Transfer Learning		Settings	Domain/ Task	Feature spaces/ Label spaces	Source Data	Target Data	Related Areas	Examples	Algorithms	References
Transfer Learning	Label-setting-based	Inductive Transfer Learning	Same or Different / Different		Unlabeled	Labeled	Self-taught Learning	Self-taught Learning	▪ Use unlabeled data to construct a new representation with *sparse coding algorithm*. ▪ Apply this new representation to labeled data. ▪ Use existing classification methods in this new space.	Self-taught Learning: http://ai.stanford.edu/~hllee/icml07-selftaughtlearning.pdf
					Labeled	Labeled	Multi-task Learning	Question Answering	▪ Build a Multitask Q/A Network trained jointly on decaNLP. - Takes in a question and context, encodes both with a BiLSTM, uses dual coattention to condition representations for both sequences, compresses all info with another two BiLSTMs, and applies self-attention - Uses a final two BiLSTMs to get representations of the question & context. - Creates Multi-pointer-generator decoder.	The Natural Language Decathlon Multitask Learning as Question Answering: https://arxiv.org/pdf/1806.08730.pdf
						Labeled + Unlabeled	General ITL	Kidney Detection	▪ Initialize the weight from Source Network parameters on the target network. ▪ Freeze the weights of the low-level layers of the network, while updating the weights of other layers until it saturates.	Understanding the Mechanisms of Deep Transfer Learning for Medical Images: https://arxiv.org/pdf/1704.06040.pdf
								Teacher-student Network	▪ Distillation between teacher and student network. - Train the teacher model to produce class probabilities by using output layer that convert the logit. - Establish the correspondence between the intermediate outputs of the student and teacher network by matching the logit. - Forward pass through the teacher network - Backpropagate through the student network.	Distilling the Knowledge in a Neural Network: https://arxiv.org/pdf/1503.02531v1.pdf
		Transductive Transfer Learning	Different / Different		Labeled	Unlabeled /Limited	Domain Adaptation	Supervised Domain Adaptation	▪ Create pairs of source and target instances to handle target data. ▪ Extend adversarial learning to align the semantic information of classes. - Create a discriminator to distinguish between samples of the source and target distributions. - Create an inference function to map a target sample to a feature space.	Few-shot Adversarial Domain Adaptation: https://arxiv.org/pdf/1711.02536.pdf
						Limited	Covariate Shift/ Sample Selection Bias	Binary Classification	▪ Build a minimax estimator, Robust Bias-aware classifier, measured by the conditional logloss (for *Sample Selection Bias*) ▪ Datasets: Generate biased subsets of used datasets as source and unbiased subsets as target samples.	Robust Classification Under Sample Selection Bias: https://papers.nips.cc/paper/5458-robust-classification-under-sample-selection-bias.pdf
		Unsupervised Transfer Learning	Same or Different / Different		Unlabeled	Unlabeled	Clustering/ Dimensionality reduction	Self-taught Clustering	▪ A clustering problem of self-taught learning ▪ Learn the feature representation of data in the source domain as auxiliary data using *Sparse Coding technique*. ▪ Build two objective functions by extending the *information theoretic co-clustering* and share feature clustering.	Self-taught Clustering: https://www.cse.ust.hk/~qyang/Docs/2008/dwyakicml.pdf
	Space-setting-based	Homogeneous Transfer Learning		Same/ Same	Labeled	Limited/ Unlabeled		Fatigue Detection	▪ CP-MDA (Limited target data): Build classifiers for all source domains to get weight value, create a learning task and build a target learner using labeled & pseudo labeled target data. ▪ 2SW-MDA (Unlabeled target data): Compute weights for source domains, modify the weights, and learn a target classifier.	Multi-source Domain Adaptation and Its Application to Detection of Fatigue: https://www.cs.ucdavis.edu/~davidson/Publications/TKDD.pdf
				Same/ Different	Labeled	Labeled		Feature Transferring	▪ Initialize the weights from a pre-trained model ▪ Use the target data to fine-tune the parameters for the target task. - Fine-tune all the layers of DNN - Freeze several layers of the DNN and only fine-tune the layers that can reduce overfitting.	How transferable are features in deep neural networks?: https://papers.nips.cc/paper/5347-how-transferable-are-features-in-deep-neural-networks.pdf
		Heterogeneous Transfer Learning		Different/ Same	Labeled	Limited		Translated Learning	▪ Build a translator to connect two different feature spaces using co-occurrence data. ▪ Use the language model and combine feature translation with nearest neighbor learning. ▪ Build a model using a Markov chain c → y → x.	Translated learning: Transfer learning across different feature spaces: https://papers.nips.cc/paper/3492-translated-learning-transfer-learning-across-different-feature-spaces.pdf
				Different/ Different	Labeled	Limited		Drug Efficacy Prediction	▪ Find a common latent input spaces using spectral mapping. ▪ Apply a clustering-based sample selection method to select new related instances. ▪ Use Bayesian-based method to find the relationship and resolve the differences in the output space.	Transfer Learning on Heterogeneous Feature Spaces via Spectral Transformation: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.298.7133&rep=rep1&type=pdf