

Original research article

# Deep stacked sparse auto-encoders for prediction of post-operative survival expectancy in thoracic lung cancer surgery

Mohammad Saber Iraj<sup>i</sup> \*

Payame Noor University, Faculty of Engineering, Department of Computer Engineering and Information Technology, Tehran, Iran

## Abstract

Lung cancer is the leading cause of cancer death in men and women. The prognostic value of survival after lung cancer surgery has an important role in decision-making for surgeons and patients. The combination of clinical features and CT scan information for diagnosis, treatment and survival of patients with lung cancer increases the accuracy of prediction using machine learning. Therefore, creating a computer intelligent method with low error and high accuracy to predict survival is an important challenge, and it is beneficial for decreasing mortality from lung cancer, and for planning treatment.

In this work, we implemented a deep stacked sparse auto-encoder (DSSAE) approach on a thoracic surgery data set for 470 patients, and our results contributing to deep learning based on 16 features were more precise than other suggested techniques for predicting post-operative survival expectancy in thoracic lung cancer surgery. The proposed method achieved a sensitivity of 94%, specificity of 82.86% and g-mean of 88.25%.

**Keywords:** Deep stacked sparse auto-encoders (DSSAEs); Lung cancer; Neural networks; Thoracic surgery

## Introduction

Lung cancer is the most common cause of death in the world and 1.4 million people die each year from lung cancer. Lung cancer is a major disease worldwide, and about 27% of all deaths from cancer are related to lung cancer (Yu et al., 2016). It is essential to identify patients selected for treatment of lung cancer through surgery who have a high survival risk after surgery. The mortality rate after lung cancer treatment for 30 or 90 days is a preferred index for the cancer population (In et al., 2016). The prediction and prognosis of the possible complications and benefits of surgery can help appropriate decisions to be made by doctors and by patients with lung cancer.

Past research evidences that soft computing techniques such as artificial neural networks (Geng et al., 2016; Luo, 2017), fuzzy logic (Kuo et al., 2015; Omiotek et al., 2013), support vector machines (Dobrowolski et al., 2016; Dolatabadi et al., 2017) and deep learning (Ciompi et al., 2015; Cosma et al., 2017; Gargeya and Leng, 2017; Pham et al., 2017; Sun et al., 2017; Yin et al., 2016; Zhang et al., 2016a), which have been used for medical prediction and decision-making, have good results using training data sets.

Simonsen et al. (2015) have expressed risk factors for post-operative pneumonia (POP) in patients undergoing therapy by lung cancer surgery within 30 days after surgery, with a

logistic regression method. They reported that previous pneumonia, advanced age, chronic pulmonary disease, obesity, atrial fibrillation and alcoholism are major factors. Möller and Sartipy (2012) declared that evolutions in modality of life for 6 months after lung cancer surgery can predict survival. They studied answers to 36-Item Short Form (SF-36) questions before and 6 months after lung cancer surgery, using a regression model. Tanvetyanon et al. (2015) offered an acceptable prediction of post-operative survival for surgery of multi-lobe non-cell lung cancers with a lack of metastatic illness. Histology, age, sex, highest N-stage, maximal T-size and laterality were effective for estimation using the regression method. Yu et al. (2016) suggested a support vector machine to specify effective features in shorter-term and longer-term survivors of lung cancer from The Cancer Genome Atlas (TCGA) and Stanford tissue microarray image databases.

Oakden-Rayner et al. (2017) alluded to a fatality classification task from chest CT images and muscle, aorta, body fat, epicardial fat, vertebral column, heart and lung segmentation maps using deep learning for 5-year mortality classification. Zhu et al. (2016a) propounded a deep convolutional neural network for survival analysis with pathological images based on National Lung Screening Trial (NLST) lung cancer data. Zięba et al. (2014) have proposed a boosted support vector machine (SVM) classifier to categorize patients into two classes: class 1 – death within 1 year after surgery, and class 2 – sur-

\* **Author for correspondence:** Mohammad Saber Iraj<sup>i</sup>, Movahhed Apartment, Shalikubi Ave, Edalate 2 alley, Golestan province, Gorgan, Iran; e-mail: [iraji.ms@gmail.com](mailto:iraji.ms@gmail.com); [iraji.ms@pnu.ac.ir](mailto:iraji.ms@pnu.ac.ir)  
<http://doi.org/10.32725/jab.2018.007>

Submitted: 2018-07-13 • Accepted: 2018-12-03 • Prepublished online: 2019-01-10

J Appl Biomed 17/1: 68–75 • EISSN 1214-0287 • ISSN 1214-021X

© 2018 The Authors. Published by University of South Bohemia in České Budějovice, Faculty of Health and Social Sciences.

This is an open access article under the CC BY-NC-ND license.

vival. They elicited nine decision rules from a thoracic surgery data set including 470 samples and 16 features for medical utilization, for predicting post-operative survival expectancy in lung cancer patients. Iraji (2017) has implemented a multi-layer architecture of sub-adaptive neuro-fuzzy inference system (MLA-ANFIS) with various combinations of multiple input features, regression, neural networks and an extreme learning machine (Macías-García et al., 2017) based on a thoracic surgery data set with 16 input features for prediction of 1-year post-operative survival expectancy in thoracic lung cancer surgery; ELM (wave kernel) has good performance.

In a study (van der Burgh et al., 2017), a deep learning technique was used as a method to indicate survival by clinical parameters and MRI data in amyotrophic lateral sclerosis (ALS).

Macías-García et al. (2017) introduced a deep learning approach to improve the quality of the genomic data used to expand immunohistochemistry signatures to predict breast cancer. Kuan et al. (2017) investigated a deep learning method for lung cancer detection using 3D lung CAT scans for determining malignancy of the cancer. Ciompi et al. (2017) classified lung cancer nodules into solid, non-solid, part-solid, calcified, perifissural and speculated nodules, using deep learning from CT scan images. Wang et al. (2017) proclaimed a system for classifying mediastinal lymph node metastasis of non-small cell lung cancer, from CT images using machine learning methods containing SVMs, random forests, deep learning, adaptive boosting and artificial neural networks. Their results indicated that the highest accuracy was by deep learning.

The Up survey discussed the usage of several available methods of lung and thoracic surgery, using soft computing techniques. Among soft computing approaches, reasons for selecting deep methodology are that it calibrates the learning function by using training data, and offers better performance in forecasting and classification.

The number of patients who die after surgery is lower than the number of patients who survive for a 1-year period. Nevertheless, new tools and methods are essential to help doctors improve their decision-making, and to predict survival expectancy of patients after lung cancer surgery, in order to minimize mortality. The motivation of this study is to plan a smart system with a deep learning method, to aid clinicians in predicting post-operative survival expectancy in thoracic

lung cancer surgery with high accuracy. In this paper, a new intelligent system based on the deep learning approach in the clinical diagnosis of thoracic lung cancer surgery is suggested, which assists doctors in patient election, and recognizes the risk of death after surgery.

## Materials and methods

### Deep learning

Deep learning is a recent development in machine learning and is used in many research fields, such as speech recognition (Martinez et al., 2017), drug discovery and toxicology (Tian et al., 2016), customer relationship management (Singh and Tucker, 2017), computer vision (Lu et al., 2017; Zhang et al., 2017), and bioinformatics (Wahab et al., 2017) with good results.

One of the characteristics of deep learning is the ability to handle data directly instead of using the extracted features of data produced by different algorithms.

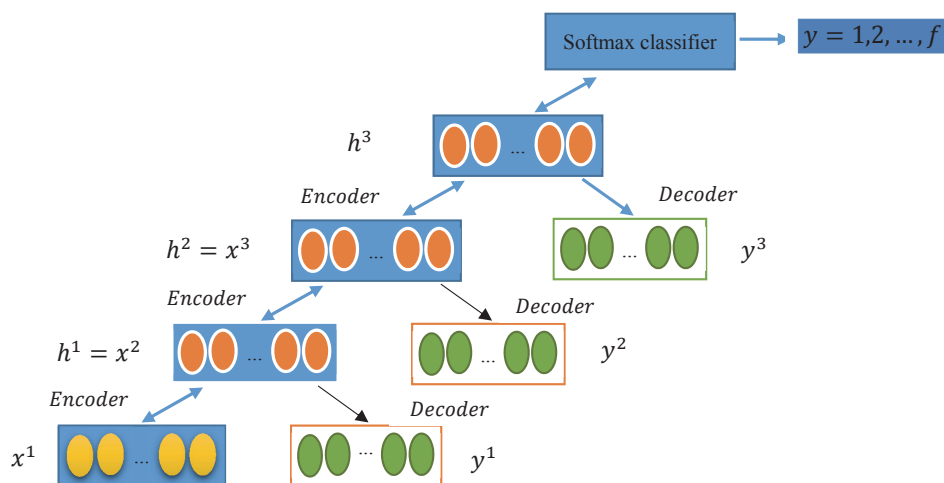
The considered deep network comprises stacked sparse auto-encoders (SSAEs) as illustrated in Fig. 1. The blue path in Fig. 1 shows the structure of a deep neural network.

An auto-encoder neural network is an unsupervised learning method that utilises the backpropagation learning approach to adjust the goal values to match the input values.

The auto-encoder trains to learn a function  $y^i \approx x^i$  and it can discover interesting structures in the input data by placing limitations on the number of hidden units, and sparsity restrictions on the hidden units.

In fact, this simple auto-encoder is very similar to PCAs, but PCAs only allow linear transformation, whereas auto-encoders are non-linear, and can learn more complex relationships between visible and hidden units; the network learns a compressed representation of the input. The auto-encoder inflicts a sparsity constraint on the hidden units, allowing it to discover interesting features in the data.

When input data are given to a deep network, it first compresses (encodes) input data, “adjusting” it to a smaller presentation, and then attempts to reconstruct (decode) it. The motivation for training the network is to minimise error in reconstruction and find the most effective compressed presentation (encoding) for the input data.



**Fig. 1.** Deep stacked sparse auto-encoders (DSSAEs) architecture

Multiple levels of auto-encoders are included, with a visible layer, a hidden layer and a reconstruction layer, at each level. The visible (input) layer of the second level auto-encoder is the hidden layer in the first level auto-encoder and the visible layer of the third level auto-encoder is the hidden layer in the second level auto-encoder, et cetera.

The first level auto-encoder is trained to obtain the first level features in the hidden layer, then the features are disseminated to the second level auto-encoder as input. In a similar manner, all of the auto-encoders are trained and will obtain higher features at the top level.

Auto-encoder training is formed of two parts: an encoder which extracts the features  $h$  from the input  $x$  and a decoder which reconstructs the output  $y$  from those extracted features  $h$ .

Encoder and decoder adaptation is represented by equations (1) and (2):

$$(1) \quad h = f(w_e x + b_e)$$

$$(2) \quad f = f(w_d h + b_d)$$

Where  $w_e$  and  $w_d$  represent the weight matrix of the encoder and decoder  $b_e$  and  $b_d$  are the bias vectors, respectively.  $f(\cdot)$  is the activation function of the auto-encoder.

The goal here is minimise the error function between the input data of visible layer and the reconstructed data of reconstruction layer. The error cost function is computed by equation 3. The prime term is an average sum of squares error and the second term is the weight degradation term, while the third term is a sparse penalty term:

$$(3) \quad c(x, y) = \frac{1}{2n} \sum_{j=1}^n \|x^j - y^j\|_2^2 + \frac{\lambda}{2} (\|w_e\|_2^2 + \|w_d\|_2^2) + \beta \sum_{j=1}^m KL(\rho \|\hat{\rho}_j)$$

where:

$n$  = the number of data inputs

$\lambda$  = the weight degradation parameter which controls the relative momentousness of the prime and the second terms

$\beta$  = a parameter controlling the weight of the sparse penalty term

$m$  = the number of neurons in the hidden layer

$KL(\rho \|\hat{\rho}_j)$  is the kullback-leibler (Cosma et al., 2017) divergence between tow bernoulli random variables with mean  $\rho$  and mean  $\hat{\rho}_j$

$\hat{\rho}_j$  is the average activation of hidden unit  $j$  or averaged over the training set and  $\rho$  is a sparsity parameter, usually a small value near to zero ( $\rho = 0.05$ )

$KL(\rho \|\hat{\rho}_j)$  is calculated using equation 4:

$$(4) \quad KL(\rho \|\hat{\rho}_j) = \rho \log \frac{\rho}{\hat{\rho}_j} + (1 - \rho) \log \frac{(1-\rho)}{(1-\hat{\rho}_j)}$$

where

$$(5) \quad \hat{\rho}_j = \frac{1}{n} \sum_{t=1}^n h_j(x^{(t)})$$

The number of auto-encoder level and hidden neurons in each level are determined by trial and error.

The high-level features or hidden neurons are given to the softmax regression (SF) model, which is a supervised learning algorithm in the deep neural network and classifies the input data.

If we consider  $\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \dots, (x^{(n)}, y^{(n)})\}$  tuples as input data, the output  $y$  can take on  $f$  different classes. We have that  $y^{(i)} \in \{1, 2, 3 \dots f\}$ . We desire our hypothesis to predict the probability that  $p(y = z | x)$  for  $z = 1, 2, \dots, f$ . Our hypothesis is defined by equation 6:

$$(6) \quad h_\theta(x^{(i)}) = \begin{bmatrix} p(y^{(i)} = 1 | x^{(i)}; \theta) \\ p(y^{(i)} = 2 | x^{(i)}; \theta) \\ \vdots \\ p(y^{(i)} = f | x^{(i)}; \theta) \end{bmatrix} = \frac{1}{\sum_{z=1}^f e^{\theta_z^T x^{(i)}}} \begin{bmatrix} e^{\theta_1^T x^{(i)}} \\ e^{\theta_2^T x^{(i)}} \\ \vdots \\ e^{\theta_f^T x^{(i)}} \end{bmatrix}$$

Where  $\theta_1, \theta_2, \theta_f$  are model parameters,  $\theta = \begin{bmatrix} \theta_1^T \\ \theta_2^T \\ \vdots \\ \theta_f^T \end{bmatrix}$ , the term

$\frac{1}{\sum_{z=1}^f e^{\theta_z^T x^{(i)}}}$  causes the distribution to sum to one and normalises it.

The cost function of the softmax regression classifier model is specified by equation 7:

$$(7) \quad J(\theta) = -\frac{1}{n} \left[ \sum_{i=1}^n \sum_{z=1}^f 1\{y^{(i)} = z\} \log \frac{e^{\theta_z^T x^{(i)}}}{\sum_{l=1}^f e^{\theta_l^T x^{(i)}}} \right] + \frac{\lambda}{2} \sum_{i=1}^f \sum_{j=0}^u \theta_{ij}^2$$

Where  $1\{\cdot\}$  is the indicator function,  $1\{\text{a true expression}\} = 1$  and  $1\{\text{a false expression}\} = 0$ . The weight decay term  $\frac{\lambda}{2} \sum_{i=1}^f \sum_{j=0}^u \theta_{ij}^2$  punishes high values of the parameters, and  $u + 1$  is dimensional of feature vectors  $x$ .

The cost function  $J(\theta)$  is certain to have a single solution using algorithms like L-BFGS, gradient descent, etc. since it is now entirely convex with the addition of the weight decay term. Further details about deep learning are found in Bengio (2009), Su et al. (2015), Zhang et al. (2016b) and Zhu et al. (2016b).

### Extreme learning machine (Macías-García et al., 2017)

An extreme learning machine (Macías-García et al., 2017) is a type of NN that is used (Equation 8) for a single-layer feed-forward NN that specifies the input weights randomly and the output weights analytically (Feng et al., 2009; Huang et al., 2006).

In the following equation,  $\alpha_k$  and  $b_k$  are learning parameters of activation functions and  $\beta_k$  is the weight of the  $k^{\text{th}}$  hidden neuron associated to the output neuron. The input data are  $(x_i, t_i)$ , and  $N$  is the number of samples where  $x_i = [x_{i2} \dots x_{in}]^T \in R^n$  and  $t_i = [t_{i1} \ t_{i2} \dots t_{im}]^T \in R^m$ .

$$(8) \quad H\beta = T$$

$$\text{So:} \quad (9) \quad H(\tilde{a}, \tilde{b}, \tilde{x}) = \begin{bmatrix} G(a_1, b_1, x_1) & \dots & G(a_L, b_L, x_1) \\ \vdots & \ddots & \vdots \\ G(a_1, b_1, x_N) & \dots & G(a_L, b_L, x_N) \end{bmatrix}_{N \times L}$$

$$(10) \quad G(a_k, b_k, x_i) = g(a_k x_i + b_k)$$

$$(11) \quad G(a_k, b_k, x_i) = g(b_k \|x_i - a_i\|)$$

$$(12) \quad \beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix}_{L \times m} \quad \text{and} \quad t = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix}_{N \times m}$$

## Results and discussion

Zięba et al. (2014) performed boosted SVM to discover rules from a real clinical data set for post-operative survival expectancy in lung cancer patients, and they extracted 16 features for predicting post-operative survival expectancy. The data set was gathered retroactively for lung samples from 1200 lung cancer patients in the years 2007–2011 at the Wrocław thoracic surgery station, which was managed by the Institute of Tuberculosis and Pulmonary Diseases in Warsaw, Poland.

The main data set contained 139 features, of which 36 were from pre-operative, 37 from peri-operative and 46 (containing 17 pathology-related) from post-operative periods.

Zięba et al. (2014) used 36 pre-operative features for the prediction of 1-year survival, and they reduced the number of features from 36 to 16 with 470 example records by the boosted SVM method.

Iraji (2017) predicted 1-year post-operative survival expectancy for thoracic lung cancer surgery by applying MLA-ANFIS, neural networks, regression and ELM, based on the same thoracic surgery data set with 16 input features.

We tried to design a careful method for predicting post-operative survival expectancy in lung cancer patients with a thoracic surgery data set (Zięba et al., 2014); 470 patient records were used, and the 16 input variable features are specified below.

- v1. DGN: Diagnosis – specific combination of ICD;
  - v2. PRE5: Volume exhaled at the end of the first second of forced expiration – FEV1 (numeric);
  - v3. PRE4: Forced vital capacity – FVC;
  - v4. PRE7: Pain before surgery;
  - v5. PRE6: Performance status – Zubrod scale;
  - v6. PRE8: Haemoptysis before surgery;
  - v7. PRE10: Cough before surgery;
  - v8. PRE11: Weakness before surgery;
  - v9. PRE9: Dyspnoea before surgery;
  - v10. PRE14: T in clinical TNM – size of the original tumour;
  - v11. PRE17: Type 2 DM – diabetes mellitus;
  - v12. PRE19: MI up to 6 months;
  - v13. PRE25: PAD – peripheral arterial diseases;
  - v14. PRE30: Smoking;
  - v15. PRE32: Asthma;
  - v16. AGE: Age at surgery;
- Output: Risk1Y: True value if died in 1-year survival period.

Of the patients with lung cancer disease in the data set, 400 patients survived (positive) and 70 patients died (negative); the number of patients who survived after surgery is greater than the number of patients who died. Choosing patients to undergo lung cancer surgery who have a lower risk of death during the short term (30-day period) or long term (1- or 5-year period) after surgery is very important (Zięba et al., 2014).

In this paper, we evaluated 1-year survival, to estimate post-operative survival expectancy in lung cancer patients.

We executed our proposed method in MATLAB version 9.1.0.441655 (R2016b) on a laptop with 1.7 GHz CPU, and we applied root mean square error (RMSE) and accuracy in order to calculate evaluation indices, to compare our method with others and to identify the best one. RMSE describes the sample standard deviation of the differences between predicted values and actual values, and accuracy demonstrates precision classification.

$$(13) \text{ Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

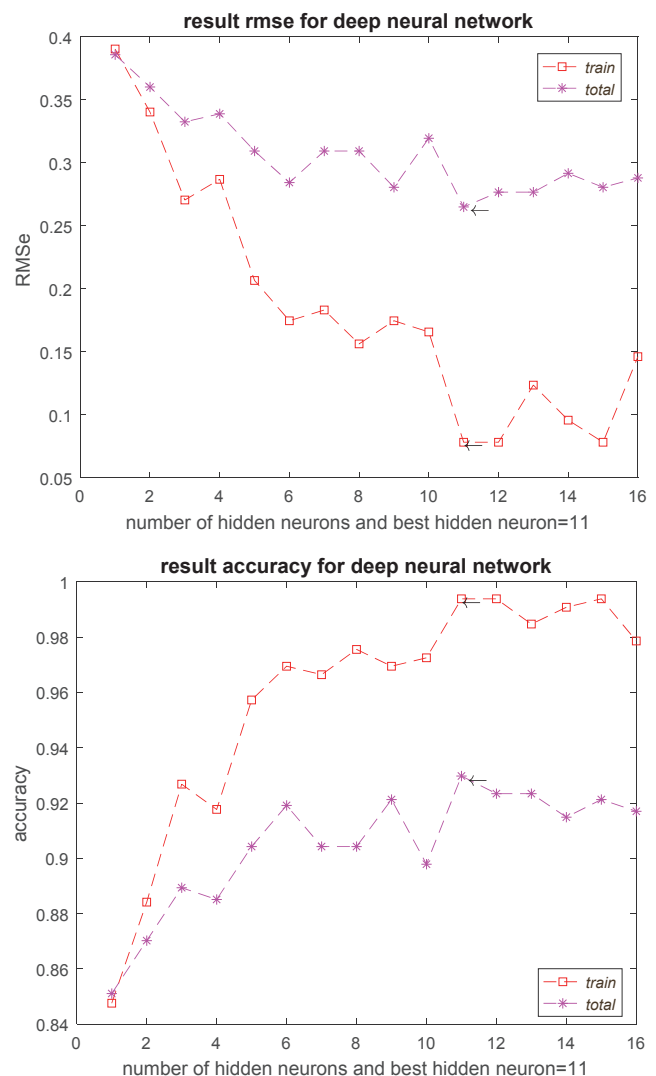
$$(14) \text{ MSE} = \frac{1}{N} \sum_1^N (\text{output}_{\text{actual}} - \text{output}_{\text{predicted}})^2$$

$$(15) \text{ RMSE} = \sqrt{\frac{1}{N} \sum_1^N (\text{output}_{\text{actual}} - \text{output}_{\text{predicted}})^2}$$

Deep stacked sparse auto-encoder (DSSAE) neural networks deploy a model of the system by using data, 70% of which was randomly selected for training data, 15% for validation and 15% for testing (Er et al., 2012; Kuncheva, 2004; Temurtas et al., 2009).

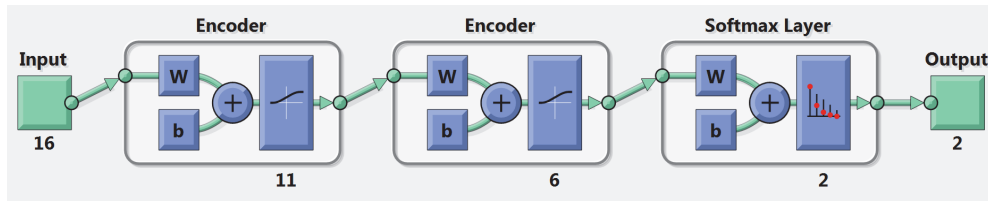
The native auto encoders utilize lesser number of neurons in the hidden layer than the input neurons in first layer to extract the useful compacted representations features from input data (Baldi and Hornik, 1989; Kramer, 1991).

Fig. 2 illustrates RMSE and accuracy between actual and predicted survival expectancy after thoracic lung cancer surgery, using DSSAEs from 1 to 16 hidden neurons. As seen in Fig. 2, the optimum number of hidden neurons is 11. DSSAE architecture 16-11-6-2 was considered (Fig. 3), and it achieved the best performance.



**Fig. 2.** RMSE and accuracy between actual and predicted survival expectancy after thoracic lung cancer surgery, using deep stacked sparse auto-encoders (DSSAEs) to determine the number of hidden neurons





**Fig. 3.** Deep stacked sparse auto-encoder (DSSAE) structure

There are many training and testing protocols (Kuncheva, 2004) but due to the same conditions in order to compare with previous works (Iraji, 2017; Zięba et al., 2014) we repeated proposed methods randomly 100 repetitions. To eradicate an over-learning phenomenon, we performed our proposed method for 100 runs with randomly selected training, validation and testing data for each simulation on the same data set.

Table 1 shows the average confusion matrix according to the predicted and actual data for training, testing, validation and all samples using DSSAEs to predict survival expectancy after thoracic lung cancer surgery.

**Table 1.** Average confusion matrix for 100 runs for training, testing, validation and all data by deep learning

Training		Predicted	
Actual		Survivor	Dead
Survivor		276	0
Dead		0	52
Testing		Predicted	
Actual		Survivor	Dead
Survivor		52	10
Dead		6	3
Validation		Predicted	
Actual		Survivor	Dead
Survivor		48	14
Dead		6	3
Total		Predicted	
Actual		Survivor	Dead
Survivor		376	24
Dead		12	58

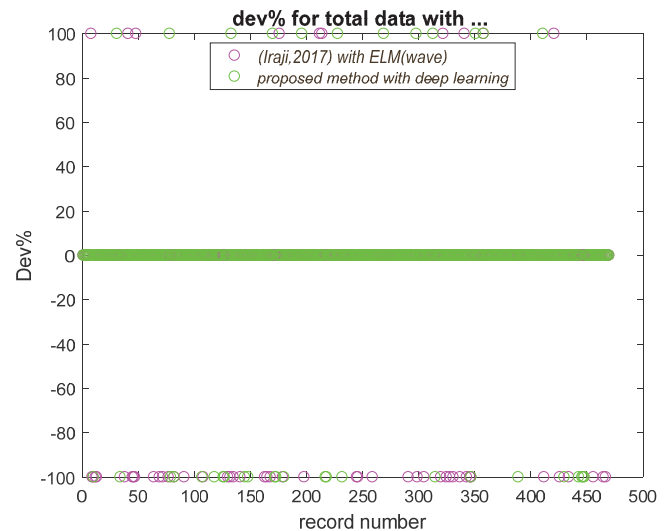
Averages of correct classification index and average RMSE for 100 runs between the proposed method and those from previous work are compared in Table 2. The average correct classification index and average RMSE were 92.34 and 0.28 after applying the DSSAE method, whereas (Iraji, 2017) obtained an average correct classification index of 88.79 and average RMSE of 0.33.

Comparison of average correct classification index and average RMSE for 100 runs between methods proves that the DSSAE system has better performance than the other methods.

Fig. 4 shows the deviation of predicted survival expectancy after thoracic lung cancer surgery from actual (dev%) for 470 records, using the DSSAE method and previous work (Iraji, 2017) with ELM; the DSSAE method has lower error values.

**Table 2.** Comparison of average correct classification index and average RMSE for 100 runs

Research reference	Method	RMSE	Correct classification index
(Iraji, 2017)	ELM (wave kernel)	0.33463	88.78724
This research	Deep stacked sparse auto-encoders (DSSAEs)	0.27675	92.34043



**Fig. 4.** Comparison of deviation of predicted survival expectancy after thoracic lung cancer surgery, using the proposed method (for colour revolution see the on-line edition)

In this paper, performance measures, namely sensitivity, specificity, geometric and mean area under the curve (AUC) were computed for forecast evaluation (Table 3). Patients surviving after thoracic lung cancer surgery are classed as positive, and dead people as negative. According to the equations below, sensitivity is the ratio of positives that are correctly recognized, and specificity is the ratio of negatives that are correctly recognized. In mathematics, the geometric mean is a type of mean or average that is defined as the  $m$ th root of the product of  $m$  numbers (equation 18).

$$(16) \quad \text{Sensitivity} = TP / (TP + FN)$$

$$(17) \quad \text{Specificity} = TN / (FP + TN)$$

where:

True positive (TP) = count of records that are correctly recognized.

False positive (FP) = count of records that are incorrectly recognized.

True negative (TN) = count of records that are correctly rejected.

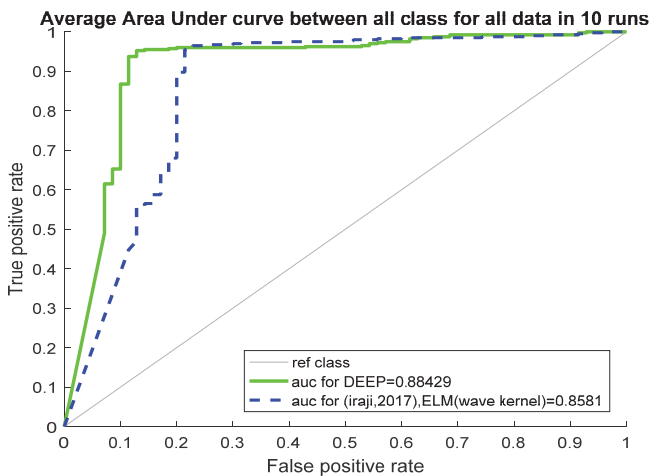
False negative (FN) = count of records that are incorrectly rejected.

$$(18) \text{ Geometric mean} = \sqrt[m]{x_1 x_2 \dots x_m}$$

Referring to Table 3, it can be observed that the DSSAE method is a better classifier than previous methods, with sensitivity of 94, specificity of 82.86 and g-mean of 88.25.

**Table 3.** Performance comparison of proposed method using average confusion matrix for 100 runs for all data

Research reference	Method	TP	TN	FP	FN	Sensitivity	Specificity	Geometric mean	Area under curve
Zięba et al., 2014	SVM	–	–	–	–	60	72	65.73	0.66
Iraji, 2017	ELM (wave kernel)	360.2	57.1	12.9	39.8	90.05	81.5714	85.7059	0.8581
This research	Deep stacked sparse auto-encoders (DSSAEs)	376	58	12	24	94.00	82.86	88.2529	0.8843



**Fig. 5.** Receiver operating characteristics (ROC) plot for the deep method

Our results show that RMSE = 0.28 (95% CI, 0.24–0.32), correct classification indicator = 92.34 (95% CI, 89.55–94.58) and g-mean = 88.25 (95% CI, 85.34–91.16), with sensitivity of 94 (95% CI, 91.20–96.12), specificity of 82.86 (95% CI, 71.97–90.82) and AUC of 0.8843 (95% CI, 0.86–0.91), so the DSSAE method is better for predicting post-operative survival expectancy in thoracic lung cancer surgery than other proposed algorithms, as that of (Iraji, 2017) has a g-mean of 85.70 with sensitivity of 90.05, specificity of 81.57 and AUC of 0.8581 using the same data set; the calculations performed on experimental results shown in Tables 1–3 prove this claim. Also, Zięba et al. (2014) found a g-mean of 65.73 with sensitivity of 60.00 and specificity of 72.00 using the same data set.

## Conclusions

People like to know how long they are likely to survive; they save themselves from dying by changing their lifestyle, and identifying the factors that affect it. To treat diseases with surgery, prognosis of post-operative survival is very effective for decision-making by doctors and patients.

In the receiver operating characteristics (ROC) plot, FP rate is drawn on the X axis and TP rate is drawn on the Y axis; AUC is an index of classification that shows the performance of methods (Carter et al., 2016). Fig. 5 shows the ROC for the deep method, and the 45° native is above it. Average values of AUC for 100 runs for the positive class were 0.8843 using the deep method, while it was reported as 0.8581 by ELM (wave) in a previous work (Iraji, 2017), and 0.66 by SVM by (Zięba et al., 2014).

The aim of this research was to create a solution to use machine learning techniques, namely DSSAEs, for prediction of post-operative survival expectancy in thoracic lung cancer surgery. We have provided DSSAEs with a 16-11-6-2 structure, using two stacked sparse auto-encoders; the advantage of the approach used is the design of a system capable of classifying survivors and non-survivors after surgery, from features extracted before surgery.

We have shown that our approach based on DSSAEs can be successful in solving the problem of predicting post-operative survival expectancy in thoracic lung cancer surgery, and the performance of the proposed method is higher than that of other solutions.

It could also be considered, using genetic algorithms to determine the DSSAE structure weights accurately, that the number of hidden neurons and counts of auto-encoders could be specified more precisely. We could change the type of learning function and cost function in DSSAEs, and better results would probably be achieved. Genomic analysis of the tumour, neoadjuvant therapy, anatomopathological findings and biomarker features were not included in this research, a restriction of this study. By training on other data sets, the efficiency of the proposed system can be enhanced to make it more effective in making decisions about post-operative survival expectancy in thoracic lung cancer surgery.

We used SAE for learning patterns from the patient data. SAEs usually create features for reconstructing the given data in unsupervised fashion. The use of SAE is helpful in extracting features taking advantage of the data characteristics, but there may be limitations in extracting “discriminative” features, which can be critical in distinguishing the classes. We should further consider the extraction of discriminative features through such as fine-tuning of the network structures.

Prediction of post-operative survival expectancy after surgery for other organs in the body could be done in future work. Other than 1-year survival, the most important thing is long-term survival of these patients; this is an issue that could be addressed by other models.

## Conflict of interests

The author have no conflict of interests to declare.

## References

- Baldi P, Hornik K (1989). Neural networks and principal component analysis: Learning from examples without local minima. *Neural networks* 2(1): 53–58. DOI: 10.1016/0893-6080(89)90014-2.
- Bengio Y (2009). Learning deep architectures for AI. *Foundations and trends® in Machine Learning* 2(1): 1–127. DOI: 10.1561/22000000006.
- Carter JV, Pan J, Rai SN, Galandiuk S (2016). ROC-ing along: Evaluation and interpretation of receiver operating characteristic curves. *Surgery* 159(6): 1638–1645. DOI: 10.1016/j.surg.2015.12.029.
- Ciampi F, Chung K, van Riel SJ, Setio AAA, Gerke PK, Jacobs C, et al. (2017). Towards automatic pulmonary nodule management in lung cancer screening with deep learning. *Sci Rep* 7: 46479. DOI: 10.1038/srep46479.
- Ciampi F, de Hoop B, van Riel SJ, Chung K, Scholten ET, Oudkerk M, et al. (2015). Automatic classification of pulmonary peri-fissural nodules in computed tomography using an ensemble of 2D views and a convolutional neural network out-of-the-box. *Med Image Anal* 26(1): 195–202. DOI: 10.1016/j.media.2015.08.001.
- Cosma G, Brown D, Archer M, Khan M, Pockley AG (2017). A survey on computational intelligence approaches for predictive modeling in prostate cancer. *Expert Syst Appl* 70: 1–19. DOI: 10.1016/j.eswa.2016.11.006.
- Dobrowolski A, Suchocki M, Tomczykiewicz K, Majda-Zdancewicz E (2016). Classification of auditory brainstem response using wavelet decomposition and SVM network. *Biocybern Biomed Eng* 36(2): 427–436. DOI: 10.1016/j.bbe.2016.01.003.
- Dolatabadi AD, Khadem SEZ, Asl BM (2017). Automated diagnosis of coronary artery disease (CAD) patients using optimized SVM. *Comput Methods Programs Biomed* 138: 117–126. DOI: 10.1016/j.cmpb.2016.10.011.
- Er O, Tanrikulu AC, Abakay A, Temurtas F (2012). An approach based on probabilistic neural network for diagnosis of Mesothelioma's disease. *Computers & Electrical Engineering* 38(1): 75–81. DOI: 10.1016/j.compeleceng.2011.09.001.
- Feng G, Huang G-B, Lin Q, Gay R (2009). Error minimized extreme learning machine with growth of hidden nodes and incremental learning. *IEEE Trans Neural Netw* 20(8): 1352–1357. DOI: 10.1109/TNN.2009.2024147.
- Gargeya R, Leng T (2017). Automated Identification of Diabetic Retinopathy Using Deep Learning. *Ophthalmology* 124(7): 962–969. DOI: 10.1016/j.ophtha.2017.02.008.
- Geng D, Zhou W, Zhang Y, Geng S (2016). Epileptic seizure detection based on improved wavelet neural networks in long-term intracranial EEG. *Biocybern Biomed Eng* 36(2): 375–384. DOI: 10.1016/j.bbe.2016.03.001.
- Huang G-B, Zhu Q-Y, Siew C-K (2006). Extreme learning machine: theory and applications. *Neurocomputing* 70(1–3): 489–501. DOI: 10.1016/j.neucom.2005.12.126.
- In H, Palis E, Merkow RP, Posner MC, Ferguson MK, Winchester DP, Pezzi CM (2016). Doubling of 30-day mortality by 90 days after esophagectomy: a critical measure of outcomes for quality improvement. *Ann Surg* 263(2): 286–291. DOI: 10.1097/SLA.0000000000001215.
- Iraji MS (2017). Prediction of post-operative survival expectancy in thoracic lung cancer surgery with soft computing. *J Appl Biomed* 15(2): 151–159. DOI: 10.1016/j.jab.2016.12.001.
- Kuncheva LI (2004). Combining pattern classifiers: Methods and algorithms. John Wiley & Sons.
- Kramer MA (1991). Nonlinear principal component analysis using autoassociative neural networks. *AIChE J* 37(2): 233–243. DOI: 10.1002/aic.690370209.
- Kuan K, Ravaut M, Manek G, Chen H, Lin J, Nazir B, et al. (2017). Deep Learning for Lung Cancer Detection: Tackling the Kaggle Data Science Bowl 2017 Challenge. *arXiv preprint arXiv:1705.09435*.
- Kuo R, Cheng W, Lien W, Yang T (2015). A medical cost estimation with fuzzy neural network of acute hepatitis patients in emergency room. *Comput Methods Programs Biomed* 122(1): 40–46. DOI: 10.1016/j.cmpb.2015.06.006.
- Lu K, An X, Li J, He H (2017). Efficient deep network for vision-based object detection in robotic applications. *Neurocomputing* 245: 31–45. DOI: 10.1016/j.neucom.2017.03.050.
- Luo Y (2017). Recurrent Neural Networks for Classifying Relations in Clinical Notes. *J Biomed Inform* 72: 85–95. DOI: 10.1016/j.jbi.2017.07.006.
- Macías-García L, Luna-Romera JM, García-Gutierrez J, Martínez-Ballesteros M, Riquelme-Santos JC, González-Cámpora R (2017). A Study of the Suitability of Autoencoders for Preprocessing Data in Breast Cancer Experimentation. *J Biomed Inform* 72: 33–44. DOI: 10.1016/j.jbi.2017.06.020.
- Martínez AMC, Mallidi SH, Meyer BT (2017). On the relevance of auditory-based Gabor features for deep learning in robust speech recognition. *Comput Speech Lang* 45: 21–38. DOI: 10.1016/j.csl.2017.02.006.
- Möller A, Sartipy U (2012). Associations between changes in quality of life and survival after lung cancer surgery. *J Thorac Oncol* 7(1): 183–187. DOI: 10.1097/JTO.0b013e3182340abb.
- Oakden-Rayner L, Carneiro G, Bessen T, Nascimento JC, Bradley AP, Palmer LJ (2017). Precision Radiology: Predicting longevity using feature engineering and deep learning methods in a radiomics framework. *Sci Rep* 7(1): 1648. DOI: 10.1038/s41598-017-01931-w.
- Omiotek Z, Burda A, Wójcik W (2013). The use of decision tree induction and artificial neural networks for automatic diagnosis of Hashimoto's disease. *Expert Syst Appl* 40(16): 6684–6689. DOI: 10.1016/j.eswa.2013.03.022.
- Pham T, Tran T, Phung D, Venkatesh S (2017). Predicting healthcare trajectories from medical records: A deep learning approach. *J Biomed Inform* 69: 218–229. DOI: 10.1016/j.jbi.2017.04.001.
- Simonsen DE, Sogaard M, Bozi I, Horsburgh CR, Thomsen RW (2015). Risk factors for postoperative pneumonia after lung cancer surgery and impact of pneumonia on survival. *Respir Med* 109(10): 1340–1346. DOI: 10.1016/j.rmed.2015.07.008.
- Singh A, Tucker CS (2017). A machine learning approach to product review disambiguation based on function, form and behavior classification. *Decis Support Syst* 97: 81–91. DOI: 10.1016/j.dss.2017.03.007.
- Su S-Z, Liu Z-H, Xu S-P, Li S-Z, Ji R (2015). Sparse auto-encoder based feature learning for human body detection in depth image. *Signal Process*, 112: 43–52. DOI: 10.1016/j.sigpro.2014.11.003.
- Sun W, Zheng B, Qian W (2017). Automatic feature learning using multichannel ROI based on deep structured algorithms for computerized lung cancer diagnosis. *Comput Biol Med* 89: 530–539. DOI: 10.1016/j.compbiomed.2017.04.006.
- Tanvetyanon T, Finley DJ, Fabian T, Riquet M, Voltolini L, Kocatürk C, et al. (2015). Prognostic nomogram to predict survival after surgery for synchronous multiple lung cancers in multiple lobes. *J Thorac Oncol* 10(2): 338–345. DOI: 10.1097/JTO.0000000000000400.
- Temurtas H, Yumusak N, Temurtas F (2009). A comparative study on diabetes disease diagnosis using neural networks. *Expert Syst Appl* 36(4): 8610–8615. DOI: 10.1016/j.eswa.2008.10.032.
- Tian K, Shao M, Wang Y, Guan J, Zhou S (2016). Boosting compound-protein interaction prediction by deep learning. *Methods* 110: 64–72. DOI: 10.1016/j.ymeth.2016.06.024.
- van der Burgh HK, Schmidt R, Westeneng H-J, de Reu MA, van den Berg LH, van den Heuvel MP (2017). Deep learning predictions of survival based on MRI in amyotrophic lateral sclerosis. *Neuroimage Clin* 13: 361–369. DOI: 10.1016/j.nicl.2016.10.008.
- Wahab N, Khan A, Lee YS (2017). Two-phase deep convolutional neural network for reducing class skewness in histopathological images based breast cancer detection. *Comput Biol Med* 85: 86–97. DOI: 10.1016/j.compbiomed.2017.04.012.
- Wang H, Zhou Z, Li Y, Chen Z, Lu P, Wang W, et al. (2017). Comparison of machine learning methods for classifying mediastinal lymph node metastasis of non-small cell lung cancer from 18 F-FDG PET/CT images. *EJNMMI Research* 7: 11. DOI: 10.1186/s13550-017-0260-9.
- Yin XX, Zhang Y, Cao J, Wu JL, Hadjiloucas S (2016). Exploring the complementarity of THz pulse imaging and DCE-MRIs: Toward a unified multi-channel classification and a deep learning

- framework. *Comput Methods Programs Biomed* 137: 87–114. DOI: 10.1016/j.cmpb.2016.08.026.
- Yu KH, Zhang C, Berry GJ, Altman RB, Ré C, Rubin DL, Snyder M (2016). Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. *Nat Commun* 7: 12474. DOI: 10.1038/ncomms12474.
- Zhang Q, Xiao Y, Dai W, Suo J, Wang C, Shi J, Zheng H (2016a). Deep learning based classification of breast tumors with shear-wave elastography. *Ultrasonics* 72: 150–157. DOI: 10.1016/j.ultras.2016.08.004.
- Zhang Q, Yang LT, Liu X, Chen Z, Li P (2017). A tucker deep computation model for mobile multimedia feature learning. *ACM TOMM* 13(3): 39. DOI: 10.1145/3063593.
- Zhang Y, Zhang E, Chen W (2016b). Deep neural network for halftone image classification based on sparse auto-encoder. *Eng Appl Artif Intell* 50: 245–255. DOI: 10.1016/j.engappai.2016.01.032.
- Zhu X, Yao J, Huang J (2016a). Deep convolutional neural network for survival analysis with pathological images. In: Tian T, Jiang Q, Liu Y, Burrage K, Song J, Wang Y, et al. (Eds). 2016 IEEE International Conference on Bioinformatics and Biomedicine, pp. 544–547. DOI: 10.1109/BIBM.2016.7822579.
- Zhu Z, Wang X, Bai S, Yao C, Bai X (2016b). Deep learning representation using autoencoder for 3D shape retrieval. *Neurocomputing* 204: 41–50. DOI: 10.1016/j.neucom.2015.08.127.
- Zięba M, Tomczak JM, Lubicz M, Świątek J (2014). Boosted SVM for extracting rules from imbalanced data in application to prediction of the post-operative life expectancy in the lung cancer patients. *Appl Soft Comput* 14: 99–108. DOI: 10.1016/j.asoc.2013.07.016.