ICIC Express Letters
Volume **12**, Number **7**, July **2018**
ICIC International ©2018 ISSN 1881-803X
pp. 645–653

TRANSFER LEARNING USING ROUGH SETS FOR MEDICAL DATA CLASSIFICATION

Mabruka Almasoudi Ahmed, Yasser Fouad Hassan and Ashraf Elsayed

Department of Mathematics and Computer Science
Faculty of Science
Alexandria University
22 El-Guish Road, El-Shatby, Alexandria 21526, Egypt
Mabroka632010@yahoo.com; { y.fouad; ashraf.elsayed }@alexu.edu.eg

Received December 2017; accepted March 2018

Abstract. *The machine learning plays the key roles in many artificial intelligence domains including classification, regression, and clustering. The traditional machine learning methods suppose that the training and test samples are drawn from the same feature space and the same distribution. With the modification of the distribution, the traditional machine learning methods need to reconstruct the models using newly collected training samples. In the real world, it is impossible or expensive to remember and label the needed training samples and rebuild the models. To address the problem, the transfer learning is proposed. Transfer learning is the capability of a system to acknowledge and apply knowledge and skills learned in past domains or tasks to new tasks or domains, which share some commonality. In this paper, the idea of transfer learning will be applied to transferring knowledge from source data set to target data set, the source data set will be used as training data and the target data set will be used as testing data to classify target data set using the features that have been transferred from the source data and the features of target data. The experimental results show that the proposed method has higher classification accuracy.*
Keywords: Transfer learning, Source data, Target data, Rough set, Set of rules

1. **Introduction.** The main idea behind the transfer learning is to excerpt labeled data or extract knowledge from some related domains to help a machine learning algorithm to achieve greater performance in the domain of interest. Thus, transfer learning can be referred to as a different strategy for learning models with minimal human supervision, compared to semi-supervised and active learning [1]. Traditional machine learning classification algorithm assumes that the training and test data are drawn from the same distribution. In contrast, transfer learning makes the domains, tasks and distribution used in training and testing be various [2]. Recently, transfer learning techniques have been applied in many real world applications. In [3,4], transfer learning techniques are proposed to learn text data across domains. In [5,6], transfer learning techniques are applied to biological fields. [7,8] are proposed for computer vision and image processing. In [9,10], transfer learning techniques are proposed to solve collaborative filtering problems. Also, transfer learning techniques are used for early diagnosis of Alzheimer's disease in [11]. Transfer learning has been used for computer-aided detection, where information from images of natural scenes has been successfully transferred for detection of lung lesions [12].

Rough set theory is a mathematical approach for handling vagueness and uncertainty in data analysis. A rough set is characterized by a pair of precise concepts, called lower and upper approximations, generated using object indiscernibilities. The essential advantage of rough set theory is that it does not require any preliminary or additional information

DOI: 10.24507/icicel.12.07.645

about data [13]. Also, its methodology is concerned with the classification and analysis of imprecise, uncertain or incomplete information and knowledge [14]. In the course of recent years, a rough set theory has turned into a most interesting tool for researchers and has been connected to numerous medical domains [15,16]. In [17] a knowledge detection method based on rough set for diseases of congenital abnormality is presented. [18] introduced a rough set method to attribute selection and prediction to breast cancer data set. [19] presented a rough set-based inference system for cardiac disorders classification. Traditional rough set for prediction of cardiovascular pathology was used in [20]. [21] introduced a rough set method to hybrid rough-objective soft set classification system for medical data set classification. Also, rough set has been used for cancer classification in hybrid rough set and heterogeneous ensemble classifiers model for cancer classification in [22].

This paper presents the transfer learning technique to evaluate the performance of transfer learning for classification of a medical data set. We have used the feature representation approach to transfer the knowledge between two data sets that aim to find a good feature representation that reduces the difference between the source and the target data. Firstly, extract these features from source data and target data. Secondly, transfer source data features to target data. Finally, use it with features of target data to classify target data. The experimental results show that the proposed transfers learning overpass the results achieved by the rough set and decision tree. The rest of this paper is organized as follows. Section 2 presents definitions of transfer learning and rough set. The detailed introduction of our improved method is described in Section 3. Section 4 displays the experiments. Results and discussion are presented in Section 5 and Section 6 gives the conclusions and future research of our work.

2. **Concepts and Definitions.**

2.1. **Transfer learning.**

Transfer learning aims to exploit the knowledge in the source domains to promote the learning tasks in the target domains, and has attracted extensive research interests recently [23].

2.1.1. *A categorization of transfer learning techniques.* There are three main research issues in transfer learning: (1) What to transfer; (2) How to transfer; (3) When to transfer. "What to transfer" requests which part of knowledge can be transferred across domains or tasks. Some knowledge is specific for individual domains or tasks, and some knowledge may be common between different domains such that they may help improve performance for the target domain or task. After discovering which knowledge can be transferred, learning algorithms need to be developed to transfer the knowledge, which corresponds to the "How to transfer" issue. "When to transfer" asks in which situations, transferring skills should be done. Likewise, we are interested in knowing in which situations, knowledge should not be transferred [24].

2.1.2. *Transfer learning approaches.* We have different approaches to transfer learning which can be summarized into four cases based on "What to transfer". The first context is instance transfer approach which assumes that certain parts of the data in the source domain can be reused for learning in the target domain by re-weighting. Instance reweighting and importance sampling are two major techniques in this context. A second case is feature representation transfer approach that uses to learn a "good" feature representation for the target domain. In this case, the knowledge used to transfer across domains is encoded into the learned feature representation. With the new feature representation, the performance of the target task is expected to improve significantly. A third case is parameter transfer approach, which assumes that the source tasks and the target tasks share some parameters or prior distributions of the hyperparameters of the models. The transferred knowledge is encoded into the shared parameters or priors. Thus, by discovering the shared parameters or priors, knowledge can be transferred across tasks. Finally, the last case is the relational knowledge transfer, which deals with transfer learning for relational domains. The basic assumption behind this context is that some relationships among the data in the source and target domains are similar. Thus, the knowledge to be transferred is the relationship between the data [25].

2.2. **Rough set theory.** Rough set theory is a new mathematical and artificial intelligence technique introduced by Pawlak in the early 1980's [16]. The technique is particularly suited to reasoning about imprecise or incomplete data, and discovering relationships among them. The main advantage of rough set theory is that it does not require any preliminary or additional information about data like probability in statistics, or the value of possibility in fuzzy set theory. Recently, there has been an increasing interest in rough set theory among researchers in modern intelligent information systems [26].

2.2.1. *The main concepts of rough set theory.*

Information system or information table

Information table is a table that contains the objects as rows and attributes as columns. It is used in the representation of data that will be utilized by rough set, where each object has a given amount of attributes.

$$\mathrm{M} = (\mathrm{O}, \mathrm{A} \cup \{\mathrm{d}\}) \tag{1}$$

where O is a non-empty finite set of objects and A is a non-empty finite set of attributes and $\{\mathrm{d}\}$ is decision attribute where $\mathrm{d} \notin \mathrm{A}$.

Indiscernibility relation

Indiscernibility relation is a central concept in rough set theory, and is considered as a relation between two objects or more, where all the values are identical in relation to a subset of considered attributes. Indiscernibility relation is an equivalence relation, where all identical objects of set are considered as elementary.

Reduction attributes

The ideas of attribute reduction are closely related to feature selection in many other fields. Many researchers have proposed a lot of methods for feature selection. Rough set theory, with its idea of reducts, becomes an attractive and potential approach for this problem [27]. In many application problems it is often necessary to maintain a concise form of the information system. One way to do this is to search a minimal representation of original data set. Reduct is a minimal subset R of initial attribute set C (conditional) such that for a given set of decision attributes D [28].

Set of rules

In this paper we have generated the set of rules based on set of reduct. IF-THEN rules are constructed by reading of the values for each attribute in the reduct (IF-part called antecedent or premise, e.g., a1 = v1 AND a2 = v2, where a1 and a2 are attributes in a reduct and v1 and v2 are attribute values) and associating them with one or more decision classes (THEN-part called consequent, e.g., d = d1 OR d = d2, where d is the decision attribute and d1 and d2 are decision classes). The THEN-part will only include one decision class unless the decision class is rough with respect to the attributes in the reduct [29].

3. **Method.** In this paper, we have proposed a new transfer learning algorithm based on a rough set theory for medical data classification. We used the feature representation approach to transfer the knowledge between two data sets as source data $\mathrm{D_S}$ and target data $\mathrm{D_T}$ that aim to find a good features representation that reduce the difference between them where the feature space of the source data $\mathrm{X_S}$ is different from the feature space of

the target data X_T where $X_S \neq X_T$. First, the features from source data D_S and target data D_T are extracted. Secondly, transfer features of source data D_S to target data D_T. Finally, use the features of source data D_S with the features of target data to classify target data D_T.

3.1. **Extract the features from source data and target data.** We have used rough set theory as a feature selection technique to generate a subset of features from the original features to extract the set of rules from source and target data by: first, transform source data into a decision table $S = (U, A \cup \{d\})$, where U is a non-empty finite set of objects and A is a non-empty finite set of attributes and $\{d\}$ is decision attribute where $d \notin A$, compute set of reduct for attributes of source data using the genetic algorithm to search a minimal representation of original data set, generate the set of rules S_S based on reduct. Second, transforming target data into a decision table $T = (U, A \cup \{d\})$, where U is a non-empty finite set of objects and A is a non-empty finite set of attributes and $\{d\}$ is decision attribute where $d \notin A$, compute set of reduct for attributes of target data using the genetic algorithm to keep attributes that representation of original data set. Generate the set of rules S_T based on reduct.

3.2. **Transfer the features as the knowledge to target data.** To Transfer the set of rules from source data to target data we have three cases to transfer (1) renumbering for attributes numbers of rules of source data based on target data attributes numbers for common attributes between a source data and target data Ac. (2) renumbering for attributes numbers of rules of source data based on target data attributes numbers for attributes that are in target data and they are not in source data by its high correlation with target data attributes numbers Ar. (3) put OR instead of AND before attributes of rules of source data which are un-relevant attributes and that are in source data and they are not in target data Au.

3.3. **Classify the target data.** To evaluate the classification of target data by computing the accuracy we have merged the transfer rules of source data with target data rules to use them to classify target data.

The proposed algorithm for transfer learning is described as follows:
Input: $S_S \Rightarrow$ set of rules of D_S, $S_T \Rightarrow$ set of rules of D_T
$S_S \Rightarrow D_S \Rightarrow$ reduct.
$S_T \Rightarrow D_T \Rightarrow$ reduct.
$A_T \Rightarrow$ attributes of D_T.
$A_S \Rightarrow$ attributes of D_S.
For i $\Rightarrow$ size (S_S)
$A_T \Rightarrow$ attributes numbers S_S [i]
If attributes of A_T = attributes of A_S **then** renumbering target data attributes based on source data attributes numbers $\Rightarrow$ Ac.
If attributes are in A_T and they are not in A_S **then** renumbering for attributes numbers of rules of source data based on its high correlation with target data attributes numbers $\Rightarrow$ Ar.
If attributes are in A_S and they are not in A_T **then** put OR instead of AND before attributes of rules of source data which are un-relevant attributes $\Rightarrow$ Au.
End
$S_K = \{Ac, Ar, Au\}$.
Merge S_K with $S_T \Rightarrow S_M$.
Classify D_T by $S_M \Rightarrow$ accuracy.
Output: accuracy.

4. **Experiments.** All the experiments in this paper are conducted on a machine with Dual Core T4400 and 2.20GHz CPU and 4GB memory running in Windows 7. The proposed algorithm has been implemented in MATLAB® 2016b and we have generated the set of rules of source data and target data using (ROSETTA system: A Rough Set Toolkit for Analysis of Data).

We test the proposed method on three data sets that contain six medical data sets that are directly obtained from UCI Machine Learning Repository and we have split source data and target data to 75% training data and 25% testing data for the three data sets.

The first data set uses Indian Liver Patient Data set as the source data that contain 11 attributes and 583 objects [30], and Liver Disorders Data set as target data that contain 7 attributes and 345 objects [31]. There are three common attributes between target data and source data as shown in Table 1.

Table 1. Source and target attributes of data 1

	Indian Liver Patient Data set attributes as source data	Liver Disorders Data set attributes as target data
1	Age	Mean corpuscular volume
2	Gender	Alkaline phosphotase
3	Total Bilirubin	Alanine aminotransferase
4	Direct Bilirubin	Aspartate aminotransferase
5	Alkaline phosphotase	Gamma-glutamyl transpeptidase
6	Alanine aminotransferase	Drinks number of half pint equivalents of alcoholic
7	Aspartate aminotransferase	Class
8	Total Protiens	
9	Albumin	
10	Albumin and Globulin Ratio	
11	Class	

The second data set uses Indians Chronic Kidney Disease as the source data that contain 25 attributes and 400 objects [32], and Statlog (Heart) Data set as target data set that contains 14 attributes and 270 objects [33]. In this data set there are two common attributes between target data and source data as shown in Table 2.

The third data set uses Hepatitis data set as the source data that contain 20 attributes and 155 objects [34], and Indian Liver Patient Data set as the target data that contain 11 attributes and 583 objects [30]. In this data set there are six common attributes between target data and source data as shown in Table 3.

5. **Results and Discussion.** Transfer learning research in artificial intelligence is commonly divided into two sub fields. One focuses on the classification, regression, and clustering tasks involving different domains, and the other one on reinforcement learning with different tasks and/or different domains [35]. In this paper, our method adopts a new classification strategy where the target data classify by source data that has different distribution from target data by using the same and different features between two data sets (source and target data) where the experimental evaluation demonstrates that the proposed method achieves significant efficiency improvements in transfer learning using rough set. There are two important factors that would impact the algorithm performance. First, the number of attributes of the source should be more than the number of target data attributes to find a return for each attribute of target data with the attribute of source data. Second, the data type of source data should be the same as target data type to make the process of transfer effective wherein data 2 although there are only two

Table 2. Source and target attributes of data 2

	Indians Chronic Kidney Disease attributes as source data	Statlog (Heart) Data set attributes as target data
1	Age	Age
2	Blood pressure	Sex
3	Specific Gravity	chest pain type
4	Albumin	Resting blood pressure
5	Sugar	Serum cholestoral
6	Red Blood Cells	Fasting blood sugar
7	Pus Cell	Resting electrocardiographic results
8	Pus Cell clumps	Maximum heart rate achieved
9	Bacteria	Exercise induced angina
10	Blood Glucose Random	Old peak
11	Blood Urea	The slope of the peak exercise ST segment
12	Serum Creatinine	Number of major vessels (0-3) colored by flourosopy
13	Sodium	Thal
14	Potassium	Class
15	Hemoglobin	
16	Packed Cell Volume	
17	White Blood Cell Count	
18	Red Blood Cell Count	
19	Hypertension	
20	Diabetes Mellitus	
21	Coronary Artery Disease	
22	Appetite	
23	Pedal Edema	
24	Anemia	
25	Class	

common attributes between target data and source data the process of transfer gives the high accuracy of classifications because almost attributes that we have matched between the source data and target data have the same type of data.

The performance of classification is given in terms of accuracy that is the proportion of the correct classified objects to the total number of object

$$\text{Accuracy} = \text{TP} + \text{TN}/\text{N} \tag{2}$$

where TP is the number of true positive, TN is the number of true negative, and N is the total number of objects in the test set.

As shown in Table 4 and Figure 1, we have compared our method with classifying target data using rough set and decision tree where we have split this data into 75% training and 25% testing for both methods. Our proposed method achieves better results than rough set and decision tree methods in performance accuracy measures which validate the efficacy of our method. Specifically, our method can achieve a classification accuracy of 63.1% in data 1, which is significantly better than rough set and decision tree methods which achieve 47.1% and 61.6% respectively. In addition, our proposed method also achieves better results than rough set and decision tree methods in data 2 and data 3 where it has classification accuracy of 90% in data 2 compared with 55.8% and 73.1% for rough set and decision tree methods and it recorded classification accuracy in data 3

Table 3. Source and target attributes of data 3

	Hepatitis data set attributes as source data	Indian Liver Patient Data set attributes as target data
1	Age	Age
2	sex	Gender
3	STEROID	Total Bilirubin
4	ANTIVIRALS	Direct Bilirubin
5	FATIGUE	Alkaline Phosphotase
6	MALAISE	Alamine Aminotransferase
7	ANOREXIA	Aspartate Aminotransferase
8	LIVER BIG	Total Proteins
9	LIVER FIRM	Albumin
10	SPLEEN PALPABLE	Albumin and Globulin Ratio
11	SPIDERS	Class
12	ASCITES	
13	VARICES	
14	BILIRUBIN	
15	ALK PHOSPHATE	
16	SGOT	
17	ALBUMIN	
18	PROTIME	
19	HISTOLOGY	
20	Class	

Table 4. The comparison between classification accuracy for different methods

	Accuracy		
	Classification method		
	Decision tree	Rough set	Transfer learning
Data 1	61.6%	47.1%	**63.1%**
Data 2	73.1%	55.8%	**90%**
Data 3	65.5%	74.6%	**81.9%**

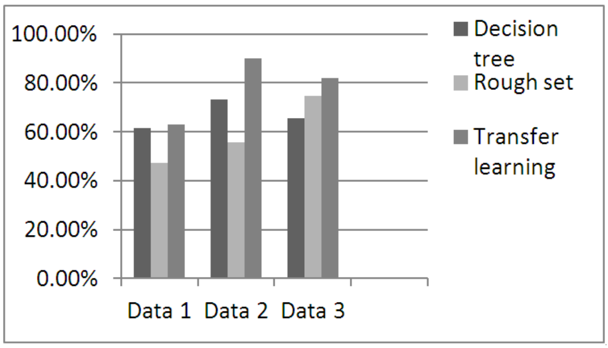

Figure 1. The comparison between rough set, decision tree and transfer learning for classification accuracy

81.9% compared with 74.6% and 65.5% by rough set and decision tree methods. It implies that transfer knowledge from source data can effectively use in classifying target data.

6. **Conclusions.** This paper presents the transfer learning classification method to classify data set from another data set and we focus on the feature representation transfer

learning setting to extract the feature from source data using rough set theory and use them to classify target data. The key idea is to first generate set of rules from source data as the features then transfer it with a set of rules of target data to classify target data and compare them with classifying the same data by rough set and decision tree. Experimental results on datasets demonstrate the validity and efficiency of our method. In the future, we plan to apply the idea of transfer learning on two data sets using rough set theory and neural network for data classification using multi source domains. Also we plan to apply the idea of transfer learning on two data sets using fuzzy rough set for data classification.

REFERENCES

[1] C. C. Aggarwal, *Data Classification: Algorithms and Applications*, CRC Press, 2014.

[2] Y. Chu, Z. Wang, M. Chen, L. Xia, F. Wei and M. Cai, Transfer learning in large-scale short text analysis, *International Conference on Knowledge Science, Engineering and Management*, pp.499-511, 2015.

[3] W. Dai, G. R. Xue, Q. Yang and Y. Yu, Co-clustering based classification for out-of-domain documents, *Proc. of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp.210-219, 2007.

[4] F. Zhuang, P. Luo, C. Du, Q. He and Z. Shi, Triplex transfer learning: Exploiting both shared and distinct concepts for text classification, *Proc. of the 6th ACM International Conference on Web Search and Data Mining*, pp.425-434, 2013.

[5] F. Zhuang, P. Luo, C. Du, Q. He and Z. Shi, Evigan: A hidden variable model for integrating gene evidence for eukaryotic gene prediction, *Bioinformatics*, vol.24, no.5, pp.597-605, 2008.

[6] D. L. Silver and R. E. Mercer, Sequential inductive transfer for coronary artery disease diagnosis, *International Joint Conference on Neural Networks*, pp.2635-2641, 2007.

[7] Y. Zhu, Y. Chen, Z. Lu, S. J. Pan, G. R. Xue, Y. Yu and Q. Yang, Heterogeneous transfer learning for image classification, *AAAI Conference on Artificial Intelligence*, pp.1304-1309, 2011.

[8] S. Chopra, S. Balakrishnan and R. Gopalan, DLID: Deep learning for domain adaptation by interpolating between domains, *ICML Workshop on Challenges in Representation Learning*, vol.2, no.6, 2013.

[9] J. Shi, M. Long, Q. Liu, G. Ding and J. Wang, Twin bridge transfer learning for sparse collaborative filtering, *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pp.496-507, 2013.

[10] B. Li, Q. Yang and X. Xue, Can movies and books collaborate? Cross-domain collaborative filtering for sparsity reduction, *International Joint Conference on Artifical Intelligence*, vol.38, no.4, pp.2052-2057, 2009.

[11] B. Cheng, M. Liu, D. Shen, Z. Li and D. Zhang, Multi-domain transfer learning for early diagnosis of Alzheimer's disease, *Neuroinformatics*, pp.1-18, 2017.

[12] H. C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues and R. M. Summers, Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning, *IEEE Trans. Medical Imaging*, vol.35, no.5, pp.1285-1298, 2016.

[13] H. S. Own, K. O. Alyahya, W. I. Almayyan and A. Abraham, Rough set-BPSO model for predicting vitamin D deficiency in apparently healthy Kuwaiti women based on hair mineral analysis, *Neural Computing and Applications*, pp.1-16, 2016.

[14] S. Rissino and G. Lambert-Torres, Rough set theory – Fundamental concepts, principals, data extraction, and applications, in *Data Mining and Knowledge Discovery in Real Life Applications*, 2009.

[15] S. Chowdhuri, P. Roy, S. Goswami, A. T. Azar and N. Dey, Rough set based ad hoc network: A review, *International Journal of Service Science, Management, Engineering, and Technology (IJSSMET)*, vol.5, no.4, pp.66-76, 2014.

[16] Z. Pawlak, Rough sets, *International Journal of Parallel Programming*, vol.11, no.5, pp.341-356, 1982.

[17] S. Tsumoto, Mining diagnostic rules from clinical databases using rough sets and medical diagnostic model, *Information Sciences*, vol.162, no.2, pp.65-80, 2004.

[18] A. E. Hassanien and J. M. Ali, Rough set approach for generation of classification rules of breast cancer data, *Informatica*, vol.15, no.1, pp.23-38, 2004.

[19] S. Mitra, M. Mitra and B. B. Chaudhuri, A rough-set-based inference engine for ECG classification, *IEEE Trans. Instrumentation and Measurement*, vol.55, no.6, pp.2198-2206, 2006.

[20] J. G. Bazan, S. Bazan-Socha, S. Buregwa-Czuma, P. W. Pardel and B. Sokolowska, Prediction of coronary arteriosclerosis in stable coronary heart disease, *Advances in Computational Intelligence*, pp.550-559, 2012.

[21] H. H. Inbarani, S. U. Kumar, A. T. Azar and A. E. Hassanien, Hybrid rough-bijective soft set classification system, *Neural Computing and Applications*, pp.1-12, 2016.

[22] M. E. Helal, M. Elmogy and R. M. El-Awady, Hybrid rough set and heterogeneous ensemble classifiers model for cancer classification, *International Conference on Advanced Intelligent Systems and Informatics*, pp.513-522, 2016.

[23] X. Hu, J. Pan, P. Li, H. Li, W. He and Y. Zhang, Multi-bridge transfer learning, *Knowledge-Based Systems*, vol.97, pp.60-74, 2016.

[24] S. J. Pan and Q. Yang, A survey on transfer learning, *IEEE Trans. Knowledge and Data Engineering*, vol.22, no.10, pp.1345-1359, 2010.

[25] K. Weiss, T. M. Khoshgoftaar and D. Wang, A survey of transfer learning, *Journal of Big Data*, vol.3, no.1, p.9, 2016.

[26] F. Zarandi, M. Hossein and A. Kazemi, Application of rough set theory in data mining for decision support systems (DSSs), *Journal of Optimization in Industrial Engineering*, pp.25-34, 2010.

[27] N. A. Tuan, Using rough set in feature selection and reduction in face recognition problem, *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pp.226-233, 2005.

[28] P. V. Anitha and P. Venkatesan, Feature selection by rough-quick reduct algorithm, *International Journal of Innovative Research in Science, Engineering and Technology*, vol.2, no.8, 2013.

[29] T. R. Hvidsten, *A Tutorial-Based Guide to the ROSETTA System: A Rough Set Toolkit for Analysis of Data*, 2010.

[30] *http://archive.ics.uci.edu/ml/datasets/ILPD+%28Indian+Liver+Patient+Dataset%29.*

[31] *http://archive.ics.uci.edu/ml/datasets/Liver+Disorders.*

[32] *http://archive.ics.uci.edu/ml/datasets/Chronic_Kidney_Disease.*

[33] *http://archive.ics.uci.edu/ml/datasets/Statlog+%28Heart%29.*

[34] *http://archive.ics.uci.edu/ml/datasets/Hepatitis.*

[35] T. T. Nguyen et al., Scalable transfer learning in heterogeneous, dynamic environments, *Artificial Intelligence*, 2015.