



Review

Volume 3 - Issue 3: 109-114 / July 2020

FEATURE SELECTION OF THYROID DISEASE USING DEEP LEARNING: A LITERATURE SURVEY

Amir MEHRNO^{1*}, Recai OKTAS², Mehmet Serhat ODABAS¹

¹Ondokuzmayis University, Institute of Science, Smart Systems Engineering, 55139, Samsun, Turkey

²Ondokuz Mayıs Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, 55139, Samsun, Türkiye

Received: February 28, 2020; **Accepted:** March 17, 2020; **Published:** July 01, 2020


Abstract

The thyroid hormone, which is secreted by the thyroid gland, helps regulate the body's metabolism. Thyroid disorders can range from a small, harmless goiter that does not need to be treated for life-threatening cancer. The most common thyroid problems include abnormal production of thyroid hormones. Overproduction of the thyroid leads to the thyroid and inadequate hormone production leads to hypothyroidism. Although the effects can be unpleasant or uncomfortable, many thyroid problems can be managed well if they are timely diagnosed and treated correctly. In this paper, the diagnosis of thyroid disease is investigated using deep learning based on the imperialist competitive algorithm feature selection method.


Keywords: Thyroid disorders, Diagnosis, Deep Learning, imperialist competitive algorithm, Feature selection method

***Corresponding author:** Ondokuzmayis University, Institute of Science, Smart Systems Engineering, 55139, Samsun, Turkey

E mail: amir.mehrno70@gmail.com (A. MEHRNO)

Amir MEHRNO  <https://orcid.org/0000-0001-9766-5487>

Recai OKTAS  <https://orcid.org/0000-0003-3282-3549>

Mehmet Serhat ODABAS  <https://orcid.org/0000-0002-1863-7566>

Cite as: Mehrno A, Oktas R, Odabas MS. 2020. Feature selection of thyroid disease using deep learning: a literature survey. BSJ Eng Sci, 3(3):109-114.

1. Introduction

The thyroid gland, one of the largest glands in the body, is located in the transverse section of the larynx. Thyroid hormones secreted by this gland have a profound effect on boosting the body's metabolism.

The thyroid secretes two important and dependent hormones, thyroxine and triiodothyronine, respectively. Thyroid secretion is mainly controlled by thyroid stimulating hormone, which is secreted from the anterior pituitary gland. Physiological effects of thyroid hormones include increased cellular metabolic activity, thyroid hormone effect on growth, respiratory changes, and

effect on cardiovascular system, muscle function and sleep. In the thyroid gland, due to the amount of hormonal secretion and symptoms present in the patient, there are generally two types of hypothyroidism and hyperthyroidism.

Thyroid function has a profound effect on many of the essential organs of the body. If the thyroid disorder is not diagnosed promptly, the patient may develop a thyroid attack or myxoma and die (Siti et al., 2008). Thyroid gland problems are so common that almost 1.5 billion people worldwide have thyroid problems. Hypothyroidism is the most common, but

hyperthyroidism is also possible. Subclinical hypothyroidism is also an important condition that affects up to 20% of people over 60 years of age.

Endocrine Clinicians agree that most patients with subclinical hypothyroidism need treatment. Although patients with this disorder can be asymptomatic, some patients have subtle symptoms, including changes in lipid metabolism, cardiac disorders, gastrointestinal, nervous and psychological disorders, and reproduction and increased risk of goitre. To understand subclinical hypothyroidism, physician education and patient awareness are essential (Mounika et al., 2013).

Diagnosis of the disease plays a major role for any physician and is essential. Thyroid disease is one of these diseases and it is difficult to predict without computer technology. The ultimate goal of a pattern recognition system is to achieve the highest possible classification rate for the problem. Since there are no classification algorithms that alone are fully suitable for all problems, combining them is suggested as a solution to increase their efficiency. Deep learning techniques can be used as a physician's assistant in the diagnosis of specific diseases to predict the likelihood of people suffering from diseases (Razia and Rao, 2016).

Deep learning methods have many applications in the medical field. In this case, medical diagnosis is made through the use of deep learning networks. As mentioned before, these include identification, segmentation, classification, prediction, and more. The results of the studied studies show that deep learning methods can be much superior to other high performance algorithms. Therefore, one can easily assume that deep learning is diversifying its applications and will continue to do so (Bakator and Radosav, 2018).

Each diagnosis should be evaluated in the context of the clinical course in question. This is especially important with algorithms where model methods and variables cannot be explicitly presented. Diagnosis of thyroid disease using deep learning algorithms has great potential. From this heuristic meta-analysis, we cautiously state that the accuracy of deep learning algorithms is equivalent to that of healthcare professionals (Liu et al., 2019).

2. Feature Selection

Understanding the features associated with a basic task for using real-world data mining algorithms is very effective. Therefore, many feature selection methods have been proposed to obtain specific features or subsets in the literature for achieving classification and clustering purposes (Kumar and Minz, 2013). The feature selection process is one of the most important steps in the field of data mining and knowledge discovery. Feature selection is a way to eliminate redundant features without reducing classification accuracy (Mousavirad and Komleh, 2013).

Among the features of feature selection, we can: reduce

data size and volume, speed up operations, improve the accuracy of data mining algorithms, and improve our learning and understanding of results. The features in our dataset can be divided into three general categories: First; Irrelevant Properties: Properties that are not loaded with information and have no relevance to data mining goals, which usually result in reduced performance of data mining algorithms. Second; Plug-in Features: Features that are related to other features that is not directly unrelated, such as features that can be used to calculate other features, and third; Related attributes: are attributes that have a great impact on the purpose of data mining and are the main purpose of attribute selection methods (Maghsoud and Ghazvini, 2018). Feature selection is the most important step in data processing because it reduces the dimensions of the data. Feature selection can be part of the criticism that should focus only on related features, such as the PCA method or the modelling algorithm. However, feature selection is usually a separate step in the whole data mining process. There are two different categories of feature selection methods, the filter approach and the packaging approach. The filter approach considers the feature selection process as a pivotal step in learning algorithms. The filter model uses evaluation functions to evaluate the performance of the feature subset classification. There are many evaluation functions such as the importance of attribute, gin, information enhancement, information acquisition and so on. A disadvantage of this approach is that there is no relationship between the feature selection process and the performance of the learning algorithms (Ha and Nguyen, 2016).

3. Deep Learning

Deep learning offers a near-human performance. In-depth learning using back propagation algorithms reveals the complex structure of large datasets to show how a machine must modify its internal parameters that use representations in the previous layer to compute representations in each layer (LeCun et al., 2015).

Deep learning refers to a neural network that has multiple layers of nonlinear processing units. Each layer receives its input sequentially from the output of the previous layer. Using these layers, the network can extract complex hierarchical features from a large volume of data. In recent years, deep learning has made remarkable advances in image classification, detection, object recognition and medical image analysis, where they have achieved excellent results by comparison or sometimes superior to human experts. Among the well-known deep learning algorithms, such as automated encoder stacking, deep Boltzmann machines, and circular neural networks, the most successful for image sharing is circular neural networks (CNNs) first introduced in 1989 by LeCun and the first successful real-world application is hand-written digit recognition by LeCun in 1998,

where it offers a fully five-layer architecture (Zhou et al., 2019).

3.1. How Deep Learning Works?

A standard neural network (NN) that generates a sequence of real-life activities includes many simple and connected processors called neurons. The input neurons are activated by sensors of the periphery, and the other neurons are activated by weight connections from the previous active neurons. Some nerve cells may affect the environment by stimulating action (Schmidhuber, 2015). Most deep learning approaches use neural network architecture, which is why it is also called deep neural network learning models (Figure 1). The word "deep" refers to the number of layers that are hidden in the neural network. Traditional neural networks only have 2-3 layers hidden, while deep neural networks can have up to 150 layers. Deep learning models are taught using large sets of labelled data and neural network architecture that learn features directly from the data without the need to extract manual attributes.

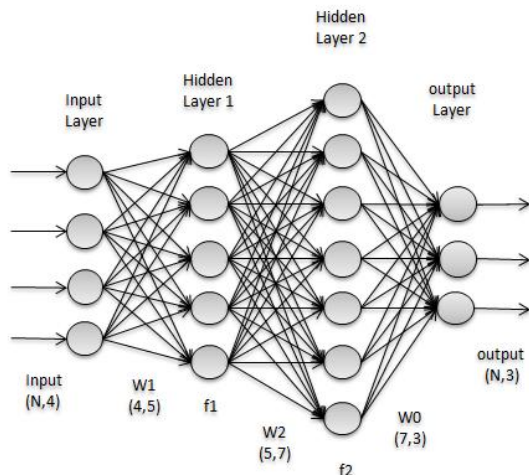


Figure 1. Neural networks, which are organized in layers consisting of a set of interconnected nodes. Networks can have tens or hundreds of hidden layers

CNN's are one of the most successful methods in pattern recognition. These structures use locally trained filters to extract visual features on the input image. This process continues until deep features have been extracted. After these steps, a decision-maker usually decides on these features. A deep convolution neural network generally implies a structure consisting of ring layers, solid layers, and a fully connected network. In this structure, the collection operations are used to extract the attributes, while the complete network is the classifier of these attributes. The fully connected segment can end with a SoftMax output layer for classification purposes. There are many popular network structures created with these layers including AlexNex, OverFeat, GoogLeNet and more. Some techniques have been developed to prevent adding

links, such as Data Enhancement, DropOut and DropConnect (Sarigül et al., 2019).

3.2. Feature Selection with Deep Learning

The main purpose of the proposed method is to analyze the contribution of each input dimension to identifying the important features (inputs) for classification. Sensitivity analysis is typically used to illustrate the importance of individual input dimensions in output with disruption to input-connected weights. The high sensitivity limit for each layer of a multilayer perceptron network determines the optimal number of neurons needed in the network. However, for ReLU, such analysis is often not required because inactive neurons may not be trained at all. For DNN, the sensitivity analysis does not exceed 1 or 2 layers. Hence, to correctly analyze the contribution of an input property, we study its activation potential (on average, over all input training values and hidden neurons) relative to the total activation potential. The greater the potential contribution of activation of an input dimension, the more likely it is to participate in hidden neural activity and, consequently, to classify it (Roy et al., 2015).

Hung and Wang suggested a feature selection method for deep learning to classify scenes. They turned a feature selection approach into a feature reconstruction problem. In a scene classification task, they select features that are more productive than discrimination. However, eliminating discriminatory features may reduce classification accuracy in a typical classification task. Ibrahim et al. have also used DBN, classical feature selection methods, and the unsupervised Active Learning method to process gene expression data. DBN was used to model high dimensional input data and to extract higher-level representation of the input data. Then statistical classical feature selection methods, such as the t-test method, were used to select features from the higher level extracted features. Ibrahim et al's method has used three cascaded independent modulus with separate computational costs. Computational costs can significantly increase when the number of training data is high. DBN is designed to work with a high number of training data samples and it requires feature selection functionality which is suitable for training large datasets, to prevent unnecessary computational cost. Deep learning methods, such as DBM, are usually composed of a high number of nonlinear processing elements that need a high number of training samples. On the other hand, the number of re-choired observations (training samples) should grow exponentially with the number of input features to train a DL method. Through feature selection, the number of input features and consequently the required number of training samples for training a DBM can be reduced. Therefore, feature selection is strongly re-choired to help deep learning methods to be trained with less training data. In this paper, a feature selection method is proposed for DBM to improve its processing ability and reduce the computational cost of

feature selection for DBM (Taherkhani et al., 2018).

4. Algorithm

Algorithms are one of the concerns of people outside computer science. As more and more of human life is done through computers and algorithms are used online and out of the home, people who were once uninterested in computer activity are more concerned about their effects. This concern has become apparent with the rise of popular and academic productions: books, articles, various editions, blog posts, Conferences and research programs are involved with algorithms that do not appear in the textbook introduction to algorithms. For computer graduates, the textbook by Korman et al is a perfect source for learning about algorithms. This massive 1,000-page book contains descriptions of the basic algorithms and data structures they work on. This method teaches how to design an algorithm, determine its performance, and prove its optimality. Knowledge about algorithms is a matter of mathematical proof, and "analyzing an algorithm means predicting the resources that the algorithm needs" (Seaver, 2014).

The algorithm is one of the mathematical and logical problem-solving methods introduced by Khwarizmi. Currently, the basis of many computer programs is Algorithm. The original design of many of the software we use is first created by the algorithm and then transformed into software and eventually into computer language.

"Computer science is nothing but the study of algorithms," says computer scientist Donald Knut. All issues resolved. Unofficially, an algorithm takes a value or set of values as input and some value or set of values as output. Therefore, an algorithm is a sequence of computational steps that converts input to output (Cormen et al., 2009).

4.1. Meta-Heuristic Algorithms

Exploration techniques are not effective for complex problems. There are two main reasons for this:

1. Although heuristic techniques usually (not always) provide an appropriate and acceptable solution to complex problems, they do not guarantee an optimal solution to complex problems.
2. Exploration techniques depend heavily on the experience and mathematical knowledge of the model developer.

Heuristic algorithms are called heuristic algorithms. The word "meta" means "beyond" or "higher level". Metaphors often do better than simple discoveries. Metaphorical algorithms use some random transactions and local search. As metaphorical algorithms perform a more thorough search, they have quickly become a desirable way of producing solutions to complex real-life problems that precise methods cannot solve. There are several classifications available for metaphor algorithms in the literature (Memari et al., 2017).

Metaphorical algorithms have been an interesting area for scholars, scientists and academics because of their unique and remarkable features and capabilities. Their ability to solve and provide optimal solutions to versatile domain problems without precise detail and problem definition is marginal to traditional techniques. Most metaheuristic algorithms, inspired by some real-world phenomena, are generally a natural optimization method. Over the past few decades, a number of meta-heuristic algorithms have been introduced and applied to different problems of different domains. This article aims to provide researchers with a comprehensive list of such algorithms that have been developed so far. These algorithms are listed in alphabetical order by the author and give a brief discussion of the method. The purpose of this work is to provide an introduction to new and interested researchers (Rajpurohit et al., 2017).

4.2. Imperialist Competitive Algorithm (ICA)

Atashpaz et al. have introduced ICA as an evolutionary optimization goal for modelling processes. Imperialist Competitive Algorithm (ICA) that results from imperialist competition, like other evolutionary algorithms, begins with the initial population known as the state. A country contains all kinds of colonies and imperialists that together form the empires. The imperialist countries are trying to conquer the other countries and make them colonies. Also, imperialist countries compete fiercely for the sake of conquering other countries. The imperialist rivalry between these empires forms the proposed evolutionary algorithm. During this competition, the weakest empires fall and the more powerful gain more power. Imperialist competition turns to conditions in which there is only one empire and the colonies have the same function and value of power as the imperialist, since a large part of the colonies can replace the imperialists (Razmjooy et al., 2013).

Specifically, ICA is used as the social counterpart of Genetic Algorithms (GA). ICA is a mathematical model and computer simulation of human social evolution. Figure 2 shows the flow chart of the imperialist competitive algorithm. Countries are like chromosomes in GA. After evaluating the performance of each country's costs, some of the best ones (in optimization terms, the lowest-cost countries) are selected to control primary empires with control over other countries (colonies). All colonies are divided among the early imperialists according to their power. The power of any country, such as the value of fitness in GA, is inversely related to its cost. The initial imperialist form is along with their colonies to form the first empires (Abedi et al., 2011).

4.3. Feature Selection ICA Algorithms

This algorithm starts with a few initial populations. Some of the best elements of the population are chosen as colonists. The rest of the population is also a colony. The colonists by their power kill these colonies in a certain way. The total power of an empire depends on its two constituent parts, the colonial state and its colonies. With

the formation of the early empires, colonial competition between them began. Any empire that fails to succeed in colonial competition and build its power is eliminated from colonial competition, so the survival of an empire depends on its ability to attract rival colonies. As such, the power of the larger and weaker empires gradually disappears during colonial competition. First, the first empires were created. All empires try to conquer the colonies of other empires. This struggle will gradually reduce the power of the weak empires and increase the power of the stronger empires, select the imperialist competition for modelling the weakest colony, the weakest empire, and then create competition among all the empires for colonization.

Each country has an n-dimensional array that contains all the attributes and the attributes that are selected have a value of one and those that are not selected receive a value of zero. It creates a set of binary numbers for a set of attributes and can be recognized as a subset of selected attributes (Maghsoudi and Ghazvini, 2018).

To start the race, after selecting a colony, the possession probability of each empire is found. The normal total cost of an empire is simply obtained by;

$$N.T.Cn = T.Cn - \max\{T.Ci\}$$

Where $T.Cn$ and $N.T.Cn$ are respectively the total cost and the most common cost of the first empires. At a normal total cost, the probability of any empire, p_{empn} is calculated as follows:

$$p_{empn} = \frac{|N.T.Nc|}{\sum_{i=1}^{imp} N.C.Ti}$$

The Roulette wheel method has been used to assign the above Colony to the empires (Mousavirad et al., 2012).

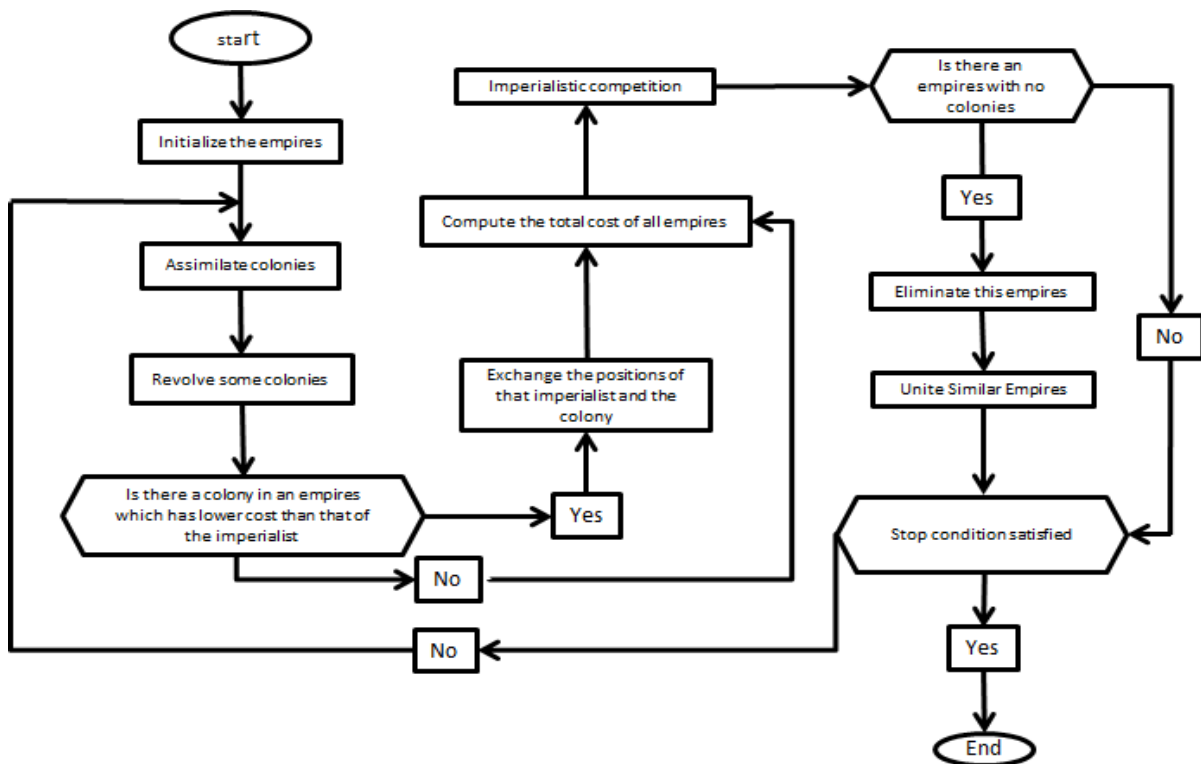


Figure 2. Imperialist-competitive-algorithm-flowchart.

6. Conclusion

The purpose of our work is to investigate specific machine learning techniques that can be used to diagnose thyroid diseases. Many achievable analyzes have been charted and used in recent years to adequately and correctly diagnose thyroid disease. The analysis shows that different technologies are used in all articles that show different accuracy. Most research papers have shown that the deep learning algorithm is superior to other techniques. On the other hand, we also add that

using support vector machine algorithm and decision tree researchers around the world have been very successful in diagnosing thyroid diseases, but it is suggested that the number of parameters used by patients to diagnose diseases is suggested. Thyroid decreases. More features mean that the patient has to perform more clinical trials that are both cost-effective and time-consuming. Therefore, there is a need to develop these types of algorithms and models for predicting thyroid disease that requires at least several individual parameters to

diagnose thyroid disease and save the patient time and money.

The future development of deep learning promises more applications in various medical fields, especially in the field of medical diagnosis. However, under current circumstances, it is not clear that deep learning can replace the role of physicians / physicians in medical diagnosis. So far, deep learning can provide good support to medical professionals.

Conflict of interest

The authors declare that there is no conflict of interest.

References

- Abdi B, Mozafari H, Ayob A, Kohandel R. 2011. Imperialist competitive algorithm and its application in optimization of laminated composite structures. *European J Sci Res*, 55: 1450-216.
- Bakator M, Radosav D. 2018. Deep Learning and medical diagnosis: a review of literature. *Multimodal Techs and Interact*, 2(3): 47. DOI: 10.3390/mti2030047.
- Cormen TH, Leiserson ChE, Rivest RL, Stein C. 2009. *Introduction to algorithms*, 3rd Edition The MIT Press, ISBN 978-0-262-03384-8.
- Ha VS, Nguyen HN. 2016. Credit scoring with a feature selection approach based deep learning. *MATEC Web of Conferences* 54: 05004. DOI: 10.1051/mateconf/20165405004.
- Kumar V, Minz S. 2014. Feature selection: A literature review. *Smart Comp Rev*, 4: 211-229. DOI: 10.1145/2740070.2626320.
- LeCun Y, Bengio Y, Hinton G. 2015. Deep Learning. *Nature*, 521: 436-444. DOI: 10.1038/nature14539.
- Liu X, Faes L, Kale A, Wagner S, Fu D, Bruynseels A, Mahendiran T, Moraes G, Shamdas M, Kern C, Ledsam J, Schmid M, Balaskas K, Topol E, Bachmann L, Keane P, Denniston A. 2019. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis 1. *The Lancet Dig Health*, 1(6): e271-e297. DOI: 10.1016/S2589-7500(19)30123-2.
- Magsudi M, Gazvini M. 2018. Metadata algorithms for dimensionality reduction and feature selection 3rd International Conference on Combination, Cryptography and Computing, Iran University of Science and Technology.
- Memari A, Robiah A, Abdul Rahim Abd. 2017. Metaheuristic algorithms: guidelines for implementation. *J Soft Comput and Dec Sup Sys*, 4: 1-6.
- Mousavirad SJ, Akhlaghian Tab F, Mollazade K. 2012. Application of imperialist competitive algorithm for feature selection: a case study on bulk rice classification. *Int J Comp App*, 40: 41-48. DOI: 10.5120/5068-7485.
- Musavirad SJ, Ebrahimpoor-Komleh H. 2013. Feature selection using modified imperialist competitive algorithm. *ICCKE*, 2013: 400-405. DOI: 10.1109/ICCKE.2013.6682833.
- Rajpurohit J, Sharma TK, Abraham A, Vaishali. 2017. Glossary of metaheuristic algorithms. *Int J Comp Inf Sys and Indust Manag App*, 9: 181-205.
- Razia SH, Rao M. 2016. Machine learning techniques for thyroid disease diagnosis: a review. *Ind J Sci and Tech*, 9(28): 1-9. DOI: 10.17485/ijst/2016/v9i28/93705.
- Razmjoooy N, Musavi BS, Soleymani F. 2013. A hybrid neural network imperialist competitive algorithm for skin color segmentation. *Math and Comp Model*, 57(3): 848-856.
- Roy D, Kodukula SRM, Chalavadi KM. 2015. Feature selection using Deep Neural Networks. 1-6. 10.1109/IJCNN.2015.7280626.
- Sarigül M, Özyildirim BM, Avci M. 2019. Differential Convolutional neural network, *Neu Net*, 116: 279-287.
- Schmidhuber J. 2014. Deep Learning in neural networks: an overview. *Neu Net*, 61. DOI: 10.1016/j.neunet.2014.09.003.
- Seaver N. 2014. *Media in transition* 8, Cambridge, MA.
- Siti F, Shurehdeli MA, Teshneh Lab M. 2008. Diagnosis of thyroid disease using probabilistic neural networks and Genetic Algorithm: 2nd Joint Congress on Fuzzy and Intelligent Systems. Iran.
- Taherkhani A, Cosma G, McGinnity TM. 2018. Deep-FS: A feature selection algorithm for Deep Boltzmann Machines. *Neurocomputing* 322: 22-37.
- Zhou T, Ruan S, Canu S. 2019. A review: Deep learning for medical image segmentation using multi-modality fusion. *Array*, 3-4: 100004.