal imaginary description for visualizing a healthy person for comparison purpose.

References

1. WHO: https://www.who.int/cancer/prevention/diagnosis-screening/breast-cancer/en/.
2. Hankey, B.F., Ries, L.A., Edwards, B.K.: The surveillance, epidemiology, and end results program. Cancer Epidemiology and Prevention Biomarkers 8(12), 1117–1121 (1999).
3. Han, J., Pei, J., Kamber, M.: Data mining: concepts and techniques. Elsevier, 2011.
4. Zhang, K., Liu, J., Chai, Y., Qian, K.: An optimized dimensionality reduction model for high-dimensional data based on restricted boltzmann machines. In: 27th Chinese Control and Decision Conference (2015 CCDC), pp. 2939–2944. IEEE, Qingdao, China (2015).
5. Katkar, J.A., Baraskar, T.: A novel approach for medical image segmentation using PCA and K-means clustering. In: 2015 International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), pp. 430–435. IEEE, Davangere, India (2015).
6. Sullivan, C.W., Leutwyler, H., Dunn, L.B., Cooper, B.A., Paul, S.M., Conley, Y.P., Levine, J.D, Miaskowski, A.: Differences in symptom clusters identified using symptom occurrence rates versus severity ratings in patients with breast cancer undergoing chemotherapy. European Journal of Oncology Nursing 28, 122–132 (2017).
7. Chen, A.T.: Exploring online support spaces: using cluster analysis to examine breast cancer, diabetes and fibromyalgia support groups. Patient education and counseling 87 (2), 250–257 (2012).
8. Sanford, S.D., Beaumont, J.L., Butt, Z., Sweet, J.J., Cella, D., Wagner, L.I., Prospective longitudinal evaluation of a symptom cluster in breast cancer. Journal of pain and symptom management 47 (4), 721–730 (2014).
9. Sarenmalm, E.K., Browall, M., Gaston-Johansson, F.: Symptom burden clusters: a challenge for targeted symptom management. A longitudinal study examining symptom burden clusters in breast cancer. Journal of pain and symptom management 47 (4), 731–741 (2014).
10. Rathnam, C., Lee, S., Jiang, X.: An algorithm for direct causal learning of influences on patient outcomes. Artificial Intelligence in Medicine 75, 1–15 (2017).
11. Fogel, D.B., Wasson III, E.C., Boughton, E. M.V., Porto, W., Evolving artificial neural networks for screening features from mammograms, Artificial Intelligence in Medicine 14 (3), 317– 326 (1998).
12. Blanco, R., Inza, I., Merino, M., Quiroga, J., Larrañaga, P.: Feature selection in bayesian classifiers for the prognosis of survival of cirrhotic patients treated with tips, Journal of biomedical informatics 38 (5), 376–388 (2005).
13. Kose, U., Alzubi, J. (eds): Deep Learning for Cancer Diagnosis. Springer (2020).

14. Zhou, Z.H., Jiang, Y., Yang, Y.B., Chen, S.F.: Lung cancer cell identification based on artificial neural network ensembles. Artificial Intelligence in Medicine 24 (1), 25–36 (2002).
15. Xu, R., Damelin, S., Nadler, B., Wunsch, D.C.: Clustering of high-dimensional gene expression data with feature filtering methods and diffusion maps. Artificial intelligence in medicine 48 (2), 91–98 (2010).
16. Li, L., Tang, H., Wu, Z., Gong, J., Gruidl, M., Zou, J., Tockman, M., Clark, R.A., Data mining techniques for cancer detection using serum proteomic profiling. Artificial intelligence in medicine 32 (2), 71–83 (2004).
17. Regnier-Coudert, O., McCall, J., Lothian, R., Lam, T., McClinton, S., NDow, J.: Machine learning for improved pathological staging of prostate cancer: a performance comparison on a range of classifiers, Artificial intelligence in medicine 55 (1), 25–35 (2012).
18. Yang, X., Cao, A., Song, Q., Schaefer, G., Su, Y.: Vicinal support vector classifier using supervised kernel-based clustering, Artificial intelligence in medicine 60 (3) (2014) 189–196.
19. Kakushadze, Z., Yu, W.: k-means and cluster models for cancer signatures. Biomolecular Detection & Quantification 13, 7–31 (2017).
20. Lisboa, P.J., Wong, H., Harris, P., Swindell, R.: A bayesian neural network approach for modelling censored data with an application to prognosis after surgery for breast cancer. Artificial intelligence in medicine 28 (1), 1–25 (2003).
21. Hinton, G.E.: A Practical Guide to Training Restricted Boltzmann Machines. In: Montavon, G.: OrrKlaus-Robert Müller, B.: (eds.) Neural Networks: Tricks of the Trade, pp. 599–619. Springer (2012).
22. Hinton, G.E., Salakhutdinov, R.R.: Reducing the Dimensionality of Data with Neural Networks. SCIENCE 313 (5786), 504–507 (2006).