# Chapter 32
# Clinical Machine Learning in Action:
## CAD System Design, Development, Tuning, and Long-Term Experience

**Yoshitaka Masutani**
*The University of Tokyo, Japan*

**Mitsutaka Nemoto**
*The University of Tokyo, Japan*

**Yukihiro Nomura**
*The University of Tokyo, Japan*

**Naoto Hayashi**
*The University of Tokyo, Japan*

## ABSTRACT

*A machine learning-based method using a database of clinical data, such as Computer-Assisted Detection/Diagnosis (CAD), is one of the next key technologies in clinical imaging. The most important issue for machine learning, based on clinical data, is construction of the database, and one of the promising improvements this technology offers is in the field of in-hospital development because of increased data accessibility and periodical updates.*

*This chapter first discusses the database problems in CAD development comprehensively. Then, it introduces the authors' integrated platform, called the Clinical Infrastructure for Radiologic Computation of United Solutions (CIRCUS), for in-hospital research, development, use, and evaluation of clinical image processing. Based on the authors' clinical experience and the data collected through the CIRCUS system, they present research results on the improvement of CAD performance as well as simulated studies for additional learning. Finally, the authors' future plans, including radiologist-CAD collaboration beyond machine learning, are also discussed.*

DOI: 10.4018/978-1-4666-3994-2.ch032

## INTRODUCTION

Machine learning is one of the key technologies in advanced medical imaging research and is expected to become essential in clinical imaging. Among the clinical applications of machine learning technology, Computer-Assisted Detection/Diagnosis (CAD) is a promising example of supervised learning based on various types of image features. However, in spite of the strong expectations by radiologists, practical problems remain unsolved in the development and use of a clinically feasible CAD system. One of the most critical issues is the database, which is indispensable for supervised learning. The database problems cover a broad spectrum of issues, such as the accessibility to clinical data and the amount of time needed for gold standard input as the supervisory signal, and so on.

To solve these problems, this group of authors has been constructing an integrated CAD system aimed at efficient development and clinical use. This system, named CIRCUS (Clinical Infrastructure for Radiologic Computation of United Solutions), includes the multi-functions of database construction and updates, a platform for new CAD application development, interfaces for CAD use in clinical routines, and additional learning. CIRCUS primarily consists of two subsystems, a database system and a clinical server system, which have been in use at the University of Tokyo hospital since January 2009. Using our experience with design, development, tuning and long-term use of our CAD system, we discuss clinical machine learning in this chapter.

The objectives of this chapter are:

- Discussion of database problems and solutions;
- Introduction of our in-hospital CAD development project, known as the CIRCUS system, by including several clinical experiences and simulation studies; and
- Discussion of the importance of in-hospital CAD development and future plans.

This chapter is organized as follows. First, the problems in CAD development, which are primarily database problems, are formulated and discussed in the next section. The discussion leads to the importance of in-hospital management of CAD development, testing, and updating. Then, the CIRCUS system is introduced and the technical specifications and the clinical experiences are described in detail. The discussion of the technical specifications includes the interface of the CIRCUS system. The CIRCUS interface is designed not only for effectiveness in displaying CAD results for clinical routines and easy entry of clinical feedback as the supervisory signal for additional learning, but also for collecting data on each radiologist's characteristics in reading clinical information. The section describing our clinical experience with CIRCUS includes the detection performance and the changes made by updating the system with more than 3200 cases of lung nodule detection in Computed Tomography (CT) images and also cerebral aneurysm detection in Magnetic Resonance Angiography (MRA) images. Based on the database and the clinical feedback, the results of simulation studies are presented to display the learning curves from initial learning to additional learning with various scenarios of the addition of more samples. The discussion on additional training leads us to suggest a better strategy for classifier data updates from the machine learning viewpoint. Finally, after summarizing the importance of in-hospital CAD development, we discuss the possibility of personalized/customized CAD as a future goal for maximizing the performance of collaborative detection in consideration of the issues of cognitive psychology.

## BACKGROUND

In medical imaging fields, the automated detection/classification of lesions by Computer-Assisted Detection/Diagnosis (CAD) is one of the most anticipated applications of supervised learning. In