

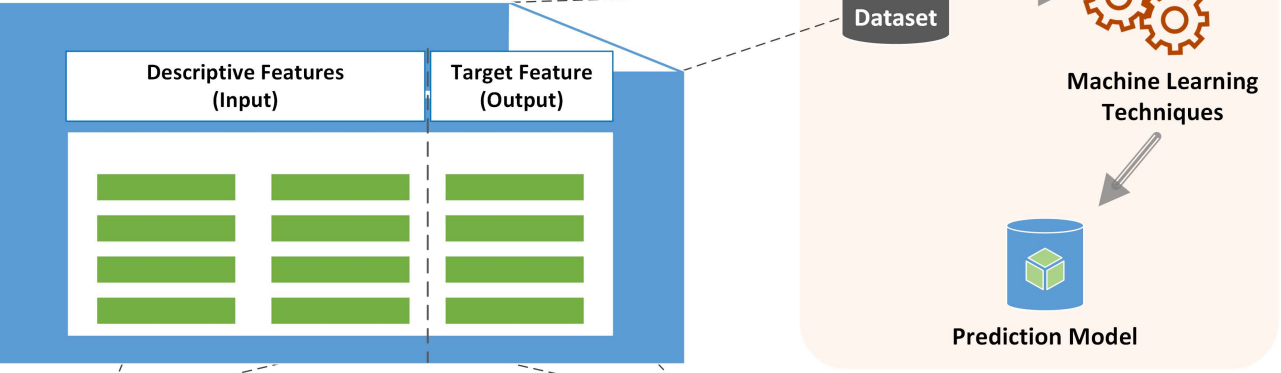
Supplemental Digital Appendix 2

**Conceptual Diagram to Illustrate How Machine Learning Techniques Can Be
Applied to Current Competence Assessment Methods, From a Systematic Review
of Literature on Using Machine Learning Techniques to Assess Physician
Competence, May 2017**

Conceptual Diagram to Illustrate Machine Learning

Machine learning (ML): an automated process that extracts patterns from data.

Supervised Learning: automatically learn a model of the relationship between a set of input and output data based on historical instances.



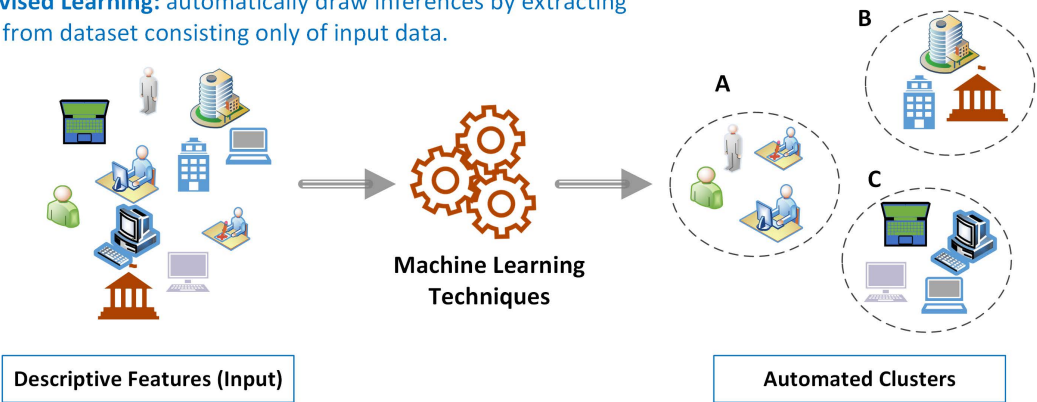
Examples of Input Features		Examples of Output Features	
Surgery	Path length, completion time, force, acceleration, number of movements	Regression: OSATS, FLS scores	Classification: novices vs seniors
Radiology	Free-text radiology reports, patient clinical data, image features	Classification: malignant vs benign tumor, BI-RADS categories, novices vs seniors, correct vs incorrect diagnosis	
Medicine	Free-text clinical notes, survey responses, audio and video features	Regression: communication quality score, performance rating, peer-assessment score	Classification: expertise level, adherence to guidelines, problem-solving correctness

Feature Extraction – most ML datasets involve a large amount of complex variables, demanding the use of dimensionality reduction techniques (e.g. LDA and PCA) to select relevant variables and/or built combinations. Common data sources used in the reviewed studies were: free-text from medical records, audio from clinician's conversations, video and motion sensor data from surgeries.

“Ground Truth” - refers to the *gold standard* measures against which the input features will be related. If it is a numerical variable, the relationship model is a regression. If categorical variable, it is a classification model. Most “ground truth” features were human-based evaluations (e.g. observation rating scales, self-assessment, expert confirmation). Other common feature was expertise level (novice vs senior).

OSATS: objective structured assessment of technical skills; FLS: fundamentals of laparoscopic surgery; LDA: linear discriminant analysis; PCA: principal component analysis; BI-RADS: breast imaging reporting and data system

Unsupervised Learning: automatically draw inferences by extracting patterns from dataset consisting only of input data.



The same input data used for supervised machine learning can be used in unsupervised models. In the reviewed studies, most unsupervised techniques analyzed statistical similarity among input features (e.g. motion sensor data, video frames, free-text) to infer clusters of physicians presenting similar competency.