A Deep Learning Based Transfer Learning Framework for Healthcare Text Analytics

Research-in-Progress

Long Xia **Alan Wang**

Weiguo Fan **Wenqi Shen**

Abstract

Transfer learning has been widely utilized in many real-world applications, including text analytics. However, our literature review suggests that it could be beneficial to conduct a more comprehensive study on transfer learning and its application. Grounded by theory of transfer of learning, we identify and re-organize different types of transfer learning approaches in a more systematical and theoretical manner. We believe different transfer learning approaches are complementary for one another rather than substitutive. As deep learning has achieved great successes and been applied in solving various healthcare related tasks, we propose a design framework that integrates deep learning model and three different transfer learning approaches and evaluate the performance of our framework in a healthcare text analytics task. Our preliminary results confirm the improved performance by incorporating the three transfer learning approaches individually and in combined. We also discuss the theoretical and practical implications and research plans.

Keywords: Healthcare analytics, transfer learning, knowledge transfer, deep learning, artificial intelligence, theory of transfer of learning

Introduction

In the last few decades, artificial intelligence (AI) has made remarkable progresses in various tasks and domains (LeCun et al. 2015). Much of the progress has come from the recent advancement in machine learning, especially in deep learning. A major advantage of deep learning is that it directly takes raw data as its input and automatically discovers the underlining features. Deep learning has achieved unprecedented performance gains in many tasks, such as computer vision and natural language processing, spanning various domains. It has also been extensively adopted in the healthcare research. A literature search shows two major medical domains that utilize deep learning approaches: medical image processing (Shin et al. 2016; Wang et al. 2016) and medical text analytics (Cheng et al. 2016; Liang et al. 2014). More recent studies applied deep learning in cancer diagnosis and staging. The results show that the deep learning-based methods can significantly reduce the workload of practitioners, and in the meantime, achieve a level of competence comparable to human experts (Esteva et al. 2017; Litjens et al. 2016). The recent success of deep learning also drew great attention from the industry. Many deep learning algorithms have been incorporated into commercial systems. As a result, three top AI experts claim, "Deep learning is making major advances in solving problems that have resisted the best attempts of the AI community for many years" (LeCun et al. 2015).

However, despite remarkable advances in AI, humans have much better learning and generalizing abilities. Based on our literature review in cognitive science, one of the integral ingredients in human learning is that people can intelligently integrate knowledge learned previously to make inference and solve new problems efficiently and effectively (Pan and Yang 2010; Xu and Tenenbaum 2007). For machine learning approaches, however, we often need to collect a large amount of data to train a new model from scratch for every new task. In addition, based on the statistical learning theory, the machine learning approaches can only perform well when the data size is large enough and the testing data and training data share similar distributions (Vapnik 1999). In such cases, transfer learning, which could potentially address this distribution discrepancy, would be desirable.

Transfer learning techniques have been widely applied to solving a large number of real-world challenges, especially in the machine learning field. When data is not sufficient enough to build a valid model for a given target task, transfer learning can be utilized to select and transfer useful knowledge learned from a related source domain to the target domain (Pan and Yang 2010). Many machine learning models adopted transfer learning to improve performance (Li et al. 2012). However, there are very few attempts of effectively using this technique in healthcare applications. In healthcare domain in particular, which is extremely knowledge-intensive, the application of transfer learning could be of more significance. However, transfer learning is not always beneficial, especially in highly divergent domains, such as healthcare domain. Transfer learning can sometimes reduce the learning performance of the target task (known as "negative transfer") (Pan and Yang 2010). In addition, existing studies on transfer learning are often limited to one specific approach and fail to consider synergistic effects by combining multiple transfer learning approaches. Moreover, the current applications of transfer learning are *ad-hoc*, without providing evidence or theoretical support on if transfer learning actually helps, when it should be applied, and how to choose from different transfer learning approaches.

Fortunately, the success of deep learning provides a unique opportunity for transfer learning as useful knowledge representations can be learned automatically, making knowledge transfer more efficient (LeCun et al. 2015). Traditional transfer learning often requires manual engineering steps to extract transferable knowledge, which requires considerable domain expertise and is extremely time-consuming (Abbasi et al. 2012). Deep learning makes it possible to extract useful and transferable features from much larger datasets, which is not possible when using a manual approach. This is further supported by recent studies showing that deep learning can learn transferable features which generalize well to new tasks (Long et al. 2015).

In this study, we try to address the above research gap. Drawing on the theory of transfer of learning, we propose a deep learning-based model and incorporating different transfer learning approaches. As a case study of the proposed model, we demonstrate the performance using a healthcare related text analytics dataset. Our results indicate that transfer learning can significantly improve the learning outcome of tasks in a knowledge-intensive domain. In addition, we find that each type of transfer learning approach can improve learning outcome of the model, while combining them more intelligently and strategically can further contribute to the performance. Our results suggest that different types of transfer learning are not substitutive for one another, but instead, can produce synergistic effects on improving the overall performance of the learning. Moreover, deep learning models are more robust to negative transfer learning mostly likely due to its ability to learn and select necessary representations and suppress irrelevant features.

The main contributions of our work are summarized as follows. First, we propose a transfer learning framework and successfully apply it in mining healthcare data to achieve better performance. The preliminary results show that the proposed transfer learning framework can improve existing state-of-the-art methods. Second, drawing on the theory of transfer of learning, we theoretically evaluate different transfer learning approaches and how they interact. This is in contrast to the existing studies, which often utilize only one transfer learning approach in an *ad-hoc* manner. However, when people try to solve new problems, they normally reply on a series of sources of prior knowledge. Thus, our study can provide important theoretical implications on transfer learning effects, model parameter transfer, and deep learning models. It also provides practical insights on building better machine learning models to solve target tasks in machine learning and healthcare AI communities.

The rest of the paper is organized as follows. We first review the existing studies on transfer learning in machine learning and healthcare AI and identify research gaps. Then, an overview of theory of transfer of learning and proposed research framework are described. We then introduce our experimental methodology and present the preliminary results. Lastly, we conclude the paper and discuss future research directions we are currently working on.

Literature Review

Transfer Learning

We often address three research questions in transfer learning: "what to transfer", "how to transfer", and "when to transfer" (Pan and Yang 2010). Many existing studies utilized transfer learning as a standard procedure in building machine learning models in both natural language processing tasks (Dai and Le 2015) and image processing tasks (Wu and Dietterich 2004). With recent success in deep learning, transfer learning has also been adopted in developing deep learning models. Transfer learning has the potential to be more important and integral for deep learning models as deep learning is extremely data-hungry and needs a much larger volume of data than traditional machine learning models and a long processing time to learn patterns. Bengio (2012) demonstrated a potential application of transfer learning and unsupervised learning in building deep learning algorithms. Shin et al. (2016) studied the effect of transfer learning and showed that the models learned from a generic image dataset significantly improved the performance of medical image detection. Long et al. (2015) demonstrated that deep learning can learn transferable features which generalize well to novel tasks.

As mentioned, one potential risk with transfer learning is that the performance of the target task is not guaranteed to be improved. If the source and target domains are largely different, the performance of transfer learning might not be desirable. This is known as negative transfer. Deep learning models have the potential to be more robust to negative transfer learning due to its ability to learn and select necessary knowledge representation and suppress irrelevant features (LeCun et al. 2015).

Theory of Transfer of Learning

Transfer learning has also been applied in information system research. From the theoretical perspective, the mechanism of transfer learning can be best explained by the theory of transfer of learning from two perspectives (Kang et al. 2017). The first perspective is from the nature of the tasks, also termed as the environmental theory of transfer of learning. It mainly focuses on the characteristics of both source and target tasks. Transfer learning can only be effective if the source task and target task share common elements and characteristics. The more similarities exist between the source task and target task, the better the transfer learning performance is. The second perspective asks that, given the fact that similarities exists between the source and target tasks, can the similarities be identified? This is defined as the cognitive theory of transfer of learning, which "argues that the likelihood of learning transfer is determined by the ability of learners in retrieving relevant prior experience stored in memory" (Kang et al. 2017). Combining these two perspectives together, we can draw the conclusion that it is not enough to just ensure similarities between the source and target tasks, we also need to ensure that the similarities can be recognized and transferred.

Based on the theory of transfer of learning, Kang et al. (2017) examines the effects of prior knowledge in information system development. They suggest that when the prior and current projects share the same knowledge elements, project teams' prior experiences can be significantly translated into performance gains for the current projects. They also explore how different types of experiences would interact with each other and how the project complexity would impact the transfer learning performance. Multi-task learning is very close to transfer learning, as it tries to learn multiple tasks simultaneously. The main difference between them is that transfer learning focuses on the performance of the target task, while multi-task learning focus on both the source and target tasks. Lin et al. (2017) proposes a Bayesian multitask learning model to coordinate a set of baseline models, one for each event, and communicate training information across the models to uncover the latent features that can benefit for each individual task. Compared with the performances obtained from the single-task model, their multi-task approach achieves an improved predictive performance.

Research Gap and Research Questions

Existing studies often utilize transfer learning to improve learning performance. However, they often fail to justify if the transfer learning actually helps to improve the performance and/or why a certain transfer learning approach is utilized. As a result, transfer learning has been commonly used in an *ad-hoc* manner, especially in healthcare domain (Li et al. 2015). In addition, extant studies often utilize one type of transfer learning and have not considered the combination of multiple transfer learning techniques (Long et al. 2015; Wu and Dietterich 2004). However, from human learning perspective, when people try to apply knowledge learned previously to solve new problems, they normally do not reply on just one source, but a series of sources of prior knowledge. Moreover, with the recent progress in deep learning, which could learn and select useful knowledge representation and suppress irrelevant features, how it can be combined with transfer learning has not been well addressed. Lastly, unsupervised learning has been recognized as being able to significantly contribute to AI study, but it has been largely overlooked because of the exceptional successes of supervised learning (Bengio 2012). However, it is believed that unsupervised learning is very close to human learning, which is largely unsupervised, and we learn things without knowing every aspect or name of every object. Although several existing approaches have adopted unsupervised transfer learning to pre-train the deep learning models, very few studies have considered combining it with other transfer learning techniques. In summary, we are trying to answer three research questions:

RQ1: Is there a systematical and theoretically grounded way to combine different transfer learning approaches for the purpose of improving performance in healthcare text analytics tasks?

RQ2: What are the relationships among different transfer learning approaches? Are they substitutive or complementary for one another with respect to the model performance?

RQ3: How to apply unsupervised transfer learning approach together with other transfer learning approaches in the deep learning-based framework?

Research Rationale and Proposed Framework

In this section, we explain our overall deep learning-based transfer learning framework to answer those three research questions. Based on our literature review and previous discussions, transfer learning has not been effectively used in healthcare applications (Li et al. 2015). Given the fact that healthcare is a knowledge-intensive domain, transfer learning could be of great significance. Moreover, there are extremely large amounts of data and knowledge sources, both of which provide important resources for transfer learning study. Thus, we select a healthcare text analytics task to evaluate our proposed framework.

Based on our discussion, deep learning provides a natural platform for future study on transfer learning, which can be used to solve "how to transfer" issue. In addition, regarding the "when to transfer" question, deep learning models are believed to be more robust to negative transfer learning due to its ability to learn and select necessary knowledge representation and suppress irrelevant features. Thus, transfer learning is more likely to only benefit the learning process. Thus, deep learning-based model provides a promising solution to these two questions, and the key question is narrowed down to "what to transfer". Thus, deep learning model has been used in our framework. We use the recurrent neural network, which is especially useful for modeling the sequence type of data such as language. More specifically, we use Long Short Term Memory (LSTM), the state-of-the-art deep learning model for natural language processing (Graves and Schmidhuber 2005). As we are focusing on text analytics tasks, using LSTM can maximize the generalizability of our model.

Drawing on the theory of transfer of learning, we need to ensure that similarities between the source and target domains exist and can be identified and utilized. Deep learning models have the ability to identify and extract useful similarities and suppress dissimilarities. Thus, it is adopted in our framework design. The existing transfer learning approaches can be grouped into four categories: instance-based, feature-based, parameter-based, and relational knowledge-based (Pan and Yang 2010). For a healthcare text analytics task, the learning processes can be split into three components. First of all, as all the text analytics tasks are based on linguistic knowledge about a language, which serves as the basic analysis units for text analytics tasks. Thus, transferring general knowledge from a language perspective provides a basis to solve text analytics tasks. In addition, two tasks in different domains could share some similarities, such as sentence structures, internal relations, as well as dissimilarities, such as distribution discrepancy, domain-specific knowledge. It is crucial to identify the similarities between the source task and the target task. It is also recognized that the discrepancy of different domains should be reduced to improve transfer learning performance (Long et al. 2015).

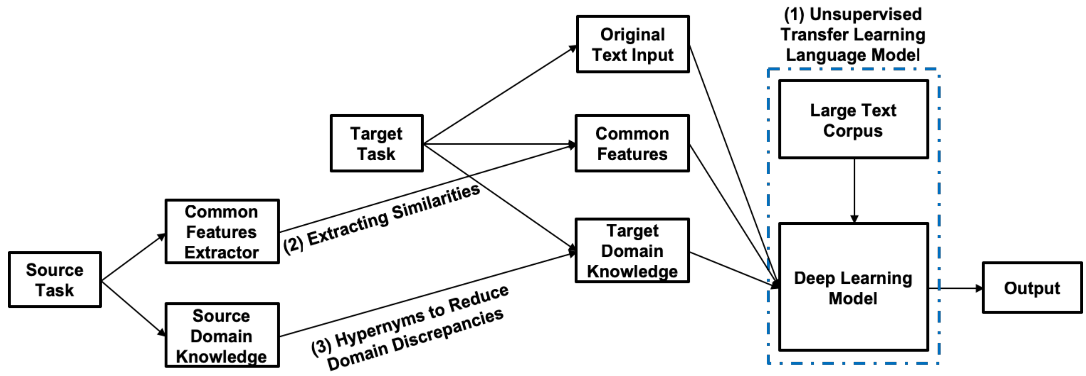

Figure 1 A Deep Learning-Based Transfer Learning Framework for Health Text Analytics

Based on these design requirements, we propose a deep learning-based transfer learning framework to perform healthcare text analytics tasks (see Figure 1). The framework combines (1) an unsupervised language model (parameter-based transfer learning) (defined as TL1); (2) a deep learning model to extract similarities between the target and source tasks (feature-based transfer learning) (defined as TL2); and (3) a relationship knowledge to reduce the domain discrepancy (relational knowledge-based transfer learning) between the target and source domains (defined as TL3). Since they are addressing different aspects of the tasks, we expect they are complementary for one another rather than substitutive. Thus, we expect this combination could achieve the best performance. The detailed implementation will be demonstrated in the next section.

Experimental Design and Preliminary Results

In this section, we demonstrate the implementation and evaluations of the proposed transfer learning framework using a healthcare text analytics task. For the language model, we adopted the approach developed by Dai and Le (2015). A LSTM-based language model is trained using a large text corpus (including 1.25 million medical articles from PubMed, text data from health websites: WebMD.com, DailyStrength.com, HealthGrades.com, American Diabetes Association forum) with 9 billion tokens. After training, the model is able to predict what comes next in a language sequence. As we are transferring the language parameters by language model, this belongs to parameter-based transfer learning.

In addition, for the common features, a widely used approach is that we train a deep learning model using similarly labeled datasets. We then apply the pretrained model on the target datasets to extract common features. As discussed previously, deep learning models can be used to automatically extract features from the raw data. Thus, we trained a LSTM model for named entity recognition task on news articles (Sang and De Meulder 2003). This task was selected because it is a similar task as our target task, adverse drug event (ADE) detection. After training the LSTM model, we used it on our target dataset to extract common features shared by two different tasks.

Domain knowledge refers to the domain-specific knowledge or context in which the information system or application operates. Consider two sentences: (1) "Tires may cause a blowout"; (2) "Metformin may cause a headache". The first sentence reveals the possible defect about the tires, while the second sentence indicates the possible defect (side effect) of the drug Metformin. Although they are referring to the similar task, these two sentences come from different domains and each has its own domain-specific knowledge. One possible way to reduce these domain discrepancies is to find the more generic hypernyms that group related elements in different domains into the same group. By apply WorkNet hypernyms (Miller 1995), we find the hypernyms for "tire" are: "hoop", "band", "strip", "artifact", "'whole", "'object", "***physical_entity***", and the hypernyms for "metformin" are: "antidiabetic", "medicine", "drug", "agent", "causal_agent", "substance", "***physical_entity***". Similarly, the hypernyms for "blowout" are: "malfunction", "failure", "happening", "event", "***psychological_feature***", and hypernyms for "headache" are: "ache", "pain", "symptom", "evidence", "information", "cognition", "***psychological_feature***". As hypernyms can capture the common generic categories for terms from different domains, they could be used to reduce the domain discrepancies.

An overview of the transfer learning approaches adopted in our framework is given in Table 1.

Table 1. Overview of Transfer Learning Approaches Adopted

Transfer Learning Approach	Source Domain	Target Domain	What Being Transferred	Type of Transfer Learning
TL 1: Unsupervised transfer learning: language model	General language	Healthcare text (language)	Language parameters	Parameter-based
TL2: Deep learning model-based common features extractor	General named entity recognition task	ADE extraction task	Common features between source task and target task	Feature-based
TL3: Hypernyms to capture domain-invariant relations	General relational knowledge	Medical domain relational knowledge	Domain-invariant relations	Relational knowledge-based

We then constructed an ADE dataset from drug review data from WebMD.com, one of the largest health information websites in the United States. We developed an automated crawler to download 5,711 diabetes drug reviews and split the reviews into a training set, a validation set, and a testing set. Each word in a sentence was labeled to one of three entity groups: D, S, or O (which stand for Drug Entity, ADE entity, and Others, respectively). Four graduate students were separated into two groups with each group annotated each word separately. A third rater will make a final decision whenever the two groups disagree. A summary of the annotated distribution is provided in Table 1.

Table 2. Label Distribution

Datasets	**Entity Classes**	**Total Count**	**Percentage**
Training set	Others	212,638	91.44%
	Drug	7,472	3.21%
	ADEs	12,432	5.35%
Validation set	Others	43,522	92.31%
	Drug	1,396	2.96%
	ADEs	2,231	4.73%
Testing set	Others	37,934	90.69%
	Drug	1,221	2.92%
	ADEs	2,674	6.39%

For evaluation and comparison purpose, we built a LSTM model without utilizing any transfer learning techniques on the ADE dataset as the baseline performance (Table 3 row 2). Consistent with prior work, we used precision, recall, and f-measure as our performance measures. In the first experiment, we evaluated the effectiveness of unsupervised learning. We applied the pre-trained LSTM language model (***Unsupervised transfer learning: language model***) on the ADE dataset, and we compared the effectiveness of the unsupervised language model against the model without unsupervised transfer learning step. The performance comparisons are included in Table 3, row 3. In experiment 2, we evaluated the effectiveness of reducing domain discrepancy. The hypernyms are intended to reduce the discrepancies between different domain-specific knowledge (***dissimilarities***). We evaluated the performance after incorporating these hypernyms features to the baseline model. The performance comparisons are included in Table 3, row 4. In experiment 3, we used the LSTM model pretrained on named entity recognition task to extract 100-demensional features (***similarities***) from ADE datasets and appended them into the baseline model. The performance comparisons are included in Table 3, row 5. Lastly, since we are interested in the relationships among different transfer learning settings, we also combined all the three components to deliver the proposed framework depicted in Figure 2. The performance is shown in Table 3, row 6.

Based on the results, each individual transfer learning approach significantly improves the model performance (all p-values are <0.0001), indicating the transfer learning can indeed contribute to the performance gains. One interesting pattern is that the hypernyms contributes more to improving the performance than the unsupervised language model and extracted common features. We believe this can explained by the knowledge-intensive nature of the task. A more comprehensive evaluation could be conducted by considering a wide spectrum of tasks in the future. In addition, combing all three transfer learning approaches achieves the best performance. Contrary to existing studies, which suggests that knowledge in one type may interfere with or substitute for the accumulation of knowledge in another type, our preliminary results suggest that different transfer learning approaches very likely complement one another, thus they produce additional synergistic effects on performance. However, additional experiments need to be conducted to further validate this statement.

Table 3. Evaluation Results

Brief Model Representation	Detailed Model Representation	Precious	Recall	F-measure	P-value for F-measure (Compared with baseline)
Baseline	Baseline LSTM model	0.601	0.645	0.622	
Baseline + TL1	LSTM with unsupervised language model	0.676	0.711	0.693	< 0.0001***
Baseline + TL2	LSTM with extracted common features	0.702	0.729	0.715	< 0.0001***
Baseline + TL3	LSTM with hypernyms	0.726	0.739	0.732	< 0.0001***
Baseline + TL1 + TL2 + TL3	LSTM with unsupervised language model, hypernyms, and common features	**0.785**	**0.798**	**0.791**	< 0.0001***

Conclusions and Future Steps

In this study, we introduce a deep learning-based transfer learning framework for solving a healthcare text analytics task. Drawing on the theory of transfer of learning, our study proposes a theoretically grounded way of conceptualizing the different transfer learning approaches and how they contribute to the learning process. Our results show that these transfer learning approaches can individually improve the performance of machine learning model. Moreover, this study provides a better appreciation of how different types of transfer learning interact with one another and their combination can further contribute to the overall performance. Finding these interaction effects among different types of transfer learning contribute significantly to transfer learning and general machine learning literatures because it enables us to systematically and theoretically understand how to apply transfer learning in building machine learning models for various different tasks.

This is a research-in-progress study. Although promising results were obtained, we are actively working on improving our proposed framework and expanding to more general scenarios. First, the current deep learning-based platform provides an excellent starting point for studying transfer learning effects in great detail. We are trying to further improve our designed artifact grounded by learning theories. In addition, we will extend our datasets to more domains and task types. We start with a task in healthcare domain as it is one of the most knowledge-intensive domains. Thus, the concept of transfer learning could be of more significance. However, the research framework we developed is generic and has the potential to be widely applied to a wider spectrum of domains. Previous studies suggested that the relative impact and importance of transfer learning dependent on the actual task complexity (Kang et al. 2017). Thus, we plan to conduct more comprehensive and robust evaluations on our proposed framework and maximize the practical impacts. Lastly, the current theories suggest that transfer between tasks takes place only if the tasks share identical elements. In this study, we start by following the traditional transfer learning route that the source and target tasks are similar tasks but from the different domains. However, with the power of deep learning, we believe we can also apply transfer learning when source and target tasks are different tasks but from the same domain. We believe it would be very promising to study the possibility of performing transfer learning at both task and domain levels.

References

Abbasi, A., Albrecht, C., Vance, A., and Hansen, J. 2012. "Metafraud: A Meta-Learning Framework for Detecting Financial Fraud," *MIS Quarterly* (36:4), pp. 1293-1327.

Bengio, Y. 2012. "Deep Learning of Representations for Unsupervised and Transfer Learning," in *Proceedings of ICML Workshop on Unsupervised and Transfer Learning*, Guyon, I., Dror, G., Lemaire, V., Taylor, G., and Silver, D. (eds.), Bellevue, WA, pp. 17-36.

Cheng, Y., Wang, F., Zhang, P., and Hu, J. 2016. "Risk Prediction with Electronic Health Records: A Deep Learning Approach," in *Proceedings of the 2016 SIAM International Conference on Data Mining*, Venkatasubramanian,S. C. and Meira, W. (eds.), Miami, FL, pp. 432-440.

Dai, A. M., and Le, Q. V. 2015. "Semi-Supervised Sequence Learning," in *Proceeding of the 29th Conference on Advances in Neural Information Processing Systems,* Cortes, C., Lawrence, N. D., Lee, D. D., Sugiyama, M., and Garnett, R. (eds.), Montréal Canada, pp. 3079-3087.

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., and Thrun, S. 2017. "Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks," *Nature* (542:7639), pp. 115-118.

Graves, A., and Schmidhuber, J. 2005. "Framewise Phoneme Classification with Bidirectional Lstm and Other Neural Network Architectures," *Neural Networks* (18:5-6), pp. 602-610.

Kang, K., Hahn, J., and De, P. 2017. "Learning Effects of Domain, Technology, and Customer Knowledge in Information Systems Development: An Empirical Study," *Information Systems Research* (28:4), pp. 797-811.

LeCun, Y., Bengio, Y., and Hinton, G. 2015. "Deep Learning," *Nature* (521:7553), pp. 436-444.

Li, W., Zhao, R., and Wang, X. 2012. "Human Reidentification with Transferred Metric Learning," in *Proceedings of 11th Asian Conference on Computer Vision*, Lee, K.M., Matsushita, Y., Rehg, J.M. and Hu, Z. (eds.), pp. 31-44.

Li, Y., Vinzamuri, B., and Reddy, C. K. 2015. "Constrained Elastic Net Based Knowledge Transfer for Healthcare Information Exchange," Data Mining and Knowledge Discovery (29:4), pp. 1094-1112.

Lin, Y.-K., Chen, H., Brown, R. A., Li, S.-H., and Yang, H.-J. 2017. "Healthcare Predictive Analytics for Risk Profiling in Chronic Care: A Bayesian Multitask Learning Approach," *MIS Quarterly* (41:2), pp. 473-495.

Litjens, G., Sánchez, C. I., Timofeeva, N., Hermsen, M., Nagtegaal, I., Kovacs, I., C., Bult, P., Van Ginneken, B., and Van Der Laak, J. 2016. "Deep Learning as a Tool for Increased Accuracy and Efficiency of Histopathological Diagnosis," *Scientific Reports* (6), p. 26286.

Long, M., Cao, Y., Wang, J., and Jordan, M. I. 2015. "Learning Transferable Features with Deep Adaptation Networks," *arXiv preprint arXiv:1502.02791*.

Miller, G. A. 1995. "Wordnet: A Lexical Database for English," *Communications of the ACM* (38:11), pp. 39-41.

Pan, S. J., and Yang, Q. 2010. "A Survey on Transfer Learning," *IEEE Transactions on Knowledge and Data Engineering* (22:10), pp. 1345-1359.

Sang, E. F., and De Meulder, F. 2003. "Introduction to the Conll-2003 Shared Task: Language-Independent Named Entity Recognition," *arXiv preprint cs/0306050*.

Shin, H.-C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., and Summers, R. M. 2016. "Deep Convolutional Neural Networks for Computer-Aided Detection: Cnn Architectures, Dataset Characteristics and Transfer Learning," *IEEE Transactions on Medical Imaging* (35:5), pp. 1285-1298.

Vapnik, V. N. 1999. "An Overview of Statistical Learning Theory," *IEEE Transactions on Neural Networks* (10:5), pp. 988-999.

Wang, D., Khosla, A., Gargeya, R., Irshad, H., and Beck, A. H. 2016. "Deep Learning for Identifying Metastatic Breast Cancer," *arXiv preprint arXiv:1606.05718*.

Wu, P., and Dietterich, T. G. 2004. "Improving Svm Accuracy by Training on Auxiliary Data Sources," in *Proceedings of 21th International Conference on Machine Learning*, Koller, D., Schuurmans, D., Bengio, Y. and Bottou, L. (eds.), New York, NY, pp. 110.

Xu, F., and Tenenbaum, J. B. 2007. "Word Learning as Bayesian Inference," *Psychological Review* (114:2), pp. 245-272.