Cancer diagnosis should involve close coordination and communication among physicians in radiology, pathology, and clinical care in order to assure timely, efficient, and medically correct treatment decisions. The authors developed the web-based RadPath system to correlate and integrate diagnostic radiology and pathology report findings. RadPath retrieves reports, images, and test results from the institution’s clinical information system to construct a database of relevant information. Radiologists can correlate imaging findings with the pathology diagnosis; they otherwise might not receive feedback about pathology results related to cases in which they have interpreted diagnostic studies or performed biopsies. Medicine’s increasing complexity and team-based management approaches have demanded improved clinical communication and information exchange. The authors’ work is a welcome addition.

Hravnak M, Chen L, Dubrawski A, Bose E, Clermont G, Pinsky MR
Real alerts and artifact classification in archived multi-signal vital sign monitoring data: implications for mining big data
J Clin Monit Comput 2016 Dec;30(6):875-88
Hospital inpatients at high risk for cardiorespiratory instability undergo continuous non-invasive monitoring of their vital signs, such as heart rate, respiratory rate, blood pressure, and peripheral arterial oxygen saturation. To assure accurate diagnosis, it is critical to distinguish truly aberrant vital sign measurements from artifacts caused by artifacts such as patient movement, electrical noise, or loose sensors. The authors used machine-learning algorithms trained on labeled vital sign data to automatically classify alerts as real or artifact. Their algorithms achieved an area under the curve (AUC) of 0.72 to 0.94 on a test dataset, which suggests that their approach could be used for accurate automated classification of vital sign alerts as real or artifact. Such information provides a “data cleaning” step that can support data mining for future model building.

A comparison of lung nodule segmentation algorithms: methods and results from a multi-institutional study
J Digit Imaging 2016 Aug;29(4):476-87
Automated systems play an increasingly important role in identifying, classifying, and measuring lesions in medical images. These systems must provide accurate and reproducible volumetric measurements for diagnosis, staging, and assessment of treatment response. For lung nodules on CT, it is often the case that images are acquired on different scanners, generated with different imaging parameters (e.g., slice thickness, reconstruction kernel), and analyzed by different nodule-measurement systems. The authors evaluated three algorithms on 52 nodules from The Cancer Imaging Archive acquired from 40 actual patients and from an anatomic phantom. The authors found considerable differences between the algorithms’ measurements, especially in a subset of heterogeneous nodules. They recommend that longitudinal studies should use the same software at all time points.
radiography in lung cancer screening, and has led to the implementation of CT-based lung cancer screening programs. However, LDCTs result in a large number of false-positive results. The authors explored dynamic Bayesian networks (DBN) to help understand how longitudinal data might inform lung cancer screening decisions. The lung cancer screening DBNs demonstrated high discrimination and predictive power with the majority of cancer and non-cancer cases, and the DBN models outperformed logistic regression and naïve Bayes approaches. This work highlights the use of a large national dataset to create a model for decision support and demonstrates the potential of a prediction model to outperform radiologist interpretation in specific instances.

Springer DB, Tarassenko L, Clifford GD
Logistic regression-HSMM-based heart sound segmentation
Apr;63(4):822-32

Segmentation of heart sounds may be relatively simple in noise-free recordings, but noise – such as speech, patient motion, and intestinal or breathing sounds – makes the task more difficult. Neural networks and other static machine-learning approaches that lack temporal duration and ordering information may not perform as well in differentiating heart sound-like noises from noisy heart sounds. The authors sought to improve identification of first and second heart sounds in phonocardiogram (PCG) data. Their model incorporated a priori knowledge about the expected duration of the heart sounds encoded in a hidden semi-Markov model (HSMM). The authors evaluated their method on a recorded PCG dataset of more than 10,000 seconds from 112 patients. Their system’s average F1 score of 96% exceeded that of 86% achieved by conventional systems. Probabilistic models, such as hidden Markov models, can improve performance over threshold-based segmentation methods in the analysis of heart sound recordings.