Cognition Inspired Diagnostic Image Quality Models

Jorge E. Caviedes, Ph.D.

Adjunct Research Professor & Lecturer

ASU-Mayo Center for Innovative Imaging

School of Computing and Augmented Intelligence

School of Biological and Health Systems Engineering

Jorge.caviedes@asu.edu

Ira A. Fulton Schools of Engineering

Arizona State University

School of Computing and Augmented Intelligence

Summary

- Problem statement
- Technical vs. diagnostic image quality
- Components of diagnostic image quality
- Subjective assessment
- Objective model
- Image feature metrics
- Discussion

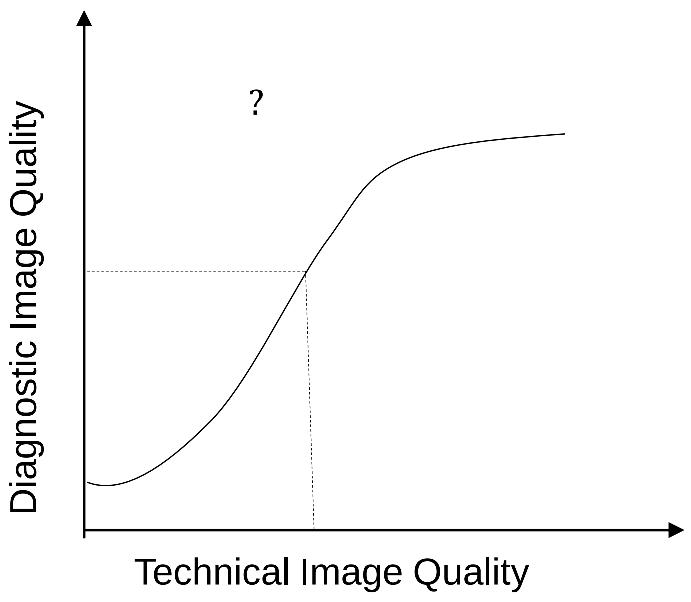
Problem Statement

- Images from full field digital mammography (FFDM) as well as its successor digital breast tomosynthesis (DBT) are the primary diagnostic tool in the detection, diagnosis, and treatment of breast cancer
- Both modalities are affected by technical recalls (TR) due to artifacts such as motion blur during image acquisition, and diagnostic errors such as false negatives, and misdiagnoses in many cases due to human, cognitive and perceptual factors related to diagnostic image quality
- Presently, there are not broadly used subjective or objective models of image quality specifically aimed at the requirements of image-based diagnosis

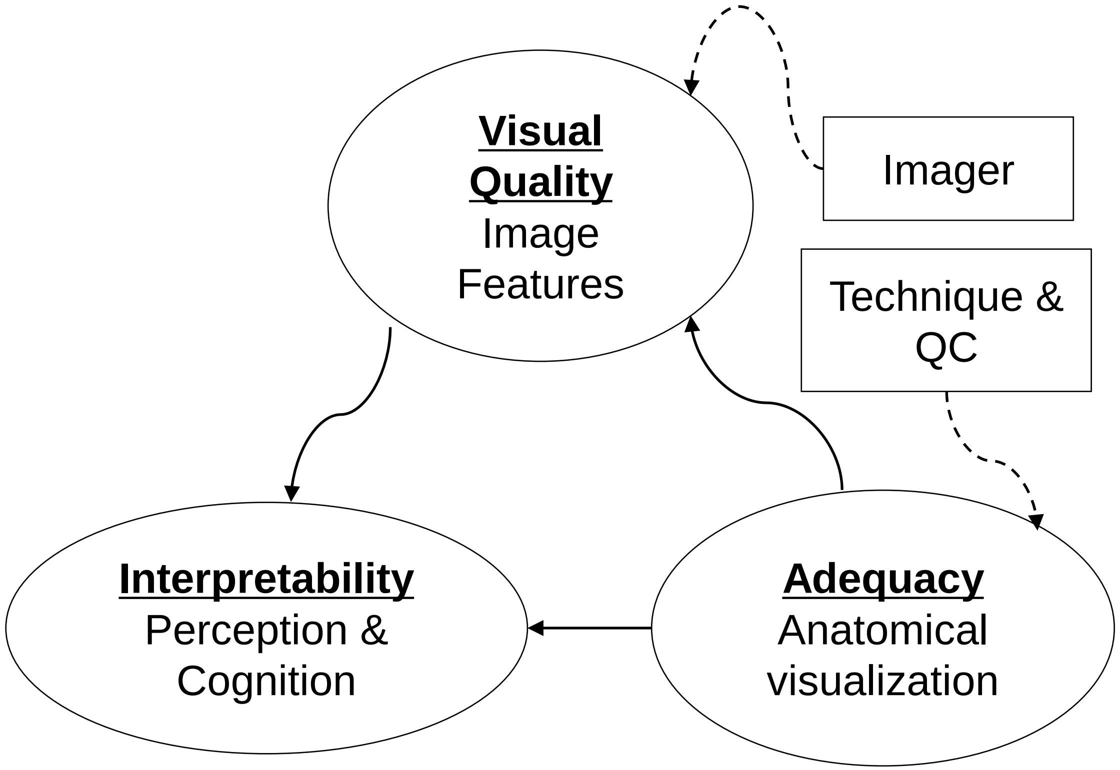
The Gap Between Technical and Diagnostic Image Quality (TIQ, DIQ)

There is a qualitative assumption of image quality monotonicity: the better the technical quality, the better the diagnostic quality.

Little is understood about the actual diagnostic quality variables and how they are affected by technical quality.



Diagnostic Image Quality



- Interpretability, directly related to the concept of lesion conspicuity, is subject to perceptual constraints including contrast sensitivity, human visual system adaptation, and workload
- Visual quality, i.e., the appearance of tissues and structures in terms of resolution, sharpness, contrast, brightness, etc., is subject to imaging constraints including sensor physics, radiation dose, safety, and patient comfort
- Adequacy depends on quality control protocols for compression and techniques such as breast positioning and breath hold to reduce motion artifacts

Objective Quality Models

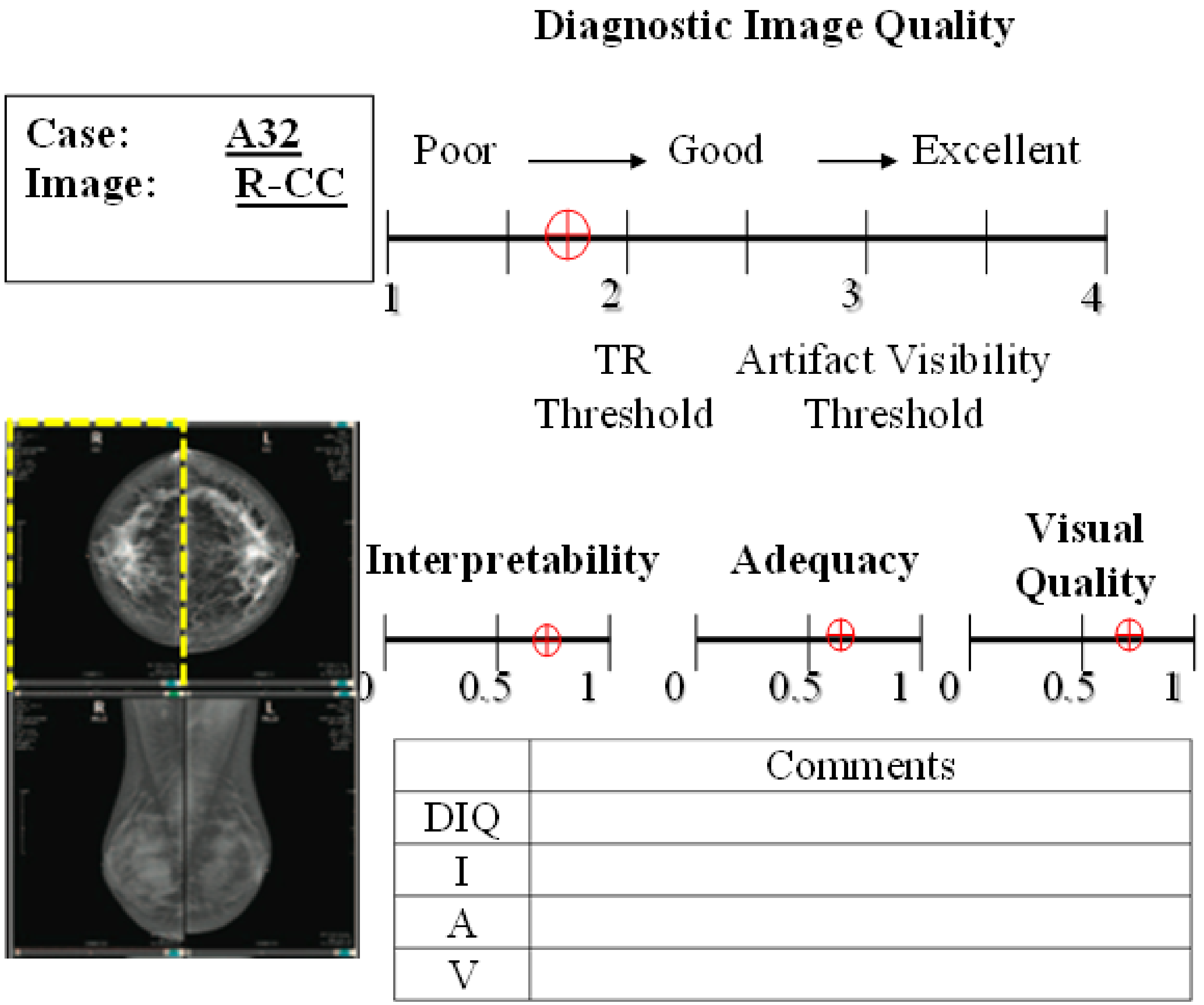
- Perceptual models have been developed to deal with the detection problem – detecting the abnormal region or structure*
- Human performance is modeled using model observers**
- The progress in objective models includes
  - Mimic human detection performance
  - Reflect improvement in signal quality resulting from new reconstruction algorithms and imaging techniques
- Generic visual quality models (aka SSIM): such models mimic mean opinion scores (MOS) from image features, but they neither address diagnostic content effect nor the impact of image quality on human diagnostic performance

**Weber's law and saliency models were not designed to detect diagnostically relevant content.*

***Model observers do not consider cognitive requirements of DIQ*

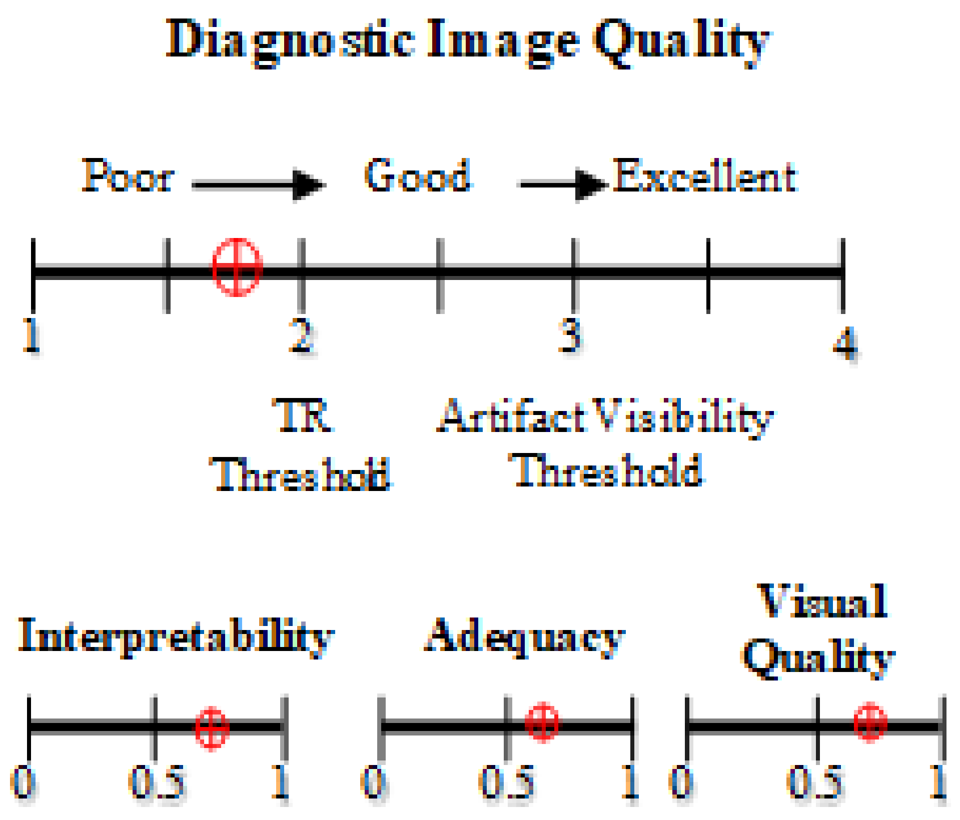
Subjective Scoring Interface Prototype

- Proposed for digital mammography, FFDM and DBT
- Inspired by generic subjective image quality evaluation methods
- Enhanced to cover 3 criteria and capture cognitive cues



A New DIQ Scale

- Our DIQ scale includes perceptual and cognitive anchor points
- Using the mouse, radiologist readers drag the markers (red encircled crosses) to score each image and can then enter comments via text or voice interface
- The web-based interface incorporates full DICOM viewer



	Comments
DIQ	
I	
A	
V	

Score Annotations Analysis: Concept Mapping

NLP can be used to structure free text by extracting concepts and their attributes

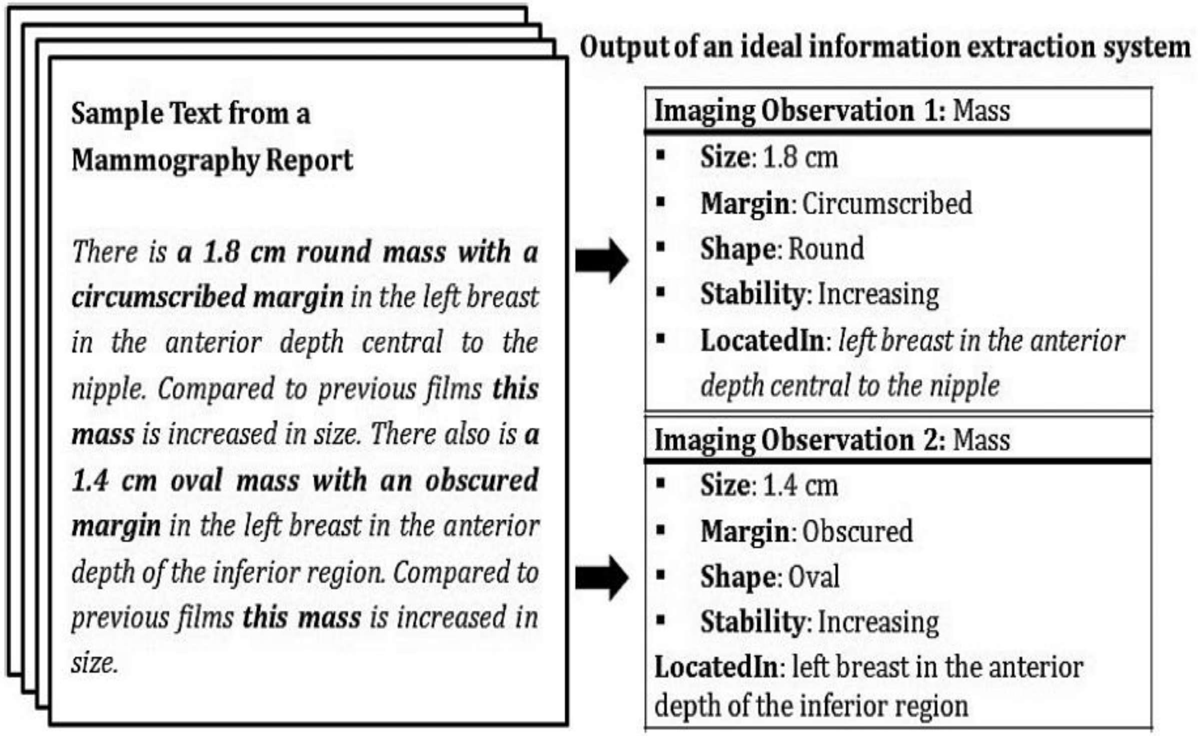


Figure 1 Example mammography report describing two different masses and the ideal output from a natural language processing system to extract the information suitable for input to a decision support system.

Selen Bozkurt, Jafi A Lipson, Utku Senol, Daniel L Rubin, Automatic abstraction of imaging observations with their characteristics from mammography reports, Journal of the American Medical Informatics Association, Volume 22, Issue e1, April 2015, Pages e81–e92, https://doi.org/10.1136/amiajnl-2014-003009

Mining for Image Features

Hypothetical case:

DIQ Descriptions

Interpretability score: 4

Description: the abnormal mass is highly visible and clearly distinguishable from the fatty tissue background without interference by obstructing tissue in front of it, its contour is well defined to allow a good segmentation and size estimation. Homogeneous and small irregular tissue structures are sufficiently visible to differentiate normal from abnormal anatomy.

Adequacy score: 5

Description: the images include full breast anatomy including well defined chest wall. The left and right images are symmetric, noise level is non-masking, no artifacts and no motion blur.

Visual quality score: 4

Description: Image brightness, contrast, and detail visibility are good in the region of interest. Smooth boundaries better visualized than irregular contours.

Image Feature Descriptions

Contrast: the difference in brightness that makes an object conspicuous over the background

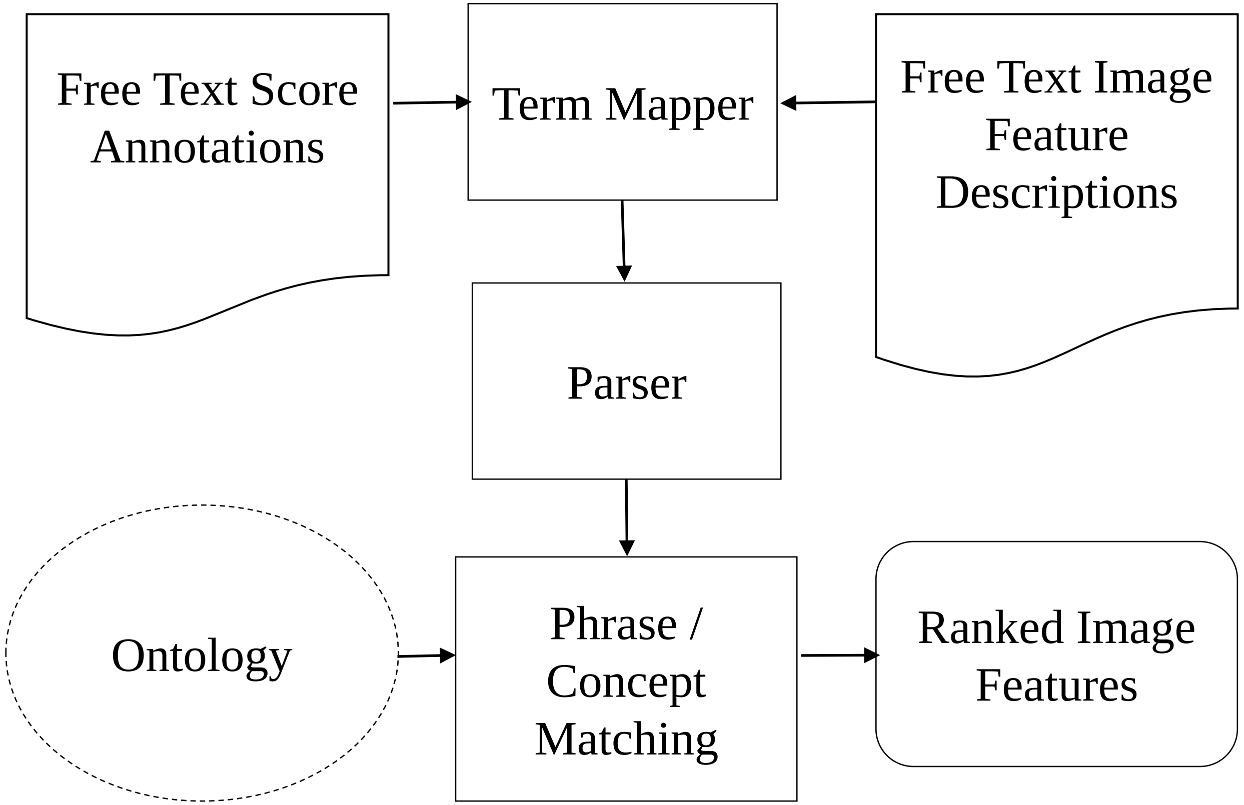
Sharpness: an image's overall clarity in terms of contour definition and contrast, perception of sharpness depends on contrast and resolution, low sharpness is called blur

Structural complexity: image statistics that describe the richness of spatial content and its visibility

Entropy of corner distribution: related to the number and degree of spatial non-uniformity of corners (or singular points) in the image.

Fractal dimension: is a ratio providing a statistical index of complexity comparing how detail in a pattern changes with the scale at which it is measured. A measure of the space-filling capacity of a pattern.

Text data mining



Using semantic similarity, we may identify relevant features

Basic approach to mining text for image features

Building an Objective DIQ Metric

- Diagnostic image quality models that have been developed to mimic unidimensional mean opinion scores (MOS) use computed global image features to build black-box algorithms
- The three-dimensional perspective we propose could potentially result in an explainable metric
- Image features that reflect content complexity have not been used in diagnostic image quality models. We expect features related to texture, fractal, and structural complexity to be relevant
- Image statistics are dependent on structural complexity and overall complexity are expected to change with tissue physiology and pathology (e.g., tumor heterogeneity)

Objective Model – Feature Engineering

Candidate features will be selected according to their discriminative power and conceptual relevance. In a preliminary study we computed the following 9 image features:

- Visual Quality - perceptual
  - Sharpness (S)
  - Contrast (C)
  - Noise (N)
- Complexity – cognitive
  - Spatial Complexity metric (SCM)
  - Local mean-luma complexity metric (LMLCM)
  - Hausdorff fractal dimension (HFD)
  - Wavelet Entropy (WE)
  - Texture (T)
  - Entropy of corner distribution (ECD)

Feature Metrics Requirements

- Linked to the score comments via text mining during model development
- Show variability and monotonicity across a varied set of images with a broad range of quality scores
- Complexity features may exhibit correlation with image content types (types of normal anatomy, types of abnormal anatomy)

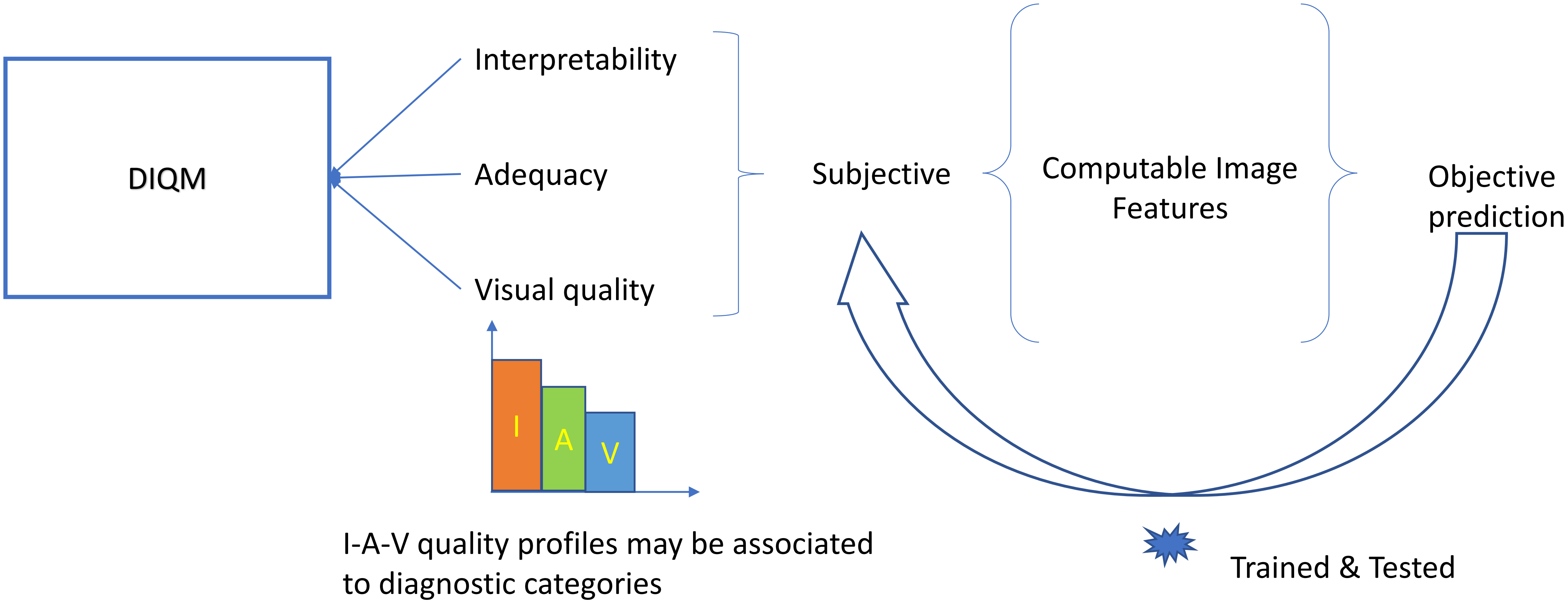
Metric Requirements – prelim. pilot study

Feature values were computed on subsets of the INbreast database and a publicly available DBT database (~200 patients, 2400 views). Anova results:

- Most of the time there was a statistically significant difference in the DBT and FFDM feature values
- For FFDM of normal and dense breast cases we found that dense breasts show no statistically significant differences among features
- Overall, DBT shows higher visual quality and higher complexity values than FFDM
- Dense breasts showed a drop in visual quality with respect to the fatty breasts

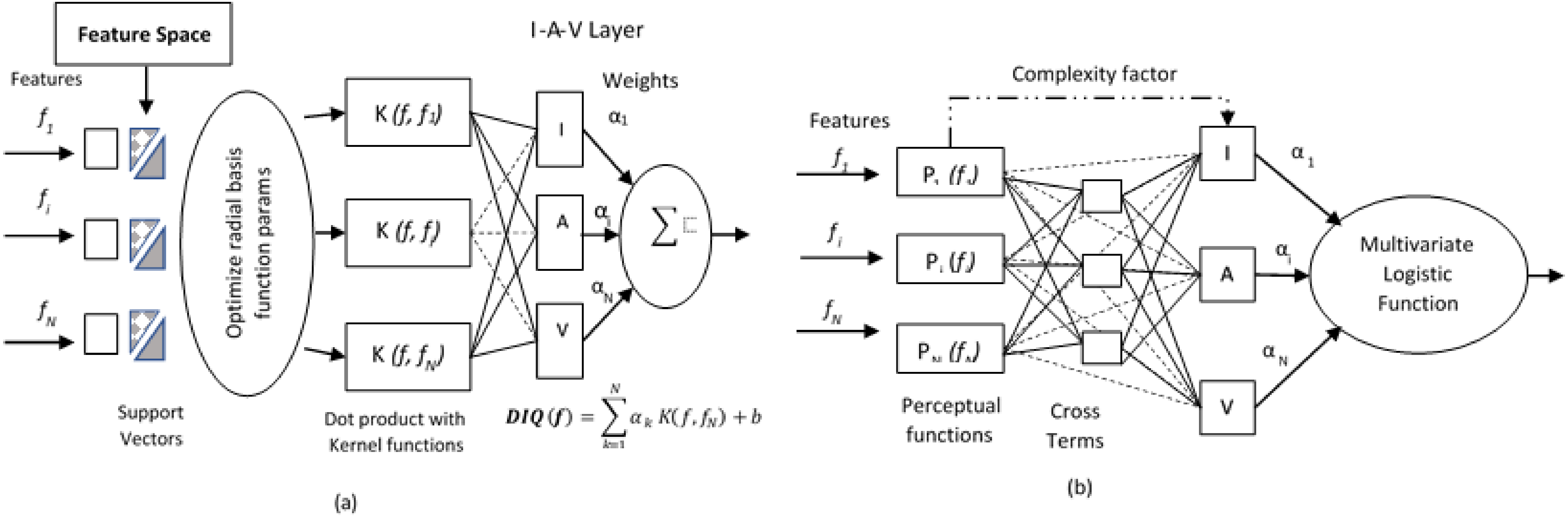
	DBT	**FFDM**	**FFDM Dense**	**FFDM Fatty**
	P-value	P-value	P-value	P-value
S	0.0042	0.0000	**0.9454**	0.0000
C	0.0006	0.0000	**0.5278**	0.0000
N	**0.1310**	0.0014	**0.6886**	0.0067
SCM	0.0000	0.0091	**0.6196**	**0.0129**
LMLCM	0.0025	0.0768	**0.1942**	**0.6537**
HFD	**0.0746**	0.0000	**0.6886**	0.0016
WE	0.0002	**0.2757**	**0.2340**	**0.8628**
T	0.0086	0.0076	**0.3619**	**0.2938**
ECD	0.0006	0.0591	**0.2272**	**0.6876**

DIQ Model

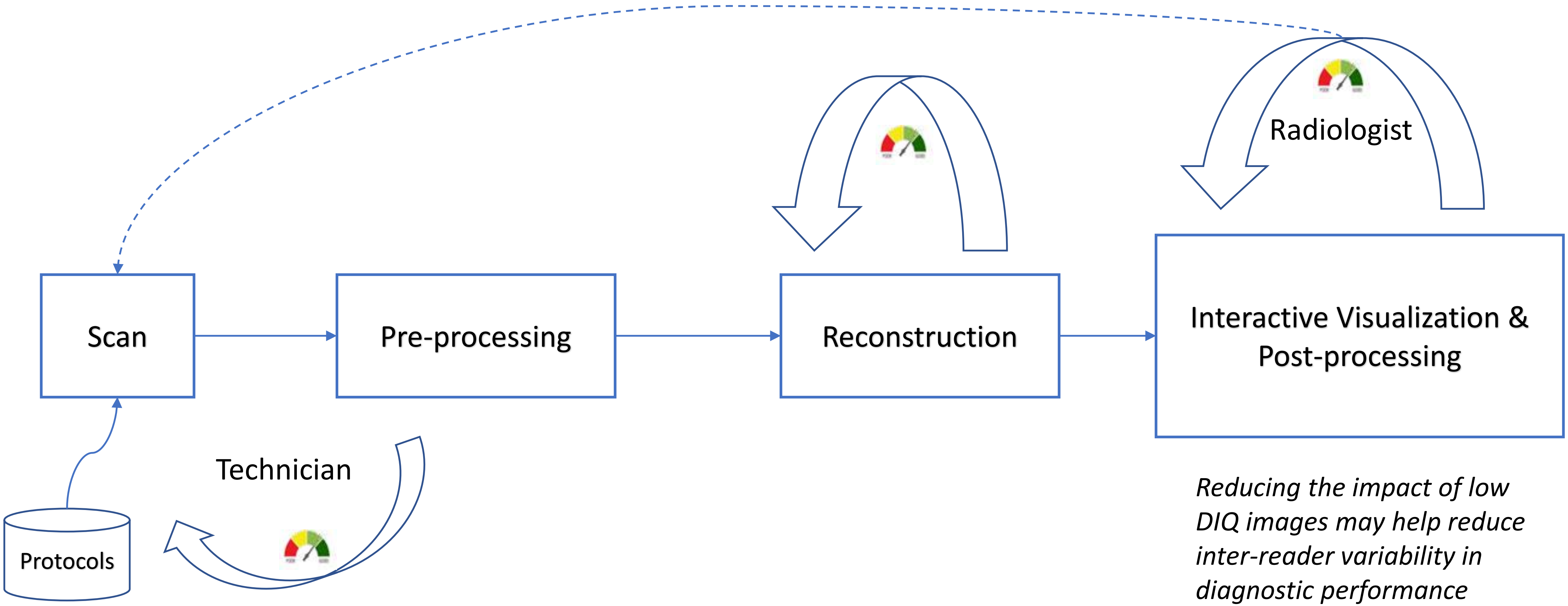

DIQM
Interpretability
Adequacy
Visual quality
I
A
V
I-A-V quality profiles may be associated to diagnostic categories
Subjective
Computable Image Features
Objective prediction
Trained & Tested

Possible DIQ Model Approaches

(a) Support vector regression model. (b) Heuristic model.



Applications: OperationalDIQMs



Reducing the impact of low DIQ images may help reduce inter-reader variability in diagnostic performance

Diagnostic Image Quality Metrics (DIQM)

Discussion:

- We propose a move from unidimensional to tridimensional, cognition inspired DIQ models
- New techniques such as NLP based mining of diagnostic quality comments alleviate the feature engineering problem
- Objective models can be made transparent by accounting for interpretability, adequacy and visual quality
- Operational models are possible by opening a window into the diagnostic quality criteria and making them available at quality control points of the diagnostic imaging chain
- DIQ models can be parametrized according to quality profiles