Health systems will also benefit from clinical laboratorians’ insight on how ML can improve patient care using laboratory data.

**Barriers to Development and Adoption**

Three categories generally describe common approaches to ML: supervised, semisupervised, and unsupervised. Supervised ML relies on a large, accurately labeled dataset to train an ML model, such as labeling images of leukocytes as lymphocytes or neutrophils for subsequent classification. Currently, the consensus is that supervised ML will generate the best models for targeted detection of known classes of data. But in many cases the data sets required are not large enough or labeled accurately enough. However, the process of curating accurately labeled datasets is difficult and time-consuming.

With EHRs, researchers certainly have greater access to data than in years past. However, health information in its native state often is insufficiently structured for the rigorous development of ML models. For example, predictive alarms and syndrome surveillance tools that use supervised ML often rely on datasets delineated by the presence or absence of clinical disease. While ICD-10 codes are a discrete data element that could be used for labeling purposes, experience at our institution indicates that ICD-10 codes are not documented reliably enough to train supervised ML models.

To avoid performance issues associated with inconsistent labels, data scientists can curate custom labels based on specific criteria to define the classes in their datasets. But criteria for defining classes are often subjective and may lack universal acceptance. For example, sepsis prediction algorithms may rely on clinical criteria of sepsis used at one institution but not another. It will become increasingly important for clinical laboratorians to consider how models are trained and which specific clinical definitions define the functional ground-truth in an ML model for the classes or disease being detected.

In addition to issues with variable criteria for clinical disease, some labels also have intrinsic variability that may preclude ML from optimal performance across institutions. Linear models such as logistic and linear regression have shown poor generalizability between institutions (13, 14). In healthcare, the problem is multifactorial and may result from population heterogeneity, or from discrepancies between the ML training population and the use case or test population. Consequently, ML models trained outside one’s institution may benefit from retraining before go-live. However, nothing in the literature supports this practice.

Lastly, the black box nature of ML models themselves poses a well-described barrier to adoption. Computer scientists have sought to elucidate how and why models arrive at the answers they generate in order to demonstrate to end users the decision points used to arrive at a given score or classification, often referred to as explainable artificial intelligence (XAI).

Proponents for XAI argue that it may help investigate the source of bias in an ML model in a scenario where a model is producing erroneous results. Ideally such a tool would also include interactive features to allow correction of the bias identified. However, as ML models become more powerful and complex, the ability to derive meaningful insight into their inner logic becomes more difficult. The practice of investigating methods for XAI is young, and its utility remains an open question.

**What’s Next?**

The powerful technology of ML offers significant potential to improve the quality of services provided by laboratory medicine. Early commercial and research-driven applications have demonstrated promising results with ML-based applications in our field. Despite nagging problems with model generalizability, oversight, and physician adoption, we should expect a steady influx of ML-based technology into laboratory medicine in the coming years. Laboratory medicine professionals will need to understand what can be done reliably with the