

To cite this paper or access its final version please use the following information:

Yanase J, and Triantaphyllou E, 2019, **A Systematic Survey of Computer-Aided Diagnosis in Medicine: Past and Present Developments**, *Expert Systems with Applications*, Vol. 138, in press. <https://doi.org/10.1016/j.eswa.2019.112821>

A Systematic Survey of Computer-Aided Diagnosis in Medicine: Past and Present Developments

Juri Yanase

jurijuriy@aol.com

Complete Decisions, LLC

Baton Rouge, LA 70810

Evangelos Triantaphyllou

etriadantaphyllou@yahoo.com

Division of Computer Science & Engineering

School of Electrical Engineering and Computer Science

Louisiana State University

Baton Rouge, LA 70803

Abstract: Computer-aided diagnosis (CAD) in medicine is the result of a large amount of effort expended in the interface of medicine and computer science. As some CAD systems in medicine try to emulate the diagnostic decision-making process of medical experts, they can be considered as expert systems in medicine. Furthermore, CAD systems in medicine may process clinical data that can be complex and/or massive in size. They do so in order to infer new knowledge from data and use that knowledge to improve their diagnostic performance over time. Therefore, such systems can also be viewed as intelligent systems because they use a feedback mechanism to improve their performance over time. The main aim of the literature survey described in this paper is to provide a comprehensive overview of past and current CAD developments. This survey / review can be of significant value to researchers and professionals in medicine and computer science. There are already some reviews about specific aspects of CAD in medicine. However, this paper focuses on the entire spectrum of the capabilities of CAD systems in medicine. It also identifies the key developments that have led to today's state-of-the-art in this area. It presents an extensive and systematic literature review of CAD in medicine, based on 251 carefully selected publications. While medicine and computer science have advanced dramatically in recent years, each area has also become profoundly more complex. This paper advocates that in order to further develop and improve CAD, it is required to have well-coordinated work among researchers and professionals in these two constituent fields. Finally, this survey helps to highlight areas where there are opportunities to make significant new contributions. This may profoundly impact future research in medicine and in select areas of computer science.

Key Words: Computer-Aided Diagnosis; Computer-Aided Detection; Expert and Intelligent Systems; Computerized Signal Analysis; Segmentation; Classification.

1. Introduction

Computer-aided diagnosis (CAD) systems in the medical field can be viewed as cutting-edge expert and intelligence systems in the interface of medicine and computer science. CAD systems in medicine may use diagnostic rules to emulate the way a skilled human expert makes diagnostic decisions. In this sense such CAD systems function as expert systems. More advanced CAD systems have the capability to analyze clinical data and infer new knowledge. In turn, this new knowledge can enhance current diagnostic rules and enable such systems

to improve their performance over time. For this to occur, these systems must also have a feedback mechanism so they can infer new knowledge from different groups of data, but most importantly, from successes and failures. In this sense this type of CAD systems may be considered as intelligent systems as they have a learning capability [Obermeyer and Emanuel, 2016]. Intelligent CAD systems use artificial intelligence (AI), data mining and machine learning approaches to analyze clinical data that often can be complex and/or massive. Such systems can offer great help in making diagnostic decisions in a wide range of diseases and medical conditions [Sun, et al., 2016], [Giger, 2018].

Our knowledge of human anatomy and physiology has increased significantly during the last century. Human body examination tools such as ones based on X-rays, ultrasounds, or magnetic resonance imaging (MRI) have improved dramatically. As an indicator of the degree of change, one may consider the changes in the world's oldest and most famous medical textbook; *the Merck Manual of Diagnosis and Therapy* [Beers and Berkow, 1999]. In 1899 its first edition had 192 pages. More than a century later, its 2011 version (19-th edition) has more than 3,750 pages [Porter, et al., 2011], [Crowley, 2011]. While numerous diseases have been discovered and the study of these diseases has progressed considerably, the medical diagnosis of many diseases has become more complex and difficult.

A number of factors have led to the development of computer-aided diagnosis in medicine. Such factors include the following: the complexity of the medical diagnosis process itself, the availability of large amounts of complex clinical data relevant to many diseases and conditions, the existence of large amounts of diagnostic knowledge (for instance, in the form of diagnostic rules), and new advances in computer science (especially in the fields of AI, data mining, and machine learning).

Advances in CAD and the increasing acceptance of its great potential in medical diagnosis has created a fertile environment for researchers in the fields of expert systems and intelligent systems to pursue new and exciting opportunities in medicine and computer science. However, the description of such developments so far has been fragmented and done in a sporadic manner.

The novelty of this study is based on the objectives it targets. The main objective of this paper is to provide a comprehensive and systematic review / survey of the most important developments related to CAD in medicine. This paper bridges a gap that currently exists in the literature and thus it can be of benefit to medical professionals who do not have a deep knowledge of computer science issues. Likewise, this paper can also be of benefit to computer scientists who are not familiar with some basic medical issues related to CAD. By better understanding the history of CAD and its current state-of-the-art, one may better recognize what are the critical issues in CAD development today and also identify opportunities for advancing this important field. The need to pursue the above goals and objectives is the motivation for this study.

There are several alternative terms that can be used to describe computer-aided diagnosis / computer-aided diagnostic systems. Such alternative terms are, for instance, expert systems (ES) in medicine [Buchanan and Shortliffe, 1984], [Waterman, 1985], [Liao, 2005], computer-aided evaluation / diagnosis, computerized sound analysis (e.g., computerized lung sound analysis (CLSA) and computerized heart sound analysis) [Gurung, et al., 2011], computerized biomedical signal analysis [Theis and Meyer-Bäse, 2010] and so on.

In the mid-1980s, medical physicists and radiologists began to focus on the aspects of computer-aided detection (CADE) and computer-aided diagnosis (CADx). The authors in [Lemaître, et al., 2015] state the following: "In fact, these investigations brought about both concepts of Computer-Aided Detection (CADE) and Computer-Aided Diagnosis (CADx) grouped under the acronym CAD." Therefore, in this study, the term "CAD" is used to describe such systems in medicine. In addition, the acronym CAD may be used to describe any computer-based technology, which can be used to support medical professionals in providing accurate diagnosis to patients.

This paper is organized as follows. Following this introduction, the next section describes the methods used to conduct the survey / review for this study. The third section describes a brief history of CAD. Section 4

describes the basic algorithmic aspects of expert systems and machine learning approaches as they are important to better understand some key CAD issues. Section 5 presents the main objectives that CAD needs to satisfy. The main application areas of CAD are illustrated in Section 6. The seventh section explains the workflow of a typical CAD system during its development stage and also under clinical use. Some of the challenges / limitations of CAD are highlighted in Section 8. Finally, this paper ends with Section 9 which presents a summary of the main contributions and describes some possible areas for future research.

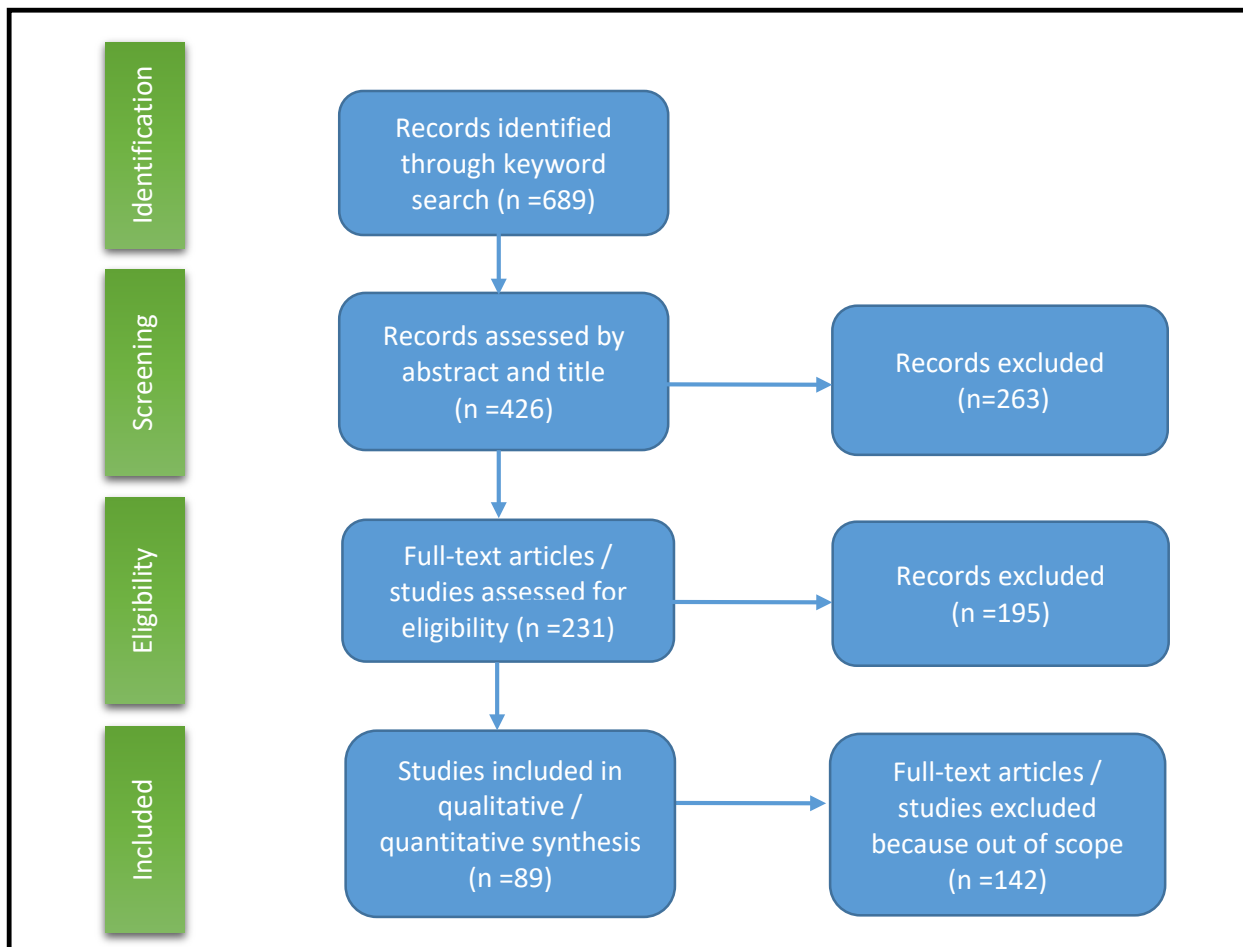


Figure 1: Flow Diagram of the Record Selection Process Based on the PRISMA Approach.

2. Methods

The authors conducted a systematic survey /review of publications that address various aspects of CAD systems in medicine. The literature search protocol followed for the review is the well-accepted PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) approach [Moher, et al., 2009]. This is shown in Figure 1. The keywords or key phrases used in the initial phase of this review are listed in Table 1. As the survey was progressing, the set of the keywords was expanded in a hierarchical manner. As more suitable publications were identified, more search terms were considered.

The 4-phase flow of the PRISMA approach (labeled “Identification,” “Screening,” “Eligibility,” “Included,” in Figure 1) was used as a model for narrowing down an initially very large number of publications. The bulk of the literature search was conducted in the spring of 2017 and the publication dates under consideration were set to be from 2013 to the present. The PubMed database was primarily used to identify such publications. Google’s Scholar function was used later in this review search as well.

The publications have been screened for inclusion eligibility. For instance, some publications were duplicates and others mentioned CAD not in fields related to detection or diagnosis in medicine. Eventually, out of a very large initial collection of publications, 89 publications were selected for analysis for this study. This number was later expanded to 251 references used in this study. Both authors were involved in the selection of the publications. After the initial search was performed, the authors identified a list of the aspects that CAD systems had to possess.

Both authors completed the analysis of the results. Publications used in this study included peer-reviewed research papers, books, conference papers, newspaper and magazine articles, and other academic literature (such as theses and dissertations). These publications had to satisfy certain criteria. They had to have appeared in reputable publication media in the field, their authors had to be from reputable institutions, and be written in English.

Author ET led the decisions regarding assessing publication media. He excluded journals described in the lists of predatory journals created by [Beall, 2017] or mentioned by Yale University [Yale University Library, 2017]. In addition, author ET referred to well-known websites for classifying author affiliations and institutional relationships as reputable. These websites included the *World Reputation Rankings 2016* [Times Higher Education, 2017], the *U.S. News Best Colleges Rankings* [U.S. News, 2017], and the *Nature Index* [2017]. Some of these aspects are qualitative and are hard to be quantified. Thus, we used a holistic approach and did not rely on a single aspect alone.

Table 1: The Keywords Used for Searching Publications.

The initial keywords / key phrases	“computer-aided diagnosis”; “computer-aided diagnostic system”; “expert system” medicine; “computer-aided detection”; “computerized sound analysis”; “computerized biomedical signal analysis”
---	--

3. A Brief History of CAD

Soon after the dawn of the computer age, in the late 1950s early biomedical researchers begun exploring the possibility of using computers to investigate and solve problems in biology and medicine. Some of those studies were ultimately directed to the development of systems for computer based medical diagnosis [Weinrauch and Hetherington, 1959], [Ledley, 1959], [Vandenberg, 1960]. These early diagnostic systems, often referred to as “*expert systems in medicine*”, used patients’ symptoms and laboratory test results as inputs to generate a diagnostic outcome [Ledley and Lusted, 1959], [Weiss, et al., 1978]. A detailed description of expert systems in CAD is provided in Sub-Section 4.1.

Some noticeable early CAD systems were the MYCIN expert system [Shortliffe and Buchanan, 1975], the INTERNIST-I expert system [Miller, et al., 1982] and later, in the middle of the 1980s, the CADUCEUS expert system [Feigenbaum and McCorduck, 1984]. However, by the early 1970s, it became clear that there were some serious limitations in delivering accurate diagnosis when using traditional methods such as flow-charts [Schwartz, 1970], [Bleich, 1972], statistical pattern-matching [Rosati, et al., 1975], or probability theory [Gorry, et al., 1973], [Szolovits, et al., 1988].

This realization of the computational limitations was unavoidable. Early expectations of the potential of the newly emerged computer based approaches (and not only in the field of medical diagnosis) tended to be overly optimistic. At the beginning of these early studies, researchers were hoping to develop entirely automatic computer-aided diagnostic systems. The expectations of people of what computers could do were frequently grossly unrealistic. This situation changed dramatically thanks to the seminal contribution in the early 1970s in theoretical computer science by Karp. In particular, his breakthrough paper “*Reducibility among Combinatorial Problems*” established for the very first time certain limitations but also opportunities that exist while developing algorithms to solve families of important computational problems [Karp, 1972]. The limitations and potential solutions of many computational problems related to medical diagnosis can be better understood in light of this

crucial development. After these early developments, researchers have instigated the development of systems that use artificial intelligence and specialized computer algorithms (i.e., pattern recognition / classification algorithms) to diagnose diseases based on patients' data [Musen, et al., 2014]. Today, CAD is considered more as an important part of a diagnostic process which also actively involves human experts [Doi, 2007].

Besides the previous developments, during the 1960s a number of radiologists had also started working on building an early form of CAD in order to detect abnormalities in medical images [Lodwick, et al., 1963], [Meyers, et al., 1964], [Winsberg, et al., 1967]. Today diagnostic radiology and medical image analysis are some of the most active research and application areas in CAD [Doi, 2007].

In the meantime, medical malpractice litigation has significantly increased since the mid-1980s [Mohr, 2000]. Rising medical malpractice liability insurance costs have caused a negative effect on the healthcare industry and thus the cost of healthcare has increased dramatically [Brennan, 2004], [Nahed, et al., 2012]. One of the motivations to develop CAD was to help physicians to avoid medical malpractice cases and thus to reduce medical healthcare costs.

4. Key Algorithmic Issues of Expert Systems and Intelligent Systems in CAD

As it was noted in the previous section, the first CAD approaches were based on expert systems. This was followed by CAD approaches based on intelligent systems. Thus, the following two sub-sections provide some key technical aspects of these two major approaches in CAD.

4.1 Expert Systems in CAD

The structure of a generic expert system is depicted in Figure 2. As it can be seen from this figure, the user interacts with the system's knowledge base via the inference engine. The inference engine may use a forward or backward chaining approach on the knowledge base to reach a decision / recommendation. The knowledge base is built by a knowledge engineer (usually a computer science person with access to the domain knowledge pertinent to the aim of the expert system). The knowledge base is comprised of a set of facts that are assumed to be true and a set of applicable diagnostic rules.

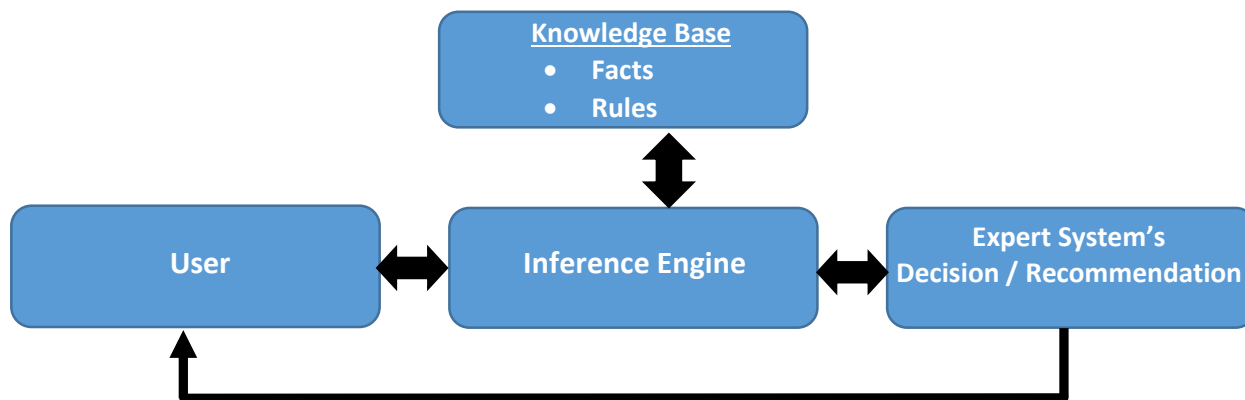


Figure 2: A Simplified Architecture of a General Purpose Expert System.

```

IF    [(Premise P1) AND (Premise P2) AND ... (Premise Pn)]
THEN  (Premise Pk)

```

An example from MYCIN:

```

IF:   [ (The infection which requires therapy is meningitis)   AND
          (Organisms were not seen on the stain of the culture) AND
          (The type of the infection is bacterial)              AND
          (The patient has been seriously burned) ]
THEN: [ (There is suggestive evidence (0.5)) (pseudomonas-aeruginosa
          is one of the organisms which might be causing the infection) ]

```

Figure 3: The General Structure of an IF-THEN type of Logical Rule and an Example from the MYCIN Expert System [Buchanan and Shortliffe, 1984, page 476].

```

Algorithm Forward_Chaining (Input_Rules1, Input_Facts1, Goal1);
Begin /* Algorithm <Forward_Chaining> */
/* It returns TRUE if Goal1 can be inferred from Input_Rules1 when applying Input_Facts1 */
Goal1 is not satisfied (= FALSE); /* Initialization */

IF (Goal1 is a member of Input_Facts1) THEN return Goal1 is satisfied (= TRUE) and STOP;
ELSE
Begin /* else #1 */
  Fire_Applicable_Rules (Input_Rules1, Input_Facts1, Output_Rules1, Output_Facts1);
  /* This fires all applicable rules in Input_Rules1. Any fired rules are removed from
  Input_Rules1 and the new set is called Output_Rules1. Any new conclusions are
  added in Input_Facts1 and the new set is called Output_Facts1. */

  /* If the reduced set of rules is empty, then stop and return
  Goal1 is not satisfied (= FALSE). We cannot prove Goal1
  Otherwise, call recursively Forward_Chaining */
  IF (Output_Rules1 = empty) THEN Goal1 cannot be proven and return
  Goal1 is not satisfied (= FALSE); STOP;
  ELSE
  Begin /* else #2 */
    (Forward_Chaining (Output_Rules1, Output_Facts1, Goal1);
    IF (Goal1 is satisfied (= TRUE) THEN STOP;
  End /* else #2 */
End /* else #1 */
End /* Algorithm <Forward_Chaining> */

```

Figure 4: The Forward Chaining Algorithm in Pseudo Code.

Usually the rules are logical structures of the IF-THEN form as shown in Figure 3. A typical rule is comprised of the antecedent part and the consequent part which is assumed to be true if the antecedent part is true. The antecedent part is a logical expression that can be either true or false. In Figure 3 the antecedent part is the conjunction of n premises denoted as premises P_1, P_2, \dots, P_n . In turn, each premise is either true or false. The consequent is a premise too. In MYCIN each rule was also associated with a probabilistic assessment (termed the “*suggestive evidence*” in Figure 3, also known as the “*certainty factor*” or CF) that denotes the likelihood the consequent will be true if the antecedent is true. Expert systems were created in an effort to emulate the

diagnostic decision-making process that humans are supposed to follow. Such knowledge by humans can be easily represented in the form of a set of IF-THEN type of rules, hence the creation of the knowledge base as a collection of such rules and facts.

When the inference engine uses forward chaining, a target goal (such as a medical hypothesis) is given. Next, the knowledge base is accessed by the inference engine to determine whether the set of rules along with the available facts, can be used to determine the true/false value of the target goal. A description of the forward chaining search approach is given in Figure 4. When the algorithm Forward_Chaining (as described in Figure 4) is applied with <Input_Rules1> being the set of the three rules shown in Figure 5, <Input_Facts1> being the facts in the same figure and <Goal1> = <Z>, then the algorithm will return value TRUE, meaning the validity of <Z> can be established from this knowledge base.

Backward chaining works in an analogous manner. Assume again the previous data. Since the goal is the search of the value of <Z>, Rule #1 is examined first. The value of its premise <Y> is still unknown and this triggers Rule #2 to be examined. This in turn triggers Rule #3 to be examined in order to determine the value of <X>. This value is true, thus now Rule #2 is fired and the value of <Y> becomes true and next Rule #1 is fired and the value of <Goal1> (i.e., <Z>), is determined to be true.

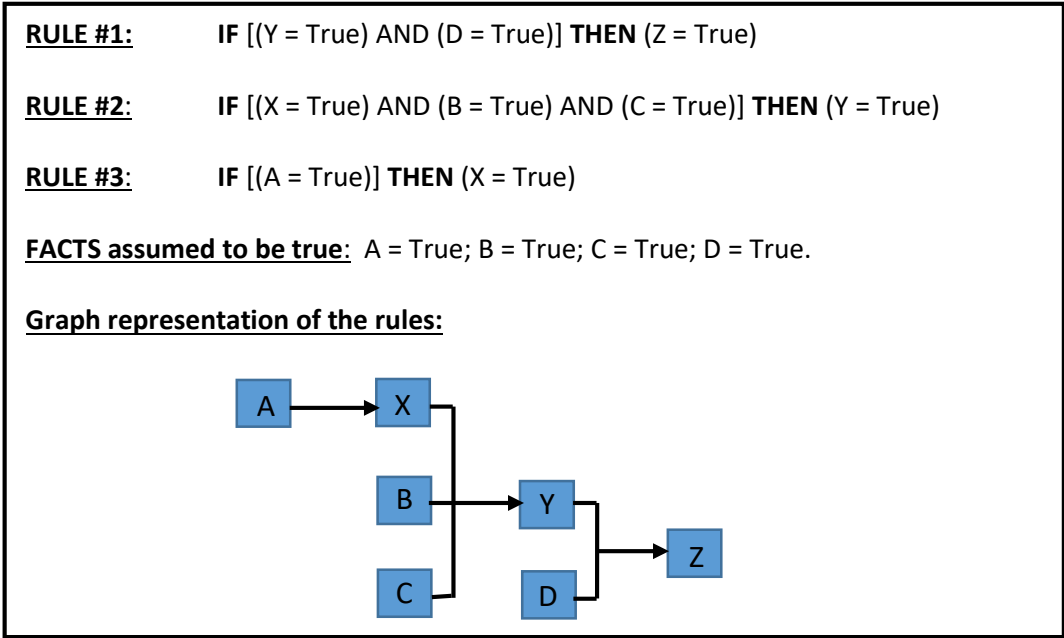


Figure 5: A Simplified Knowledge Base for Illustrative Purposes.

A key issue with the use of expert systems is how to validate and verify their knowledge bases and the way they operate. For that purpose, they were enhanced by also using an “Explanation” module (not shown in Figure 2). The problem becomes more profound when one realizes that knowledge bases were of increasing size and complexity. For instance, it was proposed that having a knowledge base with 100 million rules would result in a system of immense capabilities [Lenat and Feigenbaum, 1992].

Checking for the logical consistency of a set of interrelated logical rules results in the formulation of a satisfiability (SAT) problem [Bezem, 1988]. If one assumes only binary variables, say n of them, then the corresponding search space is of size 2^n . That is, it can become very large quickly. This is an NP-complete problem very susceptible to the “dimensionality curse” problem [Hansen and Jaumard, 1990]. Another problem related to the development and use of expert systems is the use of probabilities. For instance, the approach based on the use of the “suggestive evidence” concept in MYCIN, was criticized as ineffective [Bramer, 1982], [Torasso, 1985]. Another issue is related to the weakness of such systems to generalize to cases not explicitly described in the knowledge base. Another problem, also related to the knowledge base, is how to make updates of its

knowledge quickly and effectively [Shan and Ziarko, 1995], [Coats, 1988], [Hendriks and Vriens, 1999]. There are also questions on how to prioritize the use (firing) of its rules in order to operate more efficiently, or how to resolve ambiguities (for instance, if there are many else-if sub-structures within a single rule) and so on [Mak, et al., 1997].

For all the above reasons, a new type of computer-aided diagnosis approaches was pursued next in order to mitigate the previous problems. The new type of approaches is based on the use of machine learning techniques as it is described in the following sub-section.

```

Input:  $D$  = training dataset defined on attributes and with class values;
Output: Decision tree <Tree_ $T_0$ > capable of classifying the training dataset  $D$ ;

Algorithm Build_Decision_Tree ( $D$ )
Begin /* Algorithm <Build_Decision_Tree> */
Tree_ $T_0$  = { }; /* initialization */

/* If we have only one class (i.e.,  $D$  is "pure") or a stopping criterion is invoked, then stop.
We have reached a "leaf" node of the decision tree.
Otherwise, determine the best attribute to split the remaining of the data.
Proceed recursively until only "leaf" nodes are created and no more splits are possible */
IF ( $D$  is "pure" or a stopping criterion is invoked) THEN
    (Create a leaf node that corresponds to  $D$  and STOP); /* Tree_ $T_0$  has only this node */
ELSE
Begin /* else #1 */
    FOR each attribute in  $a \in D$  DO
    begin /* do-loop #1 */
        Compute the information-theoretic evaluative value if we split on  $a$ 
    end /* for-loop #1 */
     $a_{\text{BEST}}$  = Most promising attribute for a split;
    Tree_ $T_0$  = Create a decision node based on attribute  $a_{\text{BEST}}$ ;
     $D_i$  = Induced sub-datasets from  $D$  based on  $a_{\text{BEST}}$ ;
    FOR each sub-dataset  $D_i$  DO
    begin /* do-loop #2 */
        /* Call recursively Algorithm <Build_Decision_Tree> with argument  $D_i$  */
        Tree_ $T_i$  = Build_Decision_Tree ( $D_i$ );
        Attach Tree_ $T_i$  to the corresponding branch of Tree_ $T_0$ ;
    end /* for-loop #2 */
End /* else #1 */
RETURN Tree_ $T_0$ ;
End /* Algorithm <Build_Decision_Tree> */

```

Figure 6: Pseudo Code Description of a Generic Algorithm for Inferring a Decision Tree from Training Data.

4.2 The Emergence of Intelligent / Machine Learning Techniques in CAD

There are two main families of machine learning approaches. Those based on supervised and those based on unsupervised learning. Both approaches use datasets for training, that is, for inferring a model that could next be used on new data points and derive useful information. The inferred model is also known as "the new knowledge discovered from the data," as it can be used to assist in the diagnostic process. In supervised learning the training dataset is comprised of observations (data points) along with a class value. For instance,

one may have descriptions of mammograms some of which correspond to benign cases while the rest to malignant cases (in a simplified two-class setting).

In unsupervised learning the data are not associated with a class variable. Usually, the main goal is to explore whether the data can be grouped into clusters, which in turn may provide useful information for the diagnostic process. Another goal might be to identify patterns among the attributes of the data in the form of association rules [Hastie, et al., 2009]. Naturally, unsupervised learning is more challenging than supervised learning as there is less information available to work with.

The benefit of using machine learning approaches is that vast amounts of data can be analyzed and potentially useful information (in the form of inferred models) may be discovered. If new data arrive, one may incorporate them by simply re-training the model. Such approaches are also easier to check for data consistency, incorporate inconsistent data, deal with much higher dimensionalities, or derive probabilistic assessments. Such approaches may be characterized as “intelligent” as they can improve their performance over time by utilizing information from both successes and failures of the models they derive.

One of the most-well known supervised approaches is based on the inference of decision trees [Tan, et al., 2006]. A decision tree essentially partitions the training dataset into groups in a hierarchical (tree-like) manner where only one class is the predominant one in each group. Information along the root node and a leaf node corresponds to a pattern on the input attributes which can be used to classify new data points of unknown class value. Figure 6 provides the pseudo code of the typical approach for inferring decision trees.

```
Input: D = the training dataset /* each data point is associated with a class value */
        K = number neighbors to consider (size of dataset D >= K >=1)
        Y = a new data point of unknown class value that we need to classify.
Output:The inferred class value of the new data point Y.

Algorithm K- Nearest_Neighbors (D, K, Y, Class value of Y);
Begin /* Algorithm K-Nearest_Neighbors */
  FOR (each data point  $X_i$  in dataset D) DO
  begin /* do-loop #1 */
    Compute the distance between the new point Y to  $X_i$ ;
  end /* do-loop #1 */

  /* Find the closest K neighbors to the new point Y */
  Determine K_Closest_Neighbors = set of K data points from D that are closest to Y;
  /* Determine the majority class. Ties can be broken randomly */
  Majority_Class = the majority class of the members in K_Closest_Neighbors;
  RETURN Majority_Class as the class of the new point Y;
End /* Algorithm K-Nearest_Neighbors */
```

Figure 7: The K-Nearest Neighbors Approach to Classification.

An alternative approach is one based on geometric properties among the data. It is known as the K-nearest neighbors approach. When a new data point, say point Y, is presented for classification, all the data points in the training dataset are examined and the K points which are closest to Y are selected. The majority class of this group of K points is determined and it is assigned as the class value to the new point Y [Tan, et al., 2006]. A key question is how to select the “correct” value for K (a positive integer between 1 and the size of the training dataset). This can be done by considering many candidate values for K, and then checking the performance of the inferred model on a testing dataset (which can be different of the training dataset but also with data of known class values). The model with the best performance is used to derive the “best” K value. Another critical issue is how to define the distance between pairs of data point. There is not a universally

accepted approach for doing this and it is application depended. This is a crucial aspect for this approach and thus can be a contested issue. Figure 7 presents the pseudo code of a typical K-nearest neighbors approach.

The support vectors machine (SVM) approach is based on optimization [Tan, et al., 2006]. In the simplest form it tries to define two parallel hyperplanes (lines in 2-D) which can completely separate data grouped into two disjoint classes. Figure 8 presents an outline of this case

Input: Two disjoint sets of observations (X_i, Y_i) , where X_i is an n -dimensional vector with the attribute values of the i -th observation and Y_i is its class value ($Y_i = \{+1, -1\}$).

Output: W is an n -dimensional vector that determines the distance between two parallel hyperplanes that separate the two classes of observations.

Solve the following quadratic problem (the unknowns are $W = (w_1, w_2, w_3, \dots, w_n)$ and b):

Minimize $\|W\|^2$

Subject to: $W X_i + b \geq 1$, for all data points with $Y_i = +1$.
 $W X_i + b \leq -1$, for all data points with $Y_i = -1$.

Figure 8: The Optimization Formulation of the Support Vector Approach (SVM). The Linearly Separable Case.

If the above problem is infeasible (i.e., the data are not linearly separable), then the constraints can become “soft” by introducing some positive quantities ξ_i . Now we have a linearly non-separable case. The new objective function tries to minimize the old one plus a penalty component defined on the previous ξ_i quantities (where $\xi_i \geq 0$). The new optimization problem is as follows (where C is a user-defined parameter):

Minimize $\|W\|^2 + C \times \sum_i \xi_i$

Subject to: $W X_i + b \geq 1 - \xi_i$, for all data points with $Y_i = +1$.
 $W X_i + b \leq -1 + \xi_i$, for all data points with $Y_i = -1$.

A crucial issue with this formulation is how to choose the value of C (i.e., the weight of the part that minimizes the sum of the ξ_i values).

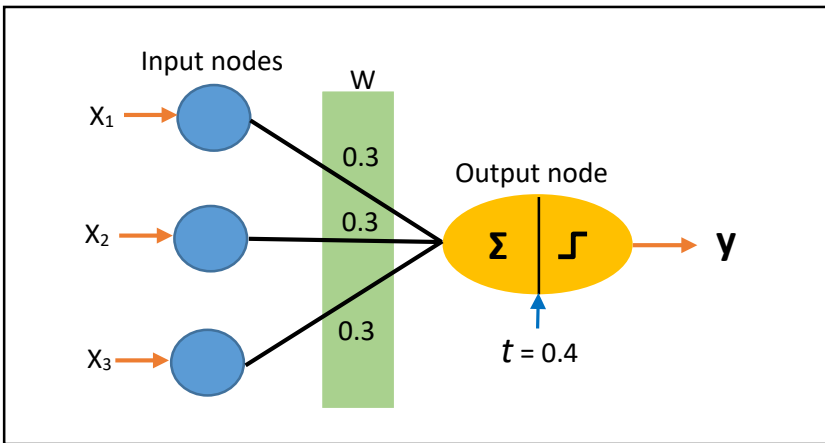


Figure 9: A Simple Network Representation of a Perceptron (based on [Tan, et al., 2006]).

Perhaps the oldest supervised machine learning approach is the one based on a concept known as the “*perceptron*.” A perceptron is an effort to emulate the way the human brain functions. It uses a simple network-based approach where each input field corresponds to an input node and a weighted approach is used to determine the predicted class value (see also Figure 9). If the weighted value of the inputs in terms of the weights in vector W is greater than or equal to a control parameter t (see also Figure 9), then this observation is assigned to the first class (in a two-class setting). It is assigned to the other class, otherwise. Figure 10 presents a pseudo

code description of the perceptron approach. It is also based on some control parameters. These parameters can be calibrated by using a testing dataset, as is the case with many such algorithms.

```

Input: D = the training dataset /* each data point  $X_i$  is associated with a class value  $Y_i$  */
           $\lambda$  = the learning rate (preset by the user. It is a real number in [0, 1])
Output: The weights  $W = (w_1, w_2, w_3, \dots, w_n)$  of a perceptron trained on dataset D

Algorithm Perceptron (D,  $\lambda$ , W);
Begin /* Algorithm Perceptron */
    K = 0; /* initializing the epoch counter */
    Randomly initialize the weights  $w^{(0)}_j$  in vector W /* should be  $\geq 0$  and add up to 1 */

    /* The stopping condition may be defined when all the data points have been
       considered and the approach run a preset number of epochs */
    DO UNTIL (stopping condition is met)
    Begin /* do-until loop */

        /* The following do-loop represents one epoch */
        FOR (each training data point  $(X_i, Y_i)$  in dataset D) DO
            begin /* do-loop #1 */
                Compute the predicted output  $\hat{Y}^{(K)}_i$  of the current data point;
                FOR (each weight  $w_j$ ) DO
                    begin /* do-loop #2 */
                        Update the weight:  $w^{(K+1)}_j = w^{(K)}_j + \lambda(Y_i - \hat{Y}^{(K)}_i)x_{ij}$ ;
                    end /* do-loop #2 */
                end /* do-loop #1 */
            K = K + 1; /* update the epoch counter */
        End /* do-until loop */
    RETURN vector W with all the weights;
End /* Algorithm Perceptron */

```

Figure 10: Pseudo Code of the Perceptron Learning Approach.

A generalization of the simple perceptron idea is what is known as the artificial neural networks (ANN). This is like having multiple perceptrons. Figure 11 presents a graphical representation of an ANN. The main difference now is the introduction of the hidden layer. An ANN with a single hidden node is the same as the previous perceptron. An ANN can be inferred (i.e., its weights W can be determined) from a dataset in a way that is more general than that for a perceptron but the results are more powerful too [Tan, et al., 2006].

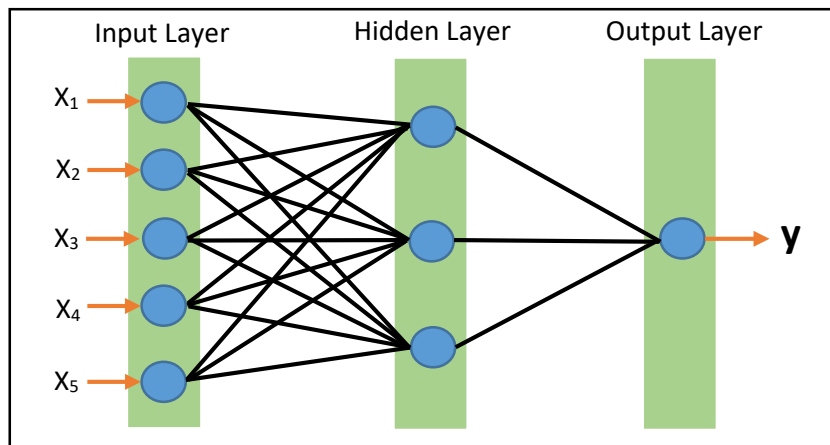


Figure 11: A simple Artificial Neural Network (based on [Tan, et al., 2006]).

The K-means approach is an unsupervised learning one [Tan, et al., 2006]. No class values are considered now. The goal is to determine clusters such that each cluster is comprised of highly similar members (data points), while members across different clusters are not as similar. Thus, a way for defining a distance between pairs of data points is needed. This is a crucial aspect with such approaches (as was the case with the K-nearest neighbors approach too). Figure 12 presents a pseudo code description of this method.

```

Input: Dataset D, K = number of clusters; /* data points do not have class values */
Output: Clusters and their corresponding centroids;

Algorithm K-Means_Clustering ( D, K, Clusters, Centroids);
Begin /* Algorithm K-Means_Clustering */
Randomly assign K data points as the initial centroids  $C_1, C_2, C_3, \dots, C_K$ ;

REPEAT UNTIL (centroids do not change OR a preset number of iterations has been reached)
Begin /* repeat-loop #1 */
  FOR (each data point  $X_i$  in dataset D) DO
  begin /* do-loop #1 */
    Compute the distance of point  $X_i$  to each centroid;
    Determine the closest cluster when comparing the above distances;
    Assign point  $X_i$  to the closest cluster;
  end /* do-loop #1 */

  FOR (each centroid  $C_1, C_2, C_3, \dots, C_K$ ) DO
  begin /* do-loop #2 */
    Updated centroid = mean of all points assigned to that cluster
  end /* do-loop #2 */
End /* repeat-loop #1 */
RETURN the K clusters and their corresponding centroids;
End /* Algorithm K-Means_Clustering */

```

Figure 12: The K-Means Approach to Clustering (an Unsupervised Machine Learning Approach).

As it is explained in the following sub-sections, machine learning approaches are used in many aspects of medical diagnosis. Thus, they are integral components of many CAD systems.

5. The Objectives of CAD

CAD can bring several benefits to the current diagnostic practice in clinics and hospitals. However, to be successful, CAD needs to meet certain objectives as it is described next.

5.1 Managing Large Volumes of Clinical Data

In order to provide accurate clinical diagnosis to a patient, medical professionals frequently have to analyze various types of clinical data. A description of clinical data is summarized in Table 2 [Data Resources in the Health Sciences, 2017]. Such data are usually a patient’s clinical information including laboratory test results, other findings and physical symptoms. Other data can be a patient’s medical history with his/her past medication records and the history of his / her current problems. Genetic aspects can be critical elements of some diseases too. Thus, a patient’s family medical history may be essential as well. In addition, a patient’s social status, where he/she lives or has lived, diet and eating habits, and lifestyle can be important key factors for the diagnosis of his / her medical problems.

Table2: Data Description for CAD Use.

<p>Clinical Data</p> <ul style="list-style-type: none"> Electronic Health Records (EHRs) Patient / Disease Registries 	<p>Clinical Data: Information collected during the course of the patient’s care.</p>
	<p>Electronic Health Records (EHRs): A digital version of a patient’s health records including physical test results, the patient’s history, symptoms, medical images, etc.</p>
	<p>Patient / Disease Registries: A special database or group of databases that contain information (such as medical conditions, exposure to certain environmental factors, geographic location, etc.) about people diagnosed with a specific type of disease.</p>

Many hospitals, medical facilities, and clinics are transferring their patients’ records from paper into digital files and they also employ electronic health records (EHRs). CAD systems may connect disease / patient registries, which hold valuable information about diseases and may help to diagnose the problem(s) that a patient has.

Clinical data keep becoming more sophisticated, complicated and increase dramatically in size [Bates, et al., 2014]. As such data become too large for humans to comprehend, it becomes too difficult for medical professionals to understand the whole spectrum of a patient’s conditions and diagnose the problem in a timely manner. Thus, CAD may have a remarkable capability to process and analyze large volumes of clinical data by applying computing power and using highly specialized computer algorithms.

5.2 Objective and Quantitative Judgments

Traditional human-based diagnostic approaches mainly depend on the judgments of the healthcare professionals, which sometimes can be subjective. If there is no gold standard or no established level of agreement about the criteria for a diagnosis, two different physicians may diagnose differently based on their understanding and experiences. In addition, decisions by physicians may be influenced by many kinds of issues, such as who the patients are, when, where, or why and how physicians make diagnostic decisions [Schiff, et al., 2009]. For instance, in a study by Elmore and his associates [Elmore, et al., 1994] it was found that 90% of radiologists had recognized fewer than 3% of the mammograms that they had screened 5 months prior to that time, while 10% had claimed to have recognized about 25% of the cases. Other similar studies are reported in [Kopans, 1994], [Gurney, 1994], [Boone, 1994], [Kovalerchuk, et al., 2000]. Sometimes data are fuzzy and then

one still needs to express them in an objective manner that is inherently consistent. In [Kovalerchuk, et al., 1997] or [Triantaphyllou, 2010] a fuzzy logic based approach is presented for describing mass micro-lobulation as part of breast cancer diagnosis.

Some healthcare professionals can be inexperienced and/or poorly trained. For example, the false positive rate of HIV rapid tests is quite high. One of the main reasons is that frequently these tests are run by staff members who do not meet sufficient qualifications [Grusky, et al., 2007]. Training procedures provided to community-based clinic staff fail to meet the standards set forth by the U.S. Center for Disease Control and Prevention (CDC) [Grusky, et al., 2007]. If test results are ambiguous and staff members are not qualified to administer and interpret the results, then the probability of having an inaccurate interpretation of the results will increase [Klarkowski, et al., 2009].

Another example is the basic auscultatory skills that medical professionals should have mastered. Auscultation is listening to the sounds produced by various organs of the human body and this practice is one of the most fundamental procedures in medical diagnosis. Using a stethoscope, medical professionals listen to the internal sounds of the human body and can detect any abnormal sounds. This basic skill, however, requires years of training and experience to master. Although this skill is very important, it is very difficult to teach in a structured way. The majority of internal medicine and cardiology programs offer poor training or no training at all. Consequently, it has been documented that some primary care physicians have poor auscultatory skills [Grusky, et al., 2007], [Roy, et al., 2002], [Roy, 2003], [Mahnke, et al., 2004]. In addition, traditional auscultation may involve subjective judgments by healthcare professionals, which can lead to an inconsistent observation and interpretation of the sounds [Kiyokawa, et al., 2001], [Abbas and Bassam, 2009].

Fatigue and distraction due to overwork or a night shift may also negatively impact healthcare professionals' performance and decisions [Petrick, et al., 2013]. Many medical trainees and nurses work long hours. It is common that trainees work more than 80 hours per week and even 100 to 120 hours per week, which is unfortunately typical [Landrigan, et al., 2004]. Under such circumstances, human errors are an unavoidable reality. Even under normal conditions, it is very difficult to quantify patients' information and accurately diagnose. Sleep deprivation of health care professionals may cause exhaustion and confusion when they are engaged in medical decision-making. As a result, under such conditions they often reach wrong diagnostic decisions [Gaba and Howard, 2002].

On the other hand, CAD does not depend on a single healthcare professional's analysis and skills. CAD can assist making more objective diagnostic decisions consistently. Furthermore, CAD can narrow the gap between experienced and inexperienced healthcare professionals in medical diagnosis.

5.3 Effectiveness and Efficiency

CAD systems can be cost-effective, especially if a disease is detected at an early stage and is treated before it progresses to advanced stages. A disease, which does not show any alarming signs or symptoms, is called an asymptomatic disease. Breast cancer and pancreatic cancer are some of the examples of asymptomatic diseases at early stages.

Many asymptomatic diseases in their early stages are easily treatable. If a person does not receive proper treatment, however, the disease would progress and can cause severe organ or nerve damage and eventually the person can become extremely sick. Unfortunately, treatments for such diseases in their later stages can be very expensive and not as effective. One of the most common treatments for a disease at a later stage is to have surgery, which is an invasive procedure. Surgery and other invasive procedures may involve high costs of many types such as physical pain, emotional distress, trauma and financial burden. There are many studies indicating the increasing risk of invasive procedures [Leape, et al., 1991], [Kable, et al., 2002], [Vincent, et al., 2004]. In addition, treatments for diseases at later stages can be often less effective and more likely not to be able to rebuild the full functionality of the organs that have been affected [Crowley, 2011]. In order to diagnose a disease

in its initial asymptomatic stages, screening is essential. Thus, many hospitals, clinics and screening centers are already employing CAD for breast cancer screening [Doi, 2007].

There is a limitation of the human ability to find abnormalities even when humans use specialized equipment [Kiyokawa, et al., 2001], [Reichert, et al., 2008]. However, CAD can interface with specialized equipment and, for instance, detect micro-calcifications that healthcare professional may fail to recognize [Helvie, et al., 2004], [Nishikawa, et al., 2012]. In addition, CAD may improve the workflow of a diagnostic procedure and reduce the search time for unclear calcifications [Lehman, et al., 2015].

6. Some Representative Application Areas of CAD

As CAD is used to diagnose a plethora of diseases and medical conditions of the human body, a question that arises is how to categorize its numerous application areas. There are two basic venues to do so. One is based on the specific diseases and medical conditions that CAD is used for. An alternative approach is to consider the various data types and data sources used in CAD. From the computer science point of view, the second approach seems to be the natural way to consider. Figure 13 depicts a conceptual organization of how CAD is applied today in numerous application areas. This figure presents three major groups as follows. The first group considers cases where the input data deal with sounds and signals related to various human organs. The second group considers cases where the input data are generated by analyzing images pertinent to the human body. There is a third group which considers the rest of the cases such as when special lab tests are performed (for instance, the VAP cholesterol test for lipid analysis).

The subdivisions under each one of these three major groups may also be combined with each other for the purpose of having a CAD system for a given disease or medical condition. For instance, a CAD system, say for the diagnosis of heart related conditions, may require the analysis of sound data (such as sounds generated when blood flows through arteries), electric signals (such as the ones generated by the heart itself), ultrasound images of the heart and lab data (such as the ones of VAP lipid testing). The sub-sections that follow discuss some representative CAD application areas in terms of these three major groups.

6.1 Sound and Signal Analysis

Electrical impulses can provide important medical information regarding various body functions [Crowley, 2011]. Digital signal-processing (DSP) has been a widely available approach for analyzing sounds created by the human body. Although advanced medical imaging techniques can obtain more detailed information, they have some key limitations. The major limitation is the high cost required with the need to have specially trained medical professionals. Another major limitation is the required complex devices. Such devices are usually available in large medical facilities at major urban centers.

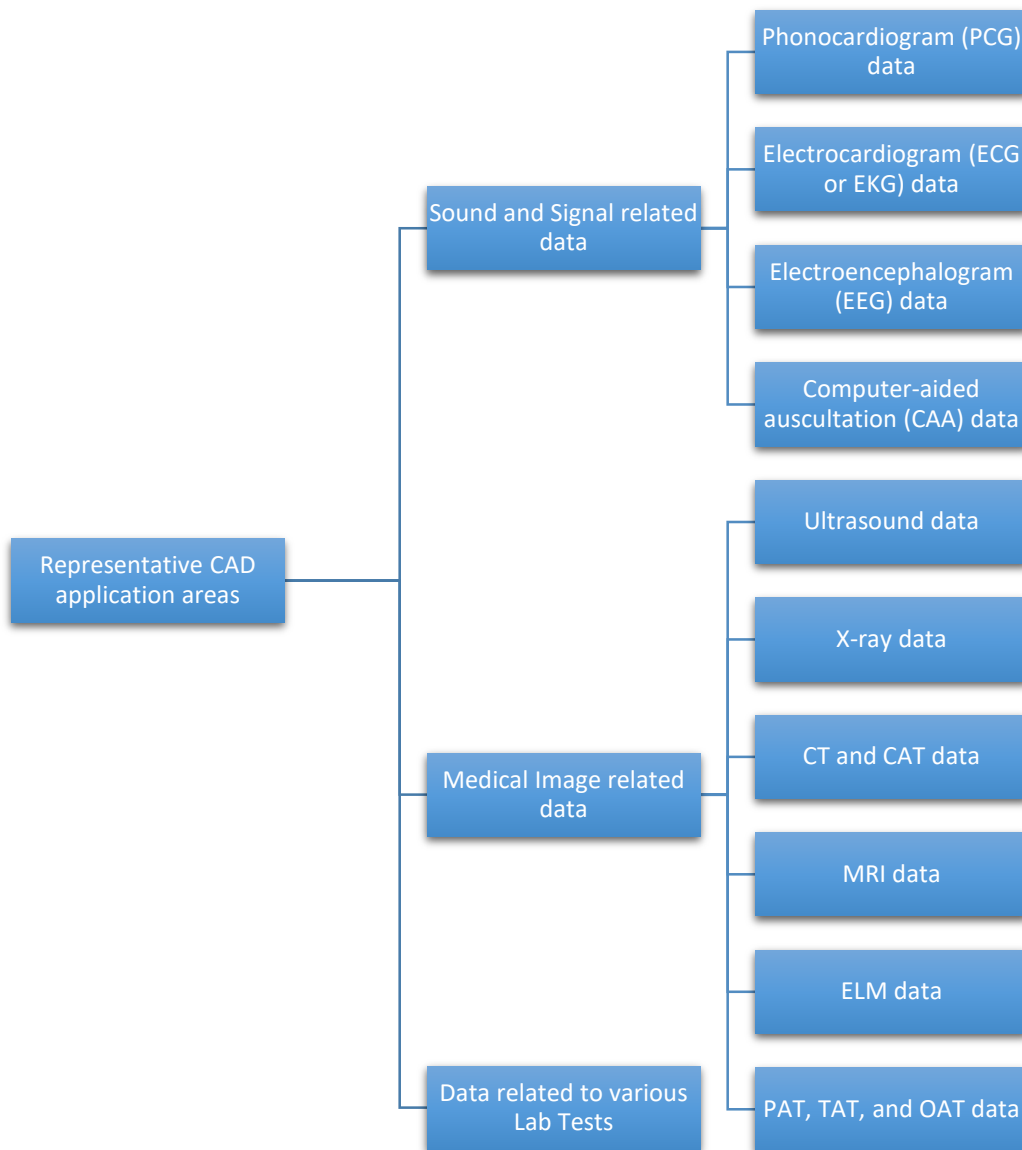


Figure 13: A Conceptual Organization of CAD Application Areas based on the Used Types of Data Sources.

On the other hand, CAD that uses sound and/or signal data has a great potential to expand the application domain. Devices for the acquisition of sounds and/or signals are usually portable. Therefore, they can be of great value in emergency medicine. Table 3 summarizes the major sound / signal acquisition methods used today by CAD. These methods are described in the following sub-sections.

Table 3: Date Acquisition Methods for Sounds and Signals.

Data Acquisition Method	Representation
Phonocardiogram (PCG)	Records of the sounds produced by the heart.
Electrocardiogram (ECG or EKG)	Records of the electrical activity of the heart.
Electroencephalogram (EEG)	Records of the electrical activity in the brain.
Computer-aided auscultation (CAA)	Recordings of heart or lung sounds obtained by using an electronic stethoscope.

6.1.1 Phonocardiogram (PCG) and Electrocardiogram (ECG or EKG) Analysis

Heart disease has consistently been the leading cause of serious sickness and death all over the world. Many heart problems can be diagnosed by analyzing heart sounds besides other types of data. These sounds are produced by heart beats and the circulation of the blood in the body. Thus, analyzing heart related sounds is a key task for detecting early stages of heart disease [Grusky, et al., 2007].

Heart sounds indicate conditions of blood flow. Normal adult heart sounds mainly consist of two periods, which are the systole and diastole, and four types of sounds. The first sound (S1) is a low frequency sound and it is caused by turbulence when the mitral and tricuspid valves close. S1 occurs at the beginning of the systolic period.

The second sound (S2) is a sound of lower frequency than that of S1. It is caused by the closure of the aortic and pulmonic valves which occur at the end of the systolic period. The S1 and S2 sounds occur in sequence with each heartbeat. Occasionally, there may be a third sound (S3). The S3 sound is of lower frequency than that of the S2 sound. The S3 sound occurs at the beginning of the diastolic period and after the S2 sound. The S3 sound is believed to be caused by the vibration of the valves, supporting structures, and the ventricular walls. Usually young people produce S3 sounds. If an S3 sound reappears later in life, it may be a signal of heart problems, such as congestive heart failure. If a fourth sound (S4) is detected, it can be a sign of heart problems [Reed, et al., 2004]. An abnormal S4 sound occurs later in the heart cycle. The S4 sound is caused by conditions that increase resistance to ventricular filling, such as a weak left ventricle. Other sounds such as murmurs, clicks, rubs and snaps can be abnormal sounds too [Reed, et al., 2004].

Phonocardiography is one of the techniques used for cardio diagnosis. It is used to record heart sounds and represent them on paper for later analysis. According to the Miller-Keane Encyclopedia and the Dictionary of Medicine, phonocardiography is the graphic recording of heart sounds and murmurs. This technique is used when it is difficult to distinguish abnormal heart sounds or murmurs by using a stethoscope. It can detect sonic vibrations from the heart via a highly sensitive microphone. Next, the vibrations are converted into electrical signals which are fed into a galvanometer, where they are recorded on paper or shown on a screen.

A PCG is a graphic record that is obtained from phonocardiography. Figure 14 shows some representative patterns of heart sounds produced by a normal (at the top of Figure 14) and several abnormal heart conditions. A disadvantage of PCG signals is that they can be too sensitive to noise which can become an obstacle. Noise can be created by various internal sources such as respiration and laughing and by external conditions such as other people speaking or having a door closing. Therefore, excluding noise from PCG signals is very critical for the accurate analysis of heart sounds [Nunes, et al., 2015].

Electrocardiography is another diagnostic procedure for examining heart activity electronically. Usually around 10 electrical wires (electrodes) are attached on select locations of the skin of a human subject and key electrical activity related to the state of the heart is recorded. The result is the formation of electrocardiograms (ECG or EKG), which can be displayed as diagrams on paper or a computer monitor.

Figure 15 shows typical PCG and ECG diagrams. Amplitude peaks in an ECG diagram are termed as the R and T wave peaks. As Figure 15 shows the R wave peaks of an ECG correspond to the S1 sounds of the PCG. Furthermore, the T wave peaks in an ECG are completed right before a period of diastole starts. An ECG is usually the first step in examining heartbeats. An ECG measures whether the pattern of heartbeats is normal or irregular, which may be related to some heart disease.

Lately, research using PCG diagrams has slowly declined while research using ECG is being recognized more. However, PCG research can provide significant information regarding the condition of heart valves and the hemodynamics of the heart, which can lead to more accurate diagnoses of various valve and heart diseases [Gupta, et al., 2007], [Uğuz, 2012]. Research based on the analysis of both ECG and PCG signals may reach highly accurate results in the diagnosis of heart disease [Safara, et al., 2013].

In order to analyze ECG and PCG signals, such signals first need to be normalized and after that to be segmented. Algorithms that accomplish these tasks include digital envelope detection. That is, they can detect the lower and upper peaks of the waves. Several kinds of classification techniques can be employed to classify such data [Gupta, et al., 2007], [Abbas and Bassam, 2009], [Safara, et al., 2013].

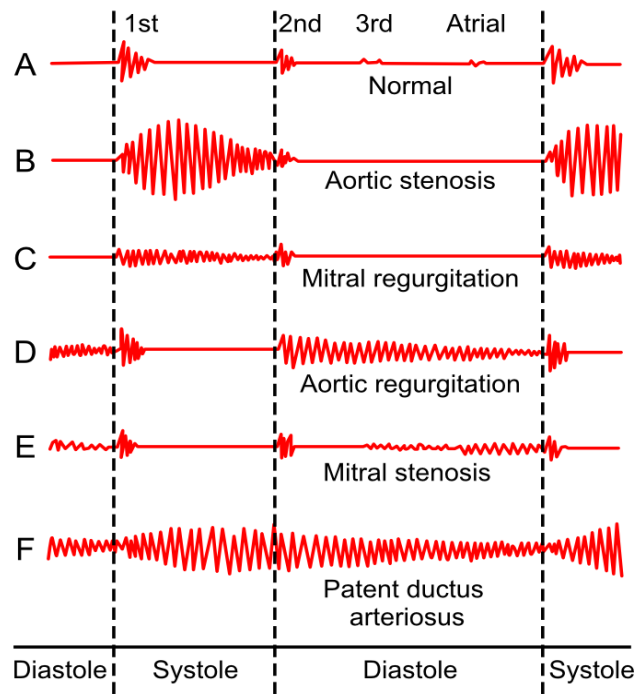


Figure 14: A PCG with Normal and Abnormal Heart Sounds [Madhero, 2010].

6.1.2 Electroencephalogram (EEG) Analysis

CAD has also been applied to detect brain disorders / abnormalities by using electroencephalogram (EEG) signals. EEG signals are derived by performing an electrophysiological test that analyzes electrical activity in the brain. Brain cells, called neurons, communicate with each other through electrochemical processes. A computer, by means of electrodes attached to the scalp, can detect these processes and the resulting signals appear as wavy lines on an EEG. EEG signals have been used to detect brain conditions, especially epilepsy and other seizure disorders. In addition, EEG signals are a reliable means for detecting other brain conditions too, such as memory problems, brain tumors, causes of sleep disorders, strokes, dementia and many other conditions of the cerebral structure. Several algorithms have been developed recently to identify and analyze EEG signals for detecting such brain conditions [König, et al., 2005], [Nunez and Srinivasan, 2006], [Ang, et al., 2010].

More recent studies which also use EEG signals have made contributions in topic areas related to mental disorders which are inherently difficult to quantify objectively. These include the auditory processing disorder (APD) [Ludlow, et al., 2014], [Snyder, et al., 2015] and the attention deficit hyperactivity disorder (ADD or ADHD) [Ahmadlou and Adeli, 2010], [Lenartowicz and Loo, 2014], [Snyder, et al., 2015]. In summary, using EEG data via computerized approaches as part of CAD offers a very promising prospect to accurately diagnose a wide spectrum of brain and mental disorders.

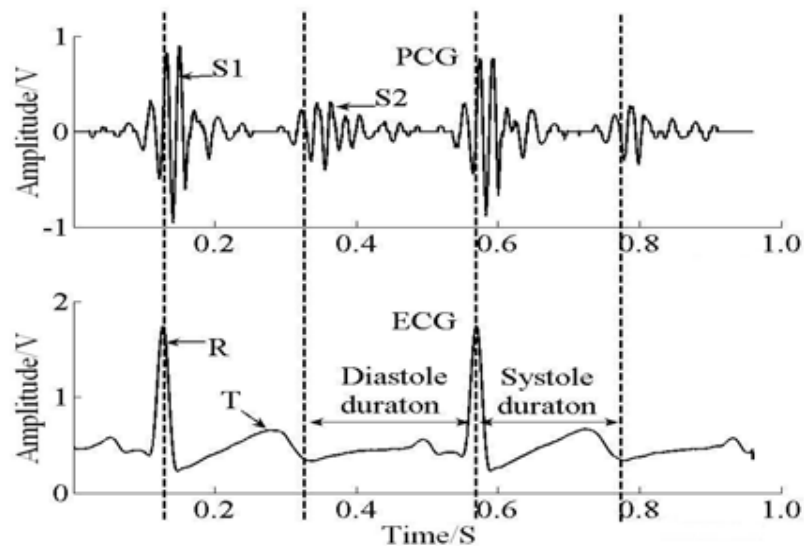


Figure 15: Typical PCG and ECG Diagrams [Han, et al., 2008].

6.1.3 Computer-Aided Auscultation (CAA)

Auscultation of the chest has been routinely an important clinical tool and is still the most common method for evaluating lung function [Fernandez-Granero, 2015]. Using a stethoscope to listen breathing is a routine procedure of clinical examination [Reichert, et al., 2008], [Bohadana, et al., 2014]. Although lung diseases are some of the most common medical conditions, people may underestimate their seriousness. Infections, air pollution (e.g. smoking and indoor / outdoor air pollution) and genetics are considered to be responsible for most lung diseases. Hundreds of millions of people suffer from lung diseases, which kill four millions of people every year worldwide [Ferkol and Schraufnagel, 2014].

Lung sounds can be categorized into two general types. These are normal sounds and abnormal sounds including wheezes, coarse crackles, fine crackles, and rhonchi [Mikami, et al., 1987]. The abnormal sounds can be signs of lung diseases, such as pneumonia, emphysema, chronic obstructive pulmonary disease (COPD), and asthma. The traditional chest examination by a stethoscope is, however, not sufficiently effective to accurately diagnose lung diseases. There are often inconsistent and subjective issues involved which may inhibit the accurate diagnosis and interpretation of lung sounds [Murphy, et al., 2004]. There are several exciting studies using CAA to collect lung sounds and analyze the data using computer algorithms for the diagnosis of lung disease or other respiratory disorders [İçer and Gengeç, 2014]. Computerized lung sound analysis (CLSA) can help to detect abnormal lung sounds related to some respiratory disorders [Gurung, et al., 2011]. Mastering traditional auscultation is difficult. Since nurses and paramedics can be easily trained to use CAA systems, CAD systems which use CAA can be applied to telemedicine [Weinstein, et al., 2014] and thus they can open totally new possibilities in this area.

6.2 Medical Image Analysis

Detecting asymptomatic diseases at their early stages is very difficult. Screening by using medical image analysis may be the only practical means to detect such conditions. In addition, medical images can contain significantly more valuable information needed to identify abnormalities in the human body, than any other form of examination.

A simple Google search using the key phrase “Computer-aided Diagnosis AND medicine” returned 283,000 hits as of April 27, 2019. Searching by using the two key phrases “Computer-aided Diagnosis AND medicine AND images” returned 232,000 hits. These results indicate the significance that imaging plays in CAD.

CAD systems which use medical imaging have been one of the most active and prominent study areas in recent years [Sotiras, et al., 2013]. The computerized module for detecting abnormalities in marked regions of medical images is extremely important in CAD systems. Often that module can be viewed as an independent computer-aided detection system. There are several diagnostic tools available for medical imaging analysis. Table 4 used as sources webpages of the Mayo Clinic, WebMD along with results in [Suetens, 2017] and [Dhawan, 2011] and it summarizes the major diagnostic methods used in applications of CAD in this area.

Table 4: Diagnostic Test Methods for Medical Imaging.

Diagnostic Test Methods	Means Used to Derive the Image	Remarks
Ultrasound (i.e., Echocardiogram)	High-frequency sound waves to produce images of structures in the human body.	<ul style="list-style-type: none"> • No risk of ionizing radiation exposure. • Less expensive than other medical imaging methods.
X-rays (i.e., chest x-rays, mammography and fluoroscopy, etc.)	Electromagnetic waves.	<ul style="list-style-type: none"> • Used for many screenings. • Risk of ionizing radiation exposure.
Computed Tomography (CT), or Computerized Axial Tomography (CAT)	X-rays and a computer to create cross-sectional images of the body.	<ul style="list-style-type: none"> • Used to identify various kinds of cancer and diseases that regular X-ray imaging cannot detect. • Higher risk of ionizing radiation exposure.
Magnetic Resonance Imaging (MRI)	Magnetic field and radio waves and a computer to produce detailed cross-sectional images of the body.	<ul style="list-style-type: none"> • No risk of ionizing radiation exposure. • Expensive, thus there is limited accessibility.
Dermatoscopy, or Epiluminescence microscopy (ELM)	Uses oil immersion techniques to create images for the detection and analysis of skin lesions.	<ul style="list-style-type: none"> • Used for the diagnosis of melanoma.
Photoacoustic tomography (PAT), Optoacoustic tomography (OAT), or Thermoacoustic tomography (TAT)	A combination of optical and ultrasound techniques. Laser pulses are delivered into tissues to acquire biomedical images.	<ul style="list-style-type: none"> • No risk of ionizing radiation exposure. • It is currently expensive.

6.2.1 Ultrasound Analysis

Analyzing images generated with ultrasound methods has several advantages. First, it has no risk of the ionizing radiation exposure that occurs with, for instance, X-ray imaging and thus it causes no harmful effects to the human body. Second, when it is compared to other medical imaging approaches, it is more cost effective. Ultrasound imaging systems can be portable and can capture still images and also motion (video) in real-time. In addition, ultrasounds can monitor the movement of blood as it flows in the vessels. Ultrasound devices can be directly transported to a trauma patient’s bedside to quickly examine the patient’s situation. Thus, ultrasound methods can be critically useful in emergency medicine. In particular, several studies have been conducted for developing CAD systems that use ultrasounds with an approach called Focused Assessment with Sonography for Trauma (FAST) [Matsushima and Frankel, 2011], [Marsousi, et al., 2014].

When ultrasounds are compared to X-rays, ultrasounds can result in clearer images of soft tissues and internal organs and parts of the body. Such organs can be the organs and the main veins in the abdomen.

Echocardiograms are images of the heart created by using ultrasound waves. They are also known as sonograms of the heart. They can be used to determine the size and shape of the heart and visualize the blood

flowing through blood vessels. Information derived by combining ECG diagrams along with echocardiograms can be used for the diagnosis of cardiomyopathy, or heart muscle disease [Chiarugi, et al., 2010], [Di Bella, et al., 2015].

However, ultrasound imaging can have some limitations. Ultrasound waves cannot penetrate dense objects (i.e., bones or teeth). For such cases, other imaging techniques, such as X-ray, CT and MRI techniques can be used. In addition, ultrasound waves can be distorted by air or gas pockets. As before, for such cases X-ray, CT, and MRI based techniques are more appropriate, unless new research can develop computerized approaches that can alleviate the previous distortion problem.

6.2.2 X-Ray Analysis

X-ray imaging is the most frequently used modality of medical imaging. X-rays can be used to examine many parts of the body, such as bones, teeth, lungs, heart and the abdomen. X-rays are effective for the analysis of hard tissues or structures (i.e., bones, teeth) and also for the analysis of air or gas-filled organs (i.e., the lungs). X-rays use ionizing radiation to produce images of internal structures of the body. Although screening mammography involves ionizing radiation exposure, the risk is very low compared to the benefits [Yaffe and Mainprize, 2011].

A mammogram, which is an X-ray image of the breast, has been the most common screening tool for the early detection of breast cancer [Ayer, et al., 2010]. A mammogram can capture the entire breast at once while an ultrasound based image captures only a localized area of the body. The concept of using a computer to detect abnormalities in a mammogram first appeared in 1967 [Winsberg, et al., 1967]. The earliest study of having a CAD system, which uses digital mammography was developed in 1993 [Ayer, et al., 2010]. Since then, using digital mammography for breast cancer screening and diagnosis has been one of the most established research areas of CAD in radiology. New research may lead to ways for minimize exposure to radiation without diminishing the diagnostic capability.

6.2.3 Computed Tomography (CT) Scan or Computerized Axial Tomography (CAT)

A computed (also known as computerized) tomography (or CT) scan is a special form of imaging which uses multiple X-ray images taken from different angles and then processing them by a computer to create 3D images [Kak and Slaney, 2002]. CT scans can generate detailed and more clear images of internal organs, the head, chest, sinuses, spine, bones, soft tissues and blood vessels when compared to images generated with traditional x-ray approaches.

During the last few decades, the number of CT scan practices has noticeably increased in the U.S. [Hall and Brenner, 2008]. Advantages of CT scanning, like X-ray imaging, is the fast, painless and non-invasive approach used. It is often considered as the best method for detecting various kinds of cancer since CT images can help physicians to spot abnormalities or tumors and precisely define their size and location [Kak and Slaney, 2002]. However, this comes with concerns stemming from the higher dose of radiation used in CT scans when it is compared to traditional X-ray approaches [Nickoloff and Alderson, 2001]. As with X-ray analysis, future research in this area may lead to ways for minimizing radiation exposure without diminishing the diagnostic capability.

6.2.4 Magnetic Resonance Imaging (MRI)

MRI is effective to use when there are many potential problems in different parts of the body. It can be used to examine parts of the human body such as the brain, breasts, heart, blood vessels, abdomen, organs in the pelvic region, or other organs. MRI is the best technique for evaluating most soft-tissue masses [Subhawong, et al., 2010]. Malignant tissue has a different set of magnetic properties than surrounding normal tissue. Therefore, MRI is able to capture these differences between malignant and benign tissue [Prostate Knowledge, 2009]. Until the early 2000s, however, there was strong skepticism about the reliability of MRI as a diagnostic

tool. Accurate detection of abnormalities depended largely on the facility and the skills and experience of the radiologists [Prostate Knowledge, 2009], [Rifkin, et al., 1990], [Engelbrecht, et al., 2001].

MRI technology has significantly improved in the past two decades. The latest MRI devices generate remarkably clear images of the internal body. There are several studies comparing breast MRI, mammography, and ultrasound imaging. These studies indicate that MRI is much more effective than mammograms and ultrasound images [Kuhl, et al., 2005], [Morrow, et al., 2011], [Fallenberg, et al., 2014]. MRI is especially effective for the diagnosis of neurological disorders and degenerative disorders (e.g., multiple sclerosis, Parkinson's disease and Alzheimer's disease).

However, MRI can be expensive. This is due to the use of rather complex devices and the need to have personnel with specialized skills to operate such devices. Furthermore, MRI is not recommended if there are implanted devices (such as a heart pacemaker) or when the patient is an infant (because it is difficult for them to stay still). Although such devices are not easily accessible or applicable, various studies have been done to develop CAD systems with MRI capabilities. Some application areas are in the diagnosis of breast cancer, dementia, and prostate cancer [Bhooshan, et al., 2014], [Bron, et al., 2015], [Litjens, et al., 2015].

6.2.5 Dermatoscopy or Epiluminescence Microscopy (ELM)

The prevention of melanoma and non-melanoma skin cancer (NMSC) is rather simple. However, it is often ignored. Now NMSC is the most common cancer in the U.S. [Guy, et al., 2015]. Dermatoscopy can be used for the diagnosis of skin lesions. Very often malignant skin lesions (e.g., melanoma) are difficult to be identified with the naked eye. There are several studies of CAD which uses dermatoscopy interpretation [Hoffmann, et al., 2003], [Schmid-Saugeona, et al., 2003], [Abbas, et al., 2012].

Interpretation by digital dermatoscopy has consistently contributed to the diagnosis of melanoma. In recent years the application domain of dermatoscopy has been expanded. New application areas include inflammatory diseases, parasitoses, hair and nail abnormalities, and neoplastic and other skin disorders [Micali and Lacarrubba, 2016]. Dermatoscopy is painless, non-invasive and easy to be performed. Dermatoscopy has the potential to further advance the development of CAD for the diagnosis of skin related problems including many types of cancer.

6.2.6 Photoacoustic Tomography (PAT), Optoacoustic Tomography or Thermoacoustic Tomography

PAT has been rapidly developed in the past few decades and became an established imaging modality in clinical practice [Xia, 2014]. PAT employs a combination of optical and ultrasound techniques. Thus, it has the same advantage that ultrasound has. First, it imposes no risk of ionizing radiation exposure. PAT captures more clear images of soft tissues. PAT can be used to detect skin cancer, brain cancer, ovarian cancer, prostate cancer and other types of cancer [Wang, 2014]. Some of the limitations of PAT are the same as those for ultrasound. PAT waves can be distorted by air or gas-filled organs. Thus, they cannot effectively penetrate gas-filled cavities or the lungs. Research in this area may have the same goals as research on ultrasound analysis.

6.3 Analysis of Lab Tests

The results of common lab tests along with physiologic measurements, such as heartbeat rate, respiratory rate and blood pressure, can be very important input data for CAD. Typical lab tests examine samples of blood, urine, or body tissues. Usually, results from such tests provide simple clarifications to basic questions such as whether key indicators fall within normal range or whether a specific substance is present or not.

Some lab tests can be quite reliable and provide physicians critical information for diagnosis. Let us consider for demonstrative purpose the case of sepsis. Sepsis is a severe and often a mortal complication which occurs after an infection takes place [Martin, et al., 2003]. The diagnosis of sepsis is difficult since symptoms of sepsis have similarities with those of other conditions from non-infection related diseases [Tang, et al., 2007]. Moreover, there is not a single test which can determine that a patient has sepsis. However, results from a group of simple blood tests may reveal the possibility of having sepsis [Drewry, et al., 2015].

The combination of various lab tests along with image data can provide a more accurate diagnosis for sepsis. Conditions of sepsis patients at an ICU (intensive care unit) may often change rapidly. For such cases lab tests may provide a more effective approach for monitoring the status of the patient when compared to imaging approaches alone [Miller, et al., 2015]. Thus, having a CAD system that considers lab test results and imaging results comprehensively provides a highly accurate and timely approach for diagnosing sepsis [Miller, et al., 2015]. Future research in this area may focus on finding new lab tests that could be incorporated into the CAD process.

7. The Workflow of CAD

A typical CAD system, like most traditional data mining systems [Triantaphyllou, 2010], usually follows a five-step (or stage) approach during its development phase as indicated in Figure 16. These steps are ordered as follows: Data acquisition, data pre-processing, processing, visualization / presentation and the validation / assessment of the final results. There is also the possibility of moving backwards (not shown in Figure 16 for simplicity of the presentation) of this sequence, if feedback generated at any stage of this process warrants to do so. Feedback may take place among tasks within a stage or among different stages. The less human interaction takes place with these steps, the more automatic the CAD system is. The first four steps shown in Figure 16 are followed in this order when a CAD system is used in a clinical setting (i.e., after it has been developed). These steps are analyzed more in the following sub-sections.

We conjecture that future CAD systems will use a workflow architecture which will be based on feedback significantly more extensively. Thus, they will have stronger learning capabilities by analyzing data generated during their operation. The emphasis on learning from past experience, via a feedback mechanism, is the foundation for having true intelligent CAD systems. Thus, we call this type of next generation CAD systems the *intelligent CAD* systems.

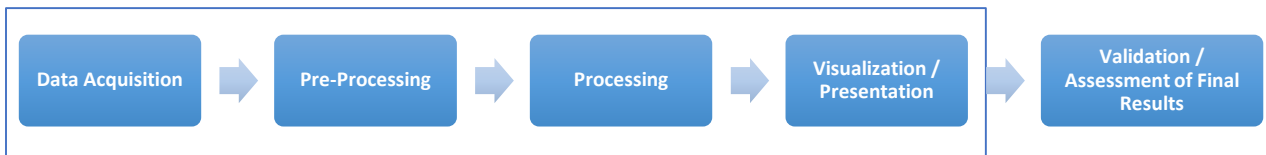


Figure 16: The Workflow Structure of a Typical CAD system.

7.1 Data Acquisition

At this step the CAD system acquires all or most of the data that may be required for the diagnosis. These data may come from the sources described earlier in Table 2. Such data may be acquired all together at once or at a sequence of stages as the diagnostic process proceeds by starting with a simple diagnostic hypothesis and then moving to more complex hypotheses [Trapeznikov and Saligrama, 2013], [Thung, et al., 2017]. Usually, such additional data require progressively higher costs to be acquired. The data at a given stage depend on the data of previous stages and the targeted diagnostic hypothesis.

7.2 Pre-Processing

During the pre-processing step, any raw data (such as lung sounds, heart signals, or medical images) need to be normalized. This means that any missing data cases need to be addressed. Also any outliers need to be identified and any noise in the raw data needs to be reduced or be eliminated. This can be done by means of filtering algorithms [El-Dahshan, et al., 2014], [Schindera, et al., 2013].

Areas in medical images that are not of particular interest are trimmed off to remove unwanted distortions, and enhance some of the image features essential for the next steps [Way, et al., 2006]. Some fundamental image pre-processing methods are brightness transformations, geometric transformations, local

neighborhood pre-processing, and image restoration [Sonka, et al., 2014]. During this step possible data outliers are identified and removed [Aggarwal, 2015], [Hodge and Austin, 2004].

Medical data are notorious for having records with missing values. This is a rather frequent problem with any data processing approach. Methods for dealing with missing data are discussed in [Collins, et al., 2015], [Harrell, 2015], and [Seera and Lim, 2014].

When a disease or medical condition occurs very infrequently, then data can be severely imbalanced. That is, records that represent cases with the disease or condition of interest are a small number when compared to the rest of the records. Algorithms that deal with such cases are described in [Maldonado, et al., 2014], [Razzaghi, et al., 2015], [López, et al., 2013], and [Şeref, et al., 2017].

7.3 Processing

Processing involves three sub-steps. These are segmentation, feature extraction / selection and classification (see Figure 17). These sub-steps are discussed next.

7.3.1 Segmentation

Segmentation is applicable when we have to process signal data or images. Segmentation is the process of dividing the data into several subsets (i.e., segments) based on some conditions. This is motivated by the fact that successfully partitioned data may be easier for analysis.



Figure 17: The Three Sub-Steps of the Processing Step.

Segmentation is considered as the most important task in computer vision [Jain, 1989] and one of the most recognized research areas in medical image analysis [Sonka, et al., 2014], [van Ginneken, et al., 2011], [Yang, et al., 2017]. Having a good segmentation approach may lead to a more objective diagnostic performance. If segmentation is inadequate, then a CAD system may misdiagnose, for instance, a malignant tumor as benign or vice versa. Many types of rules, methods and algorithms have been developed to handle segmentation. Typically, segmentation in medical imaging starts with the identification of objects (areas) of interest. The next task is to divide the image based on some criteria and to extract objects from the image's background by using some algorithms. Such algorithms may use thresholding [Armato, et al., 1999], [Way, et al., 2006], region growing methods [Brown, et al., 2001], [Way, et al., 2006], [Croisille, et al., 1995], level set methods [Malladi, et al., 1995], [Way, et al., 2006], [Lee, et al., 2001], watershed [Ng, et al., 2006], statistical region merging [Celebi, et al., 2008] or lesion methods [Yuan, et al., 2007].

Multidimensional images have become more common in clinical practice and are important for medical diagnosis [Bühler, et al., 2005]. One of the biggest challenges in segmentation is the fact that while methods and devices for developing multidimensional images have improved, advanced segmentation algorithms for multidimensional images have not been developed accordingly [Bühler, et al., 2005], [Kumar, et al., 2013], [Kumar, et al., 2016]. Recent developments in this area focus on two main trends regarding segmentation approaches for image analysis: (i) *Semantic segmentation* and (ii) *instance segmentation* [Girshick, et al., 2014], [Long, et al., 2015].

Semantic segmentation labels each pixel with an object-class (for instance, this pixel belongs to an organ, to a lesion, or to a lesion within an organ, etc.) or to a non-object class (e.g., it is on the background). But it does

not assign a pixel to a specific organ or lesion etc. On the other hand, instance segmentation is more complex. Now pixels are labeled separately to each specific object (for instance, this pixel is part of organ #1, or lesion #3, etc.). That is, instance segmentation aims at solving both object detection and semantic segmentation [Romera-Paredes and Torr, 2016], [Goodfellow, et al., 2016].

An emerging trend in this area is the use of *deep learning* approaches. Deep learning is a subset of machine learning and it is based on densely connected artificial neural networks [Goodfellow, et al., 2016], [Charniak, 2018]. Such networks use multiple layers and pathways of nodes for the data to spread [LeCun, et al., 2015]. Convolutional Neural Networks (CNN) is one of the basic approaches in deep learning. There are numerous variants of CNN as depicted in Table 5. By using deep learning techniques, some recent studies show promising results for semantic segmentation and instance segmentation for 2D and 3D images [Christ, et al., 2016], [Shin, et al., 2016], [Kamnitsas, et al., 2017], [He, et al., 2017] (see also Table 5).

Table 5: Some Deep Learning Approaches Regarding Image Segmentation.

Deep Learning Approach	Some Related Studies
Convolutional Neural Networks (CNN)	[Moeskops, et al., 2016], [Christ, et al., 2016], [Kamnitsas, et al., 2017], [Shin, et al., 2016]
Deep Convolutional Neural Networks (DCNN)	[Li, et al., 2015]
Fully Convolutional Networks (FCN)	[Long et al., 2015], [Milletari, et al., 2016], [Dolz, et al., 2018], [Christ, et al., 2016]
Region Based Convolutional Neural Networks (R-CNN)	[Girshick, et al., 2014]
Faster R-CNN	[Ren, et al., 2017]
Mask R-CNN	[He, et al., 2017]

In the case of using signal data, currently there are two basic approaches to segmentation. These two basic segmentation approaches for signal analysis are: (i) *Constant segmentation* and (ii) *adaptive segmentation*. The constant segmentation approach divides a signal into segments of a constant length. This is the simplest form of segmentation but it is not as effective. Most signals from the human body are non-stationary, which means that their statistical properties (e.g., mean value, dispersion or frequency) change over time. In the adaptive segmentation approach the signal boundaries, such as amplitude, frequency and wavelength are detected automatically and are divided into segments which may be of different lengths that are relatively stationary [Kirlangic, et al., 2001]. Currently, there are many algorithms for adaptive segmentation, such as the Matching Pursuit algorithm [Nieblas, et al., 2013], hidden Markov models [Schmidt, et al., 2010], and the tunable-Q wavelet transform [Patidar and Pachori, 2013].

7.3.2 Feature Extraction / Selection

As it was discussed in Sub-Section 5.1 (“Managing Large Volumes of Clinical Data”) data for medical diagnosis may come from a wide spectrum of sources. For instance, some data may come from personal or family history. Other data may come from demographics. Other sources for data might be pathology tests, lab results, imaging data (such as X-ray images or MR images) or signal data (such as ECG or EEG data). Data may even come from the analysis of patient testimonies in the form of natural language transcripts. In order for such data to become suitable for computer analysis they need to be processed and transformed into a form that can be analyzed by diagnostic algorithms.

The main task at this phase is to extract the pertinent features or attributes. Feature extraction transforms the raw data into more useful and non-redundant information which can be easily processed and interpreted [Kohavi and John, 1997], [Tang, et al., 2014]. For instance, when an image is considered (such as in the case of analyzing a tumor), an algorithm should take as input the raw data of the image and first extract the perimeter of the tumor. After the perimeter is extracted, another algorithm may identify the geometric features of the tumor. Such features may provide information on how many lobulations exist, how deep or wide they are

and so on (for instance, see [Kovalerchuk, et al., 1996] or [Triantaphyllou, 2010] for an example of such algorithm). For the case of ECG analysis one may define the signal segments described in Sub-Section 6.1.1 (“Phonocardiogram (PCG) and Electrocardiogram (ECG or EKG) Analysis”).

As result of the diversity of the data sources that lead to the feature extraction, features may be described in terms of deterministic or stochastic (probabilistic) data. Features may be crisp or even fuzzy. Crisp features may be continuous (such as body temperature readings) or discrete. Other ways for categorizing features is to group them as ordinal (when one may order them) or non-ordinal. Binary or non-binary, linear or hierarchical or combinations of the previous types.

Naturally, one wishes to capture all needed aspects to reach the final diagnosis, especially if the diagnosis involves an exploration among many diagnostic possibilities. This may lead to a very large number of features. From the algorithmic point of view, this may impact the storage or processing complexity of the algorithms as many of them are very sensitive to the dimensionality of the data. This is known as the “*dimensionality curse*.” The dimensionality is usually described in terms of the number of data instances and the number of features involved. Therefore, a crucial issue is how to select the most appropriate features such that the accuracy of the final diagnosis is not compromised and at the same time, the amount of data that need to be processed is not too high. Thus, the next step, known as *feature selection* is a very important step as well. Another relevant issue is how to merge / combine simple features to create fewer features which possess higher diagnostic power [Nixon and Aguado, 2012].

The process of selecting only relevant information, and removing irrelevant or low significance information from the original input data, is called feature selection [Kohavi and John, 1997], [Liu and Motoda, 2012], [Tang, et al., 2014]. There are many approaches to feature selection. The most well-known ones are three: filtering, wrapping and embedding methods [Liu, et al., 2010]. *Filtering* is the simplest approach. One examines the attributes one-at-a-time and tries to decide which one is more likely to possess the best chances to assist the most in the diagnostic process. Examples of filtering approaches are ones based on information gain, correlation coefficients, variance thresholding etc. [Yang and Pedersen, 1997]. For instance, when the variance is considered, each feature is analyzed in terms of the variance of the values it assumes in a dataset. If that variance is high enough, this is considered as a promising aspect and the feature is selected. One may proceed in a sequential manner by selecting the top K features (i.e., the features that have the K highest variances) or use a threshold value for the variance for a feature to be included. Such method, and for that matter most of the other methods, are heuristic in nature. This is a greedy approach and thus might not be optimal.

Similarly, there are alternative *wrapper methods* [Kohavi and John, 1997]. Such methods may be based on recursive feature elimination, sequential feature selection, or ones based on genetic algorithms. Such methods consider one feature at a time and examine what impact it has when one considers the classification / diagnostic task. They are sequential and also heuristic in nature.

The most well-known *embedded method* is the L1 regularization method based on the *LASSO approach*. The LASSO approach has its foundation on the seminal work by Tibshirani and it was introduced in [Tibshirani, 1996] when selecting the best features for a regression model. The main idea can be described as follows:

Suppose that given is a dataset (X^i, y_i) , for $i = 1, 2, 3, \dots, N$, where $X^i = (x_{i1}, x_{i2}, x_{i3}, \dots, x_{ip})$ are the predictor variables and y_i are the responses. Then, the ordinary linear regression model considers the following formula:

$$\beta_{i1}x_{i1} + \beta_{i2}x_{i2} + \beta_{i3}x_{i3} + \dots + \beta_{ip}x_{ip} + \beta_0 = \bar{y}_i, \text{ or: } \sum_{j=0}^{j=p} \beta_j x_{ij} = \bar{y}_i,$$

where \bar{y}_i is the predicted value which we would like to be as close to the real value y_i as possible.

The regression approach determines the β_j coefficients such that the following sum of squared errors is minimized:

$$\text{Minimize } \sum_{i=1}^{i=N} (y_i - \bar{y}_i)^2, \quad \text{or equivalently: Minimize } \sum_{i=1}^{i=N} (y_i - \sum_{j=0}^{j=p} \beta_j x_{ij})^2.$$

The above formulation is a standard quadratic unconstrained optimization model. The LASSO approach introduces a constraint which needs to be satisfied simultaneously with the previous minimization goal. The new model is as follows:

$$\text{Minimize } \sum_{i=1}^{i=N} (y_i - \sum_{j=0}^{j=p} \beta_j x_{ij})^2 \quad \text{subject to: } \sum_{j=0}^{j=p} |\beta_j| \leq t$$

The main idea here is to have a predictive model that satisfies two goals simultaneously which are contradicting each other. One is for the model to be highly accurate (which requires many features to be present) and the other is to be highly simple (which requires few features). The LASSO approach uses the previous constraint in the form of a penalty function in addition to the traditional goal to minimize the sum of the squared errors.

This extra constraint requires the sum of the absolute values of the attribute coefficients (in a regression model) to be less than or equal to a parameter, say t . If the value of t is small enough, some of the coefficients will be forced to be equal to zero in order for the constraint to be satisfied. The smaller the value of t is, the more coefficients will be forced to be equal to zero, and vice versa. The group of features which have been assigned zero coefficients in the regression model will be excluded as a sub-group and not one-at-a-time (i.e., they are captured by the “lasso” and are removed).

The value of the control parameter t can be determined as the smallest t value such that the accuracy of the resulting model is satisfactory. This can be determined by successively reducing the value of t until the appropriate model (in terms of the number of included attributes) is reached. The features that survive this process are the selected ones to be included in the model. Using feature extraction / selection can save storage space and computing time for the next step which is the classification process. This in turn, can lead to the creation of CAD systems capable of powerful performance.

7.3.3 Classification

During this step the extracted / selected features are used by the appropriate classification model. Depending on the aim of the CAD system, different classification techniques (i.e., classifiers) are applied. It is very important to identify the aim of the CAD. For example, it has to be clarified whether the CAD system is required to differentiate normal heart sounds from abnormal heart sounds or whether it is used to diagnose a specific heart disease [Leng, et al., 2015] and so on.

As it was mentioned in Sub-Section 4.2, there are two major types of machine learning approaches that can be used: supervised learning and unsupervised learning approaches. There are various supervised algorithms such as those based on the inference of decision trees [Podgorelec, et al., 2002], support vector machines [Suykens and Vandewalle, 1999], random forests [Breiman, 2001], [Liaw and Wiener, 2002], linear regression [Seber and Lee, 2012], neural networks [Hansen and Salamon, 1990], or K-nearest neighbor approaches [Korn, et al., 1998], [Chen, et al., 2015]. Tables 6 and 7 summarize some representative recent studies of supervised learning algorithms developed for medical signal analysis and medical image analysis, respectively.

Often time in medicine unsupervised learning is used to discover new knowledge after supervised learning was applied [Deo, 2015]. A survey on the use of machine learning approaches for medical diagnosis can be found in [Kononenko, 2001] or in [Erickson, et al., 2017].

The previous approaches are based on crisp logic. Such approaches can be deterministic and/or stochastic in nature. However, some aspects in medicine and healthcare, by nature, are based on subjective or fuzzy data. Thus, the best way to deal with such data is to use approaches based on some type of fuzzy logic.

The foundation of any fuzzy approach is the notion of a *fuzzy set*. It is Zadeh who in his seminal work [Zadeh, 1965] first introduced this concept as an extension of the traditional concept of crisp sets. A fuzzy set

has members too, but now each member, say denoted as x , is associated with a membership value $\mu(x)$, where $0 \leq \mu(x) \leq 1$. After some long period of time using this idea without a modification, some operations research / computer science researchers started exploring ways to extend it. Atanassov in [Atanassov, 1986; and 1999] introduced the concept of *intuitionistic fuzzy sets*. Unlike Zadeh’s original definition of a fuzzy set, now each member x of an intuitionistic fuzzy set is associated with the traditional membership value $\mu(x)$ and also a non-membership value denoted as $\nu(x)$. Both $\mu(x)$ and $\nu(x)$ values are ≥ 0 , but a key property is to have $\mu(x) + \nu(x) \leq 1$ (see also Figure 18).

The concept of intuitionistic fuzzy sets has been extended too. First Smarandache introduced the concept of *neutrosophic fuzzy sets* in [Smarandache, 1998]. A neutrosophic fuzzy set can be viewed as an intuitionistic fuzzy set with the addition of a value that expresses how much undecided is a person regarding the membership of an entity x in such set. This is expressed by a third variable, say $i(x)$, called the *indeterminate* value, which takes on values from the interval $[0, 1]$. The $i(x)$ concept is different than the non-membership $\nu(x)$ concept. Obviously, the smaller the other two values $\mu(x)$ and $\nu(x)$ values are, the bigger the value $i(x)$ can be. However, these three membership values do not have to add up to 1.

Table 6: Some Key Approaches of Supervised Learning Developed for Medical Signal Analysis.

Approach	Objective of the Approach	Some Related Studies
Convolutional Neural Networks (CNN)	Classification in ECG	[Acharya, et al., 2017], [Kiranyaz, et al., 2015]
Artificial Neural Networks (ANN)	Classification in ECG	[Rad, et al., 2017]
Support Vector Machines (SVM)	Classification in ECG	[Pławiak, 2018]
Universum Support Vector Machines (USVM)	Classification in EEG	[Richhariya and Tanveer, 2018]

Table 7: Some Key Approaches of Supervised Learning Developed for Medical Image Analysis.

Approach	Objective of the Approach	Some Related Studies
Convolutional Neural Networks (CNN)	Feature extraction	[Thomaz, et al., 2017]
	Chest pathology Feature selection in X-rays	[Bar, et al., 2018]
	Brain segmentation in MRI	[Moeskops, et al., 2016]
Deep Convolutional Neural Networks (DCNN)	Lung pattern classification	[Anthimopoulos, et al., 2016]
	2D Classification in CT	[Miki, et al., 2017]
	2D Segmentation in CT	[Li, et al., 2015]
Fully Convolutional Networks (FCN)	Segmentation	[Long, et al., 2015], [Milletari, et al., 2016]
	3D Segmentation in MRI	[Dolz, et al., 2018]
	3D Segmentation in CT	[Christ, et al., 2016]

Recently, Yager introduced another extension of intuitionistic fuzzy sets known as the *Pythagorean fuzzy sets* [Yager, 2013; 2014; and 2016]. This name comes from the requirement that the two membership values of

a traditional intuitionistic fuzzy set need to satisfy the following requirement $\mu(x)^2 + v(x)^2 \leq 1$, which reminds the Pythagorean theorem. This is a more tolerant (flexible) requirement from the previous requirement $\mu(x) + v(x) \leq 1$ (see also Figure 18). A systematic comparison between intuitionistic fuzzy sets and Pythagorean fuzzy sets is given in [Peng and Yang, 2015], while a comprehensive review on Pythagorean fuzzy sets can be found in [Peng and Selvachandran, 2017].

The above concepts can be applied when a human enters (subjective) data or on image analysis for medical applications. A recent extensive review of neutrosophic fuzzy sets and related topics can be found in [Peng and Dai, 2018]. Some issues of traditional fuzzy sets applied to multi-criteria decision-making can be found in [Triantaphyllou and Lin, 1996] or [Triantaphyllou, 2000]. An application of traditional fuzzy logic on a medical problem is described in [Kovalerchuk, et al., 1997] or [Triantaphyllou, 2010]. There the authors propose a fuzzy logic based approach for characterizing breast masses in terms of their lobulation characteristics as part of a diagnostic procedure for breast cancer. The previous variants of fuzzy sets, may be used in fuzzy decision-making, fuzzy clustering, fuzzy image analysis, and fuzzy attribute extraction and selection, all of which are of vital significance in medical CAD [Das, et al., 2016], [Deng, et al., 2016], [Lin, et al., 2016], [Kuo, et al., 2018], and [Vlachos and Sergiadis, 2007]. A concern with fuzzy approaches is the rather complex mathematical issues that may be involved and how easily and reliably the pertinent data can be elicited when human judgment needs to be considered.

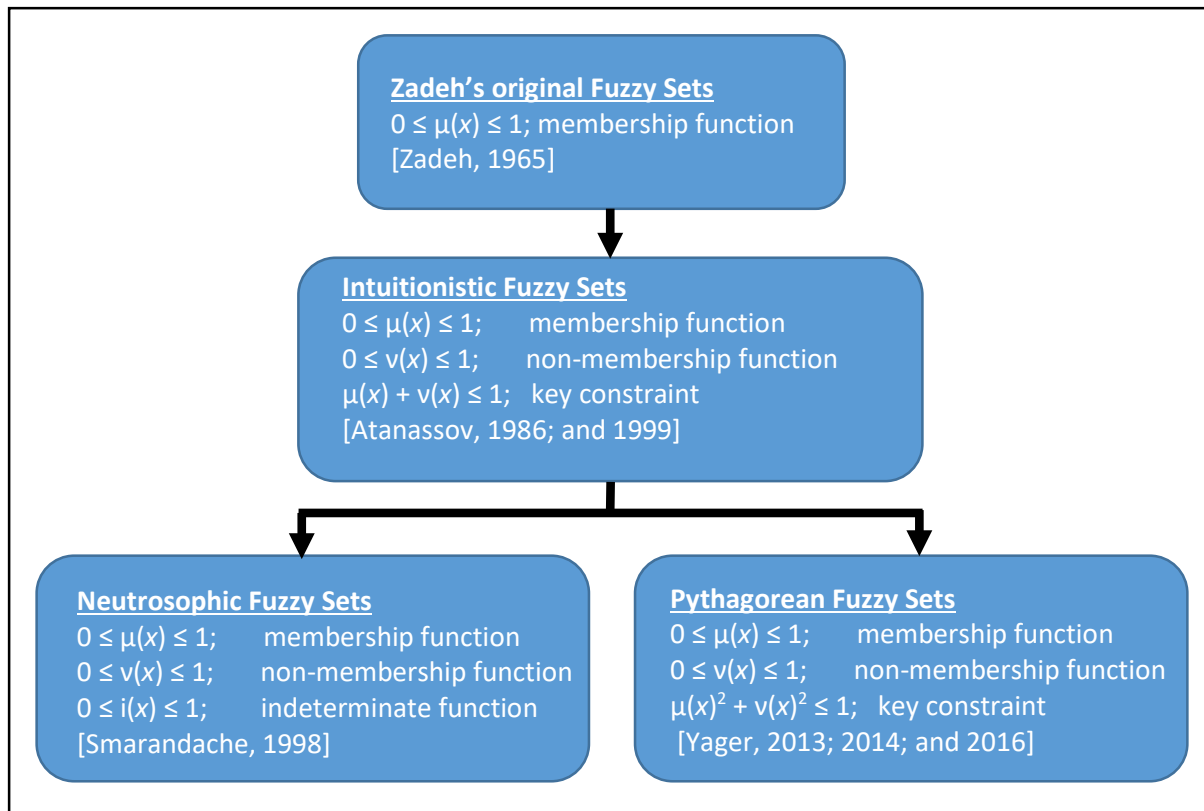


Figure 18: Some Key Developments in the Introduction of Fuzzy Set Concepts.

Recent studies on the emerging sub-field of deep learning approaches show some promising results related to CAD systems (i.e., in segmentation, feature extraction / selection, or classification). Related are some of the studies stated in Tables 5, 6, and 7. In general, regarding classification the fact is that some algorithms perform better than others depending on the application for not well-understood reasons. However, there is not a single best algorithm for all CAD systems. Therefore, more research needs to be done in this important

area including the aspects of supervised / unsupervised learning, crisp / fuzzy approaches and also to further explore the potential of deep learning approaches in CAD.

7.4 Visualization / Presentation

Visualization and the way results are presented is a significant function in CAD. Visualization can assist physicians to easily comprehend important information for diagnostic purposes. Using graphical representations and various types of visualization / presentation techniques, data and the results of classification are presented in ways that are more effective. In this manner physicians may be able to discover potentially new important aspects not recognized explicitly by the CAD system so far.

Visualization may involve creating new images (in 2D or 3D plus a time dimension) using the patient's data over time to understand the progress of the patient's disease [Bühler, et al., 2005]. Visualization may include image fusion. This is the process of creating a new image by combining data that are retrieved from different image modalities and then rebuilding high quality multidimensional images from, say, 2D images [Bühler, et al., 2005]. Visualization can provide an alternative view of the data or offer the opportunity to discover completely new and useful knowledge of the issues involved [Zhang, et al., 2006], [Ward, et al., 2010]. This stage offers a high opportunity to incorporate human feedback.

7.5 Validation / Assessment of the Final Results

This important step takes place during the development phase of a CAD system. It focusses on the validation and the performance assessment of the CAD system. It is the opportunity to provide critical feedback that could improve the entire CAD workflow and potentially go back and adjust some of the previous steps. This step is critical for creating truly intelligent systems (see also the introduction of Section 7). There are several methods which can be used to assess the diagnostic performance of a CAD system. An individual assessment approach may deal with only some of the aspects of a CAD system's performance. However, it may fail to consider some other important information. Therefore, one may need to consider more than one performance assessment approach in order to adequately assess the diagnostic performance of a CAD system. The following approaches are the major ways used today to evaluate CAD.

7.5.1 Receiver Operating Characteristic Curve (ROC), and the Area Under the Curve (AUC)

Having a CAD system and then providing medical professionals with diagnostic results, even if the results are frequently accurate, is not sufficient when one wishes to assess the overall effectiveness of the CAD system [van Ginneken, et al., 2011]. This is why a number of approaches for assessing the effectiveness of diagnostic systems (or classifiers) have been proposed in the literature. Some fundamental concepts need to be introduced first before these approaches are presented.

Of key significance here is the concept of the *confusion matrix* (see also Table 8). In the realization of the confusion matrix depicted in Table 8, we consider the case of checking whether a patient has a particular disease or not. The confusion matrix describes all four possible scenarios that exist when a classifier makes predictions and one compares them with the actual values. Thus, a given case has to be one of the following mutually exclusive and exhaustive cases: true positive, false positive, false negative, or true negative. These four concepts are next used to define the notions of *sensitivity* and *specificity* (see also Table 9). Sensitivity and specificity are the most commonly used statistical diagnostic measures [Griner, et al., 1981], [Summers, 2002].

Table 8: The Confusion Matrix [Lalkhen and McCluskey, 2008].

	Patient has the Disease	Patient does not have the Disease
Test is Positive	TP (true positive)	FP (false positive)
Test is Negative	FN (false negative)	TN (true negative)

Table 9: Key Statistical Measures Derived from the Confusion Matrix.

Explanation	Name	Calculation Formula
True Positive Rate (TPR)	Sensitivity (or Recall)	$TP / (TP + FN)$
True Negative Rate (TNR)	Specificity	$TN / (FP + TN)$
False Positive Rate (FPR)	(1- Specificity)	$FP / (FP + TN)$
Positive Predictive Value (PPV)	Precision	$TP / (TP + FP)$
Negative Predictive Value (NPV)		$TN / (TN + FN)$
Proportion of records that correspond to a disease	Prevalence	$(TP + FN) / (TP + FN + TN + FP)$

The *receiver operating characteristic (ROC)* curve is a graph which can be used to assess the diagnostic performance of CAD systems. The ROC curve displays the relation between the true positive rate (i.e., sensitivity) when it is plotted against the false positive rate (i.e., $1 - \text{specificity}$). When a CAD system has high specificity, the price we often pay is to have low sensitivity [D'Orsi, 2001]. The ultimate goal of a CAD system is to be able to produce both high sensitivity and also high specificity. In terms of Figure 19, this means that a highly effective CAD system has an ROC curve that lies closely to the upper left corner of the area. For instance, in Figure 19 a CAD system with ROC equal to the curve denoted as B is better than a CAD system with the ROC curve denoted as C. An ideal CAD system would correspond to the curve comprised by the vertical segment with coordinates (0, 0) and (1, 0) followed by the horizontal segment with coordinates (1, 0) and (1, 1). This is denoted as A at the upper left corner in Figure 19.

High sensitivity of detection with the smallest possible number of false positives (i.e., equal to 0) is ideal. However, there is usually a tradeoff between sensitivity and specificity. High sensitivity by using broader, more inclusive, classification criteria always occurs with an increase on the number of false positives [Summers, 2002].

The area under the curve (AUC) in the ROC space is another measure of test performance for a binary classifier. In terms of Figure 19 system A corresponds to an AUC with value equal to 1, system C to an AUC with value equal to 0.50, while system B corresponds to an AUC value equal somewhere between 1 and 0.50 (most likely around the value 0.85). An AUC value of 0.50 indicates that the system makes diagnostic decisions randomly with probability equal to 50% (for two-class systems). ROC curves along with AUC values can describe better how accurate a CAD system is.

Another issue here is the overfitting and overgeneralization aspects of a CAD system as these two complementary properties may affect the accuracy of the systems when they are fed with brand new data records of unknown class values. A system with too high fitting capability would classify the training data very accurately but at the same time it may be highly inaccurate when dealing with brand new data (and thus of unknown class value). On the other hand, a system with too high generalability may classify brand new data very inaccurately and training data inaccurately too. How to find an optimal balance between these two competing objectives is discussed in [Pham and Triantaphyllou, 2008a; and 2008b].

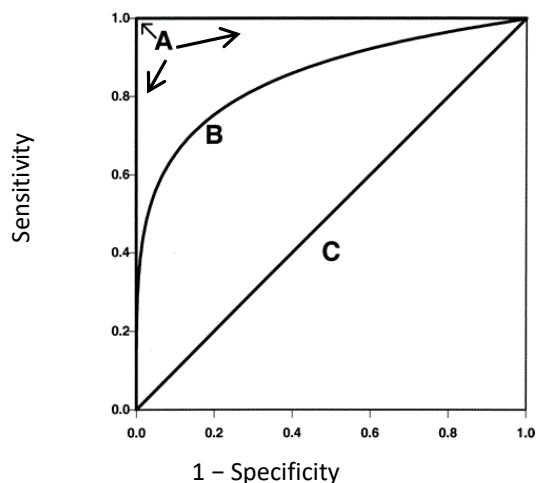


Figure 19: A Sample of Three Hypothetical ROC Curves [Zou, et al., 2007].

7.5.2 The Precision-Recall (PR) Curve

A *precision-recall* (PR) curve may also be used as an assessment tool for diagnostic systems. A PR curve (see also Figure 20) shows the relationship between precision and recall (i.e., the sensitivity or true positive rate). A PR curve is useful when a disease or medical condition is very rare and the training dataset is imbalanced [Davis and Goadrich, 2006], [Saito and Rehmsmeier, 2015a].

There are some non-trivial issues involved when plotting PR curves as described in [Davis and Goadrich, 2006]. Furthermore, the seminal paper [Saito and Rehmsmeier, 2015a] describes some important issues related to the plotting of PR curves and also some domination issues between the ROC and PR curves of the same dataset.

Figure 20 shows two precision-recall curves, which represent the performance of two systems, denoted as A and B. The dotted line P indicates the perfect performance level (i.e., when the AUC under PR is equal to 1.0). The other dotted line, denoted as R, corresponds to a system which is the same as random guessing (i.e., when the AUC under PR is equal to 0.5). The two plots in Figure 20 clearly show that system A outperforms system B, which is also supported by their AUC scores (equal to 0.7 and 0.52, respectively) [Saito and Rehmsmeier, 2015b].

7.5.3 The Predictive Values

The predictive values are another measure to evaluate the performance of a CAD system. The positive predictive value (PPV) (also known as the *precision*) and the negative predictive value (NPV) (see also Table 8) provide the probability that a test by using a given CAD system would result in the true diagnosis [Altman and Bland, 1994]. The values of sensitivity and specificity do not provide this type of information.

The population of patients with a disease does not affect the result of sensitivity and specificity. However, the PPV and NPV values are strongly related to the *prevalence* of a disease in the population (see also Table 8). These values change as the prevalence changes. Prevalence is the proportion of the patients with a disease and usually it is expressed as a percentage (e.g., 5% of the people or 5 out of 100 people have a particular disease). When the PPV value increases, then the prevalence increases too. When the PPV decreases, then the prevalence decreases. At the same time, when the NPV decreases, then the prevalence increases. When the NPV increases, then the prevalence decreases.

Prevalence is usually based on a selected sample group from the entire population, which may create prevalence bias [Heston, 2011]. There are some studies which describe how to avoid prevalence bias in sample data [Heston, 2011], [Fletcher, et al., 2012], [Szklo and Nieto, 2014].

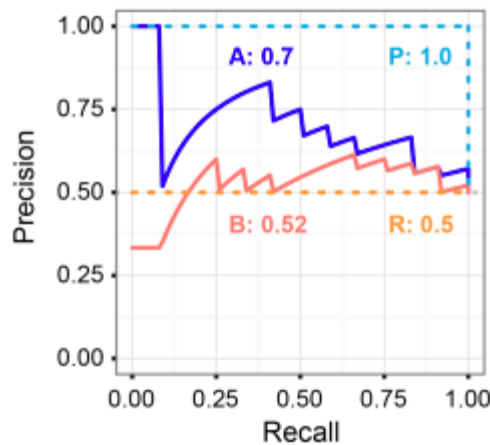


Figure 20: A Sample of Two Hypothetical PR Curves [Saito and Rehmsmeier, 2015b].

7.5.4 The Total Weighted Misclassification Cost

For illustrative purpose, consider the case of a CAD system which produces a diagnosis when it is fed with data related to breast cancer. Approximately 12% of women in the U.S. are diagnosed with breast cancer in their lifetime [DeSantis, et al., 2016]. In this scenario, if we assume that such CAD system always classifies any data record as belonging to a benign case, then its accuracy would be equal to 88%. This is true because in general 88% of the cases are benign while only 12% correspond to malignant cases in a simplified two-class setting. On the other hand, a hypothetical system, which correctly classifies all malignant cases but it also misclassifies, say, 20% of the benign cases as (false) positives would have accuracy rate equal to 80%. That is, it would appear as the first CAD system is more desirable than the second one. Obviously, this situation is highly misleading.

An alternative way for assessing the performance value of CAD systems is to consider the unit costs of making false positive (FP) and false negative (FN) diagnostic decisions. Suppose that these two unit misclassification costs are denoted as CostFP and CostFN, respectively. Then the total weighted misclassification cost (TWMC) of a CAD system can be defined as follows [Pham and Triantaphyllou, 2008a; and 2008b], [Triantaphyllou, 2010]:

$$\text{TWMC} = \text{CostFP} \times \text{FP_Rate} + \text{CostFN} \times \text{FN_Rate}.$$

Therefore, different CAD systems may also be evaluated by means of the TWMC value defined as above. The smaller the TWMC value is, the better the system is. An issue with such approaches is how one can objectively quantify the CostFP and CostFN unit misclassification costs. This issue may require a detailed cost analysis of all the issues involved when misclassifications occur.

8. Discussion

Modern CAD faces several challenges / limitations that require a considerable amount of effort to overcome. Some of these challenges / limitations are stemming from limitations related to machine learning technologies. Deep learning is an emerging subset of machine learning. Some recent developments demonstrate that deep learning approaches can be promising in key CAD areas such as segmentation, feature extraction / selection and classification [LeCun, et al., 2015]. However, there is no single algorithm, which can be applied to all aspects of CAD. New deep learning algorithms and other relevant computational approaches are continuously being developed. Designing better computational algorithms is essential for improving CAD.

Working with complex and massive clinical data can be challenging too. There are several serious problems related to clinical data. The lack of guidelines and standardization regarding electronic health records (EHRs) is crucial [Hripcsak and Albers, 2013]. Bias in training data may lead to unintended classification results. *Big data* in medicine is another challenge in CAD. Dealing with *big data* problems requires massive storage and high processing power. Moreover, analyzing complex data, such as semi-structured or unstructured data, is a big

challenge too. Very likely *big data* will continue to be a key challenge in CAD, although what is considered as “big” today it may not be considered so in the future.

CAD may provide significant diagnostic information for asymptomatic diseases at early stages. However, it may also cause overdiagnosis [Welch and Black, 2010]. Overdiagnosis occurs when a diagnosis is indeed true positive, but the diagnosis does not benefit the patient but can rather be harmful to the patient [Moynihan, et al., 2012]. For instance, frequent screening for cancer has resulted in the identification of early findings consistent with cancer. The rate of identifying such early findings has increased manifold. However, the corresponding mortality rates have not been reduced accordingly. This gap raises some serious doubts about the benefit of early screening for cancer and other diseases and conditions [Welch, 2011]. The overdiagnosis problem is very controversial. Although everybody accepts that overdiagnosis truly occurs, some researchers question its magnitude and argue that early screening is still beneficial [Kopans, et al., 2011], [Smith, et al., 2018], [Wallis, 2018]. More effective CAD systems may be able to reconcile the overdiagnosis problem. The above challenges, along with some other ones, are described in more detail in a different publication by the authors [Yanase and Triantaphyllou, 2019].

9. Concluding Remarks

The numerous and significant advances in computer science and medicine have created various exciting opportunities for expert and intelligent systems in medicine. CAD is the immediate and obvious result of work in the interface of computer science and medicine. The main objective of this paper was to provide a comprehensive overview of CAD in medicine. This paper bridged a critical gap that exists today in the literature regarding this field. It examined, in a systematic manner 251 carefully selected research papers and related materials. This paper highlighted the key advances that took place from the dawn of this field to the current state-of-the-art in this area. The contributions of this paper can be summarized as follows:

- (1) It can enable researchers in expert systems and intelligent systems to better connect their work with CAD systems in medicine. It also proposes the concept of the *intelligent CAD*.
- (2) It provides a venue for a better understanding of key CAD issues among researchers in different fields.
- (3) A researcher in one aspect of CAD, may identify potential R&D opportunities in other aspects of CAD. As an example, segmentation in digital mammography may provide help to segmentation in digital lung cancer diagnosis.
- (4) The paper highlights some opportunities for more research (such as dealing with the overdiagnosis problem).
- (5) This paper may become the foundation for many other related developments and thus be used frequently in the literature as a key reference in the field.

These contributions are significant because they explicitly provide the means to lift some of the existing barriers that inhibit progress in CAD. Thus, these contributions may accelerate progress in CAD. A fundamental barrier in CAD is what we like to call the “*Babel Tower Problem of CAD in Medicine.*” That is, ironically, the multiple advances in medicine and various areas in computer science have caused researchers to lose their understanding of the “big picture” in CAD and thus they keep working in isolation of what happens in other areas of R&D in CAD. This paper offers an opportunity to gain a comprehensive understating of the main issues of CAD systems in medicine. The contributions of this paper are also potentially very impactful because this paper has a good potential to influence research done in CAD. This may impact future research in medicine and in select areas of computer science.

A limitation of this study is that CAD is such a broad topic that it is practically impossible to cover everything in a single paper. Therefore, it is possible, that some issues that could be considered as important by some, may have not been covered at the desired depth in this paper. However, a coordinated effort was made to be as much comprehensive and prioritized as possible. Each section of this paper can be expanded into an important research topic of CAD for future work.

While the demand for CAD has significantly increased, new challenges and limitations that CAD could face in the future may emerge. Some of the challenges are not only the problems that CAD systems in medicine face but also many other expert and intelligence systems directly face as well. If the goal of developing the next generation of CAD is materialized as highlighted in this paper, then CAD will have an even more profound beneficial impact on computer science, medicine, healthcare, and society as a whole.

REFERENCES

- Abbas, A.K. and Bassam, R., 2009. Phonocardiography signal processing. *Synthesis Lectures on Biomedical Engineering*, 4(1), pp.1-194.
- Abbas, Q., Celebi, M.E, and Fondón, I., 2012. Computer - aided pattern classification system for dermoscopy images. *Skin Research and Technology*, 18(3), pp.278-289.
- Acharya, U.R., Fujita, H., Lih, O.S., Hagiwara, Y., Tan, J.H. and Adam, M., 2017. Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network. *Information sciences*, 405, pp.81-90.
- Aggarwal, C.C., 2015. Outlier analysis. In *Data mining* (pp. 237-263). Springer International Publishing.
- Ahmadlou, M. and Adeli, H., 2010. Wavelet-synchronization methodology: a new approach for EEG-based diagnosis of ADHD. *Clinical EEG and Neuroscience*, 41(1), pp.1-10.
- Altman, D.G. and Bland, J.M., 1994. Statistics Notes: Diagnostic tests 2: predictive values. *BMJ*, 309 (6947), p. 102.
- Ang, K.K., Guan, C., Chua, K.S.G., Ang, B.T., Kuah, C., Wang, C., Phua, K.S., Chin, Z.Y. and Zhang, H., 2010, August. Clinical study of neurorehabilitation in stroke using EEG-based motor imagery brain-computer interface with robotic feedback. In *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE* (pp. 5549-5552). IEEE.
- Anthimopoulos, M., Christodoulidis, S., Ebner, L., Christe, A. and Mougiakakou, S., 2016. Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. *IEEE Transactions on Medical Imaging*, 35(5), pp.1207-1216.
- Armato, S.G., Giger, M.L., Moran, C.J., Blackburn, J.T., Doi, K. and MacMahon, H., 1999. Computerized detection of pulmonary nodules on CT scans 1. *Radiographics*, 19(5), pp.1303-1311.
- Atanassov, K., 1986. Intuitionistic fuzzy sets, *Fuzzy Sets and Systems* 20(1), pp. 87-96.
- Atanassov, K., 1999. Intuitionistic fuzzy sets: Theory and applications. Physica-Verlag, New York.
- Ayer, T., Chhatwal, J., Alagoz, O., Kahn Jr, C.E., Woods, R.W. and Burnside, E.S., 2010. Comparison of logistic regression and artificial neural network models in breast cancer risk estimation. *Radiographics*, 30(1), pp.13-22.
- Bar, Y., Diamant, I., Wolf, L., Lieberman, S., Konen, E. and Greenspan, H., 2018. Chest pathology identification using deep feature selection with non-medical training. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 6(3), pp.259-263.
- Bates, D.W., Saria, S., Ohno-Machado, L., Shah, A. and Escobar, G., 2014. Big data in health care: using analytics to identify and manage high-risk and high-cost patients. *Health Affairs*, 33(7), pp.1123-1131.
- Beall, J., 2017. What I learned from predatory publishers. *Biochemia Medica*, 27(2), pp.273-278.
- Beers, M.H. and Berkow, R., 1999. The Merck Manual. *Merck Research Laboratories*, 17.
- Bezem, M., 1988, May. Consistency of rule-based expert systems. In *International Conference on Automated Deduction* (pp. 151-161). Springer, Berlin, Heidelberg, Germany.
- Bhooshan, N., Giger, M., Medved, M., Li, H., Wood, A., Yuan, Y., Lan, L., Marquez, A., Karczmar, G. and Newstead, G., 2014. Potential of computer-aided diagnosis of high spectral and spatial resolution (HiSS) MRI in the classification of breast lesions. *Journal of Magnetic Resonance Imaging*, 39(1), pp.59-67.
- Bleich, H.L., 1972. Computer-based consultation: Electrolyte and acid-base disorders. *The American journal of medicine*, 53(3), pp.285-291.

- Bohadana, A., Izbicki, G. and Kraman, S.S., 2014. Fundamentals of lung auscultation. *New England Journal of Medicine*, 370(8), pp.744-751.
- Boone, J.M., 1994. Sidetracked at the crossroads. *Radiology*, 193(1), pp.28-30.
- Bramer, M.A., 1982. A survey and critical review of expert systems research. *Introductory readings in expert systems*, pp.3-29.
- Breiman, L., 2001. Random forests. *Machine learning*, 45(1), pp.5-32.
- Brennan, T.A., 2004. Medical malpractice. *The New England Journal of Medicine*, 350(3), p.283.
- Bron, E.E., Smits, M., Van Der Flier, W.M., Vrenken, H., Barkhof, F., Scheltens, P., Papma, J.M., Steketee, R.M., Orellana, C.M., Meijboom, R. and Pinto, M., 2015. Standardized evaluation of algorithms for computer-aided diagnosis of dementia based on structural MRI: the CADDementia challenge. *NeuroImage*, 111, pp.562-579.
- Brown, M.S., McNitt-Gray, M.F., Goldin, J.G., Suh, R.D., Sayre, J.W. and Aberle, D.R., 2001. Patient-specific models for lung nodule detection and surveillance in CT images. *IEEE Transactions on Medical Imaging*, 20(12), pp.1242-1250.
- Buchanan, B.G. and Shortliffe, E.H., 1984. *Rule-based Expert Systems: The MYCIN experiments of the Stanford heuristic programming project*, Addison Wesley. Reading.
- Bühler, K., Neubauer, A., Hadwiger, M., Wolsfberger, S. and Wegenkittl, R., 2005. Interactive 3D techniques for computer aided diagnosis and surgery simulation tools. *Digital Revolution in Radiology-Bridging the Future of Health Care, 2nd edn. Springer, Heidelberg*.
- Celebi, M.E., Kingravi, H.A., Iyatomi, H., Alp Aslandogan, Y., Stoecker, W.V., Moss, R.H., Malters, J.M., Grichnik, J.M., Marghoob, A.A., Rabinovitz, H.S. and Menzies, S.W., 2008. Border detection in dermoscopy images using statistical region merging. *Skin Research and Technology*, 14(3), pp.347-353.
- Charniak, E., 2018. *Introduction to Deep Learning*. The MIT Press. Cambridge, MA, USA.
- Chen, C.H., Huang, W.T., Tan, T.H., Chang, C.C. and Chang, Y.J., 2015. Using k-nearest neighbor classification to diagnose abnormal lung sounds. *Sensors*, 15(6), pp.13132-13158.
- Chiarugi, F., Colantonio, S., Emmanouilidou, D., Martinelli, M., Moroni, D. and Salvetti, O., 2010. Decision support in heart failure through processing of electro-and echocardiograms. *Artificial intelligence in medicine*, 50(2), pp.95-104.
- Christ, P.F., Elshaer, M.E.A., Ettlinger, F., Tatavarty, S., Bickel, M., Bilic, P., Rempfler, M., Armbruster, M., Hofmann, F., D'Anastasi, M. and Sommer, W.H., 2016, October. Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3D conditional random fields. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 415-423). Springer, Cham.
- Coats, P.K., 1988. Why expert systems fail. *Financial Management*, pp.77-86.
- Collins, G.S., Reitsma, J.B., Altman, D.G. and Moons, K.G., 2015. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMC medicine*, 13(1), p.1.
- Croisille, P., Souto, M., Cova, M., Wood, S., Afework, Y., Kuhlman, J.E. and Zerhouni, E.A., 1995. Pulmonary nodules: improved detection with vascular segmentation and extraction with spiral CT. *Work in progress. Radiology*, 197(2), pp.397-401.
- Crowley, L., 2011. *Essentials of human disease*. Jones & Bartlett Publishers.
- Deng, H., Deng, W., Sun, X., Ye, C. and Zhou, X., 2016. Adaptive intuitionistic fuzzy enhancement of brain tumor MR images. *Scientific reports*, 6, p.35760.
- Das, S., Guha, D. and Dutta, B., 2016. Medical diagnosis with the aid of using fuzzy logic and intuitionistic fuzzy logic. *Applied Intelligence*, 45(3), pp.850-867.
- Data Resources in the Health Sciences, 2017. *Library Guides at University of Washington*. Accessed February 2, 2017. <http://guides.lib.uw.edu/hsl/data/findclin>
- Davis, J. and Goadrich, M., 2006, June. The relationship between Precision-Recall and ROC curves. In *Proceedings of the 23rd international conference on Machine learning* (pp. 233-240). ACM.

- DeSantis, C.E., Fedewa, S.A., Goding Sauer, A., Kramer, J.L., Smith, R.A. and Jemal, A., 2016. Breast cancer statistics, 2015: Convergence of incidence rates between black and white women. *CA: a cancer journal for clinicians*, 66(1), pp.31-42.
- Deo, R.C., 2015. Machine learning in medicine. *Circulation*, 132(20), pp.1920-1930.
- Dhawan, A.P., 2011. *Medical image analysis* (Vol. 31). John Wiley & Sons.
- Di Bella, G., Minutoli, F., Piaggi, P., Casale, M., Mazzeo, A., Zito, C., Oreto, G., Baldari, S., Vita, G., Pingitore, A. and Khandheria, B.K., 2015. Usefulness of combining electrocardiographic and echocardiographic findings and brain natriuretic peptide in early detection of cardiac amyloidosis in subjects with transthyretin gene mutation. *American Journal of Cardiology*, 116(7), pp.1122-1127.
- Doi, K., 2007. Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *Computerized medical imaging and graphics*, 31(4), pp.198-211.
- D'Orsi, C.J., 2001. Computer-aided Detection: There Is No Free Lunch 1. *Radiology*, 221(3), pp.585-586.
- Drewry, A.M., Fuller, B.M., Skrupky, L.P. and Hotchkiss, R.S., 2015. The presence of hypothermia within 24 hours of sepsis diagnosis predicts persistent lymphopenia. *Critical care medicine*, 43(6), p.1165.
- Dolz, J., Desrosiers, C. and Ayed, I.B., 2018. 3D fully convolutional networks for subcortical segmentation in MRI: A large-scale study. *NeuroImage*, 170, pp.456-470.
- El-Dahshan, E.S.A., Mohsen, H.M., Revett, K. and Salem, A.B.M., 2014. Computer-aided diagnosis of human brain tumor through MRI: A survey and a new algorithm. *Expert systems with Applications*, 41(11), pp.5526-5545.
- Elmore, J.G., Wells, C.K., Lee, C.H., Howard, D.H. and Feinstein, A.R., 1994. Variability in radiologists' interpretations of mammograms. *New England Journal of Medicine*, 331(22), pp.1493-1499.
- Engelbrecht, M.R., Jager, G.J., Laheij, R.J., Verbeek, A.L., Van Lier, H.J. and Barentsz, J.O., 2001. Local staging of prostate cancer using magnetic resonance imaging: a meta-analysis. *European radiology*, 12(9), pp.2294-2302.
- Erickson, B.J., Korfiatis, P., Akkus, Z. and Kline, T.L., 2017. Machine learning for medical imaging. *Radiographics*, 37(2), pp.505-515.
- Fallenberg, E.M., Dromain, C., Diekmann, F., Engelken, F., Krohn, M., Singh, J.M., Ingold-Heppner, B., Winzer, K.J., Bick, U. and Renz, D.M., 2014. Contrast-enhanced spectral mammography versus MRI: initial results in the detection of breast cancer and assessment of tumour size. *European radiology*, 24(1), pp.256-264.
- Feigenbaum, E.A. and McCorduck, P., 1984. The fifth generation. Addison-Wesley, Reading, MA 01867, 275 pages long.
- Ferkol, T. and Schraufnagel, D., 2014. The global burden of respiratory disease. *Annals of the American Thoracic Society*, 11(3), pp.404-406.
- Fernandez-Granero, M.A., Sanchez-Morillo, D. and Leon-Jimenez, A., 2015. Computerised analysis of telemonitored respiratory sounds for predicting acute exacerbations of COPD. *Sensors*, 15(10), pp.26978-26996.
- Fletcher, R.H., Fletcher, S.W. and Fletcher, G.S., 2012. *Clinical epidemiology: the essentials*. Lippincott Williams & Wilkins.
- Gaba, D.M. and Howard, S.K., 2002. Fatigue among clinicians and the safety of patients. *New England Journal of Medicine*, 347(16), pp.1249-1255.
- Giger, M.L., 2018. Machine learning in medical imaging. *Journal of the American College of Radiology*, 15(3), pp.512-520.
- Girshick, R., Donahue, J., Darrell, T. and Malik, J., 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 580-587).
- Goodfellow, I., Y. Benjio, and A. Courville, 2016. Deep Learning. The MIT Press. Cambridge, MA, USA.
- Gorry, G.A., Kassirer, J.P., Essig, A. and Schwartz, W.B., 1973. Decision analysis as the basis for computer-aided management of acute renal failure. *The American journal of medicine*, 55(4), pp.473-484.
- Griner, P.F., Mayewski, R.J., Mushlin, A.I. and Greenland, P., 1981. Selection and interpretation of diagnostic tests and procedures. Principles and applications. *Annals of internal medicine*, 94(4 Pt 2), p.557.

- Grusky, O., Roberts, K.J. and Swanson, A.N., 2007. Failure to return for HIV test results: a pilot study of three community testing sites. *Journal of the International Association of Physicians in AIDS Care*, 6(1), pp.47-55.
- Gupta, C.N., Palaniappan, R., Swaminathan, S. and Krishnan, S.M., 2007. Neural network classification of homomorphic segmented heart sounds. *Applied Soft Computing*, 7(1), pp.286-297.
- Gurney, J.W., 1994. Neural networks at the crossroads: caution ahead. *Radiology*, 193(1), pp.27-28.
- Gurung, A., Scrafford, C.G., Tielsch, J.M., Levine, O.S. and Checkley, W., 2011. Computerized lung sound analysis as diagnostic aid for the detection of abnormal lung sounds: a systematic review and meta-analysis. *Respiratory medicine*, 105(9), pp.1396-1403.
- Guy, G.P., Machlin, S.R., Ekwueme, D.U. and Yabroff, K.R., 2015. Prevalence and Costs of Skin Cancer Treatment in the US, 2002– 2006 and 2007– 2011. *American journal of preventive medicine*, 48(2), pp.183-187.
- Hall, E.J. and Brenner, D.J., 2008. Cancer risks from diagnostic radiology. *The British journal of radiology*, 81(965), pp.362-378.
- Han, F., Yan, H.M., Wei, X.C. and Yan, Q., 2008. Effect of mixed anesthesia on cardiac function by phonocardiogram. *Journal of Electronic Science and Technology of China*, 6(1), pp.91-94.
- Hansen, L.K. and Salamon, P., 1990. Neural network ensembles. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (10), pp.993-1001.
- Hansen, P. and Jaumard, B., 1990. Algorithms for the maximum satisfiability problem. *Computing*, 44(4), pp.279-303.
- Harrell Jr, F.E., 2015. Introduction. In *Regression Modeling Strategies* (pp. 1-11). Springer International Publishing.
- Hastie, T., Tibshirani, R. and Friedman, J., 2009. Unsupervised learning. In *The elements of statistical learning*(pp. 485-585). Springer, New York, NY.
- He, K., Gkioxari, G., Dollár, P. and Girshick, R., 2017. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 2961-2969).
- Helvie, M.A., Hadjiiski, L., Makariou, E., Chan, H.P., Petrick, N., Sahiner, B., Lo, S.C.B., Freedman, M., Adler, D., Bailey, J. and Blane, C., 2004. Sensitivity of Noncommercial Computer-aided Detection System for Mammographic Breast Cancer Detection: Pilot Clinical Trial 1. *Radiology*, 231(1), pp.208-214.
- Hendriks, P.H. and Vriens, D.J., 1999. Knowledge-based systems and knowledge management: friends or foes?. *Information & Management*, 35(2), pp.113-125.
- Heston, T.F., 2011. Standardizing predictive values in diagnostic imaging research. *Journal of Magnetic Resonance Imaging*, 33(2), pp.505-505.
- Hodge, V. and Austin, J., 2004. A survey of outlier detection methodologies. *Artificial intelligence review*, 22(2), pp.85-126.
- Hoffmann, K., Gambichler, T., Rick, A., Kreutz, M., Anschuetz, M., Grünendick, T., Orlikov, A., Gehlen, S., Perotti, R., Andreassi, L. and Newton Bishop, J., 2003. Diagnostic and neural analysis of skin cancer (DANAOS). A multicentre study for collection and computer - aided analysis of data from pigmented skin lesions using digital dermoscopy. *British J. of Dermatology*, 149(4), pp.801-809.
- Hripcsak, G. and Albers, D.J., 2013. Correlating electronic health record concepts with healthcare process events. *Journal of the American Medical Informatics Association*, 20(e2), pp.e311-e318.
- İçer, S. and Genç, Ş., 2014. Classification and analysis of non-stationary characteristics of crackle and rhonchus lung adventitious sounds. *Digital Signal Processing*, 28, pp.18-27.
- Jain, A.K., 1989. *Fundamentals of digital image processing*. Prentice-Hall, Inc.
- Kable, A.K., Gibberd, R.W. and Spigelman, A.D., 2002. Adverse events in surgical patients in Australia. *International Journal for Quality in Health Care*, 14(4), pp.269-276.
- Kak, A.C., Slaney, M. and Wang, G., 2002. Principles of computerized tomographic imaging. *Medical Physics*, 29(1), pp.107-107.
- Kamnitsas, K., Ledig, C., Newcombe, V.F., Simpson, J.P., Kane, A.D., Menon, D.K., Rueckert, D. and Glocker, B., 2017. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Medical image analysis*, 36, pp.61-78.

- Karp, R.M., 1972. Reducibility among Combinatorial Problems. In: Miller R.E., Thatcher J.W., Bohlinger J.D. (eds) Complexity of Computer Computations. The IBM Research Symposia Series. Springer, Boston, MA, pp. 85-103.
- Kiyokawa, H., Greenberg, M., Shirota, K. and Pasterkamp, H., 2001. Auditory detection of simulated crackles in breath sounds. *Chest*, 119(6), pp. 1886-1892.
- Kiranyaz, S., Ince, T. and Gabbouj, M., 2015. Real-time patient-specific ECG classification by 1-D convolutional neural networks. *IEEE Transactions on Biomedical Engineering*, 63(3), pp.664-675.
- Kirlangic, M.E., Perez, D., Kudryavtseva, S., Griessbach, G., Henning, G. and Ivanova, G., 2001. Fractal dimension as a feature for adaptive electroencephalogram segmentation in epilepsy. In *Engineering in Medicine and Biology Society, 2001. Proceedings of the 23rd Annual International Conference of the IEEE* (Vol. 2, pp. 1573-1576). IEEE.
- Klarkowski, D.B., Wazome, J.M., Lokuge, K.M., Shanks, L., Mills, C.F. and O'Brien, D.P., 2009. The evaluation of a rapid in situ HIV confirmation test in a programme with a high failure rate of the WHO HIV two-test diagnostic algorithm. *PLoS One*, 4(2), p.e 4351.
- Kohavi, R. and John, G.H., 1997. Wrappers for feature subset selection. *Artificial intelligence*, 97(1), pp.273-324.
- König, T., Prichep, L., Dierks, T., Hubl, D., Wahlund, L.O., John, E.R. and Jelic, V., 2005. Decreased EEG synchronization in Alzheimer's disease and mild cognitive impairment. *Neurobiology of aging*, 26(2), pp.165-171.
- Kononenko, I., 2001. Machine learning for medical diagnosis: history, state of the art and perspective. *Artificial Intelligence in Medicine*, 23(1), pp.89-109.
- Kopans, D.B., 1994. The accuracy of mammographic interpretation. *The New England journal of medicine*, 331(22), pp.1521-1522.
- Kopans, D.B., Smith, R.A. and Duffy, S.W., 2011. Mammographic screening and "overdiagnosis". *Radiology*, 260(3), pp.616-620.
- Korn, F., Sidiropoulos, N., Faloutsos, C., Siegel, E. and Protopapas, Z., 1998. Fast nearest neighbor search in medical image databases. University of Maryland, Digital Repository at the University of Maryland, Tech Reports in Computer Science and Engineering, Accessed April of 2017. URL: <http://hdl.handle.net/1903/805>
- Kovalerchuk, B., Triantaphyllou, E., Deshpande, A.S. and Vityaev, E., 1996. Interactive learning of monotone Boolean functions. *Information Sciences*, 94(1-4), pp.87-118.
- Kovalerchuk, B., Triantaphyllou, E., Ruiz, J.F. and Clayton, J., 1997. Fuzzy logic in computer-aided breast cancer diagnosis: analysis of lobulation. *Artificial Intelligence in Medicine*, 11(1), pp.75-85.
- Kovalerchuk, B., Triantaphyllou, E., Ruiz, J.F., Torvik, V.I. and Vityaev, E., 2000. The reliability issue of computer-aided breast cancer diagnosis. *Computers and Biomedical Research*, 33(4), pp.296-313.
- Kuhl, C.K., Schrading, S., Leutner, C.C., Morakkabati-Spitz, N., Wardelmann, E., Fimmers, R., Kuhn, W. and Schild, H.H., 2005. Mammography, breast ultrasound, and magnetic resonance imaging for surveillance of women at high familial risk for breast cancer. *Journal of clinical oncology*, 23(33), pp.8469-8476.
- Kumar, A., Kim, J., Cai, W., Fulham, M. and Feng, D., 2013. Content-based medical image retrieval: a survey of applications to multidimensional and multimodality data. *Journal of digital imaging*, 26(6), pp.1025-1039.
- Kumar, A., Kim, J., Lyndon, D., Fulham, M. and Feng, D., 2016. An ensemble of fine-tuned convolutional neural networks for medical image classification. *IEEE journal of biomedical and health informatics*, 21(1), pp.31-40.
- Kuo, R.J., Lin, T.C., Zulfia, F.E. and Tsai, C.Y., 2018. A hybrid metaheuristic and kernel intuitionistic fuzzy c-means algorithm for cluster analysis. *Applied Soft Computing*, 67, pp.299-308.
- Lalkhen, A.G. and McCluskey, A., 2008. Clinical tests: sensitivity and specificity. *Continuing Education in Anaesthesia, Critical Care & Pain*, 8(6), pp.221-223.
- Landrigan, C.P., Rothschild, J.M., Cronin, J.W., Kaushal, R., Burdick, E., Katz, J.T., Lilly, C.M., Stone, P.H., Lockley, S.W., Bates, D.W. and Czeisler, C.A., 2004. Effect of reducing interns' work hours on serious medical errors in intensive care units. *New England Journal of Medicine*, 351(18), pp.1838-1848.

- Leape, L.L., Brennan, T.A., Laird, N., Lawthers, A.G., Localio, A.R., Barnes, B.A., Hebert, L., Newhouse, J.P., Weiler, P.C. and Hiatt, H., 1991. The nature of adverse events in hospitalized patients: results of the Harvard Medical Practice Study II. *New England journal of medicine*, 324(6), pp.377-384.
- LeCun, Y., Bengio, Y. and Hinton, G., 2015. Deep learning. *nature*, 521(7553), p.436.
- Lee, Y., Hara, T., Fujita, H., Itoh, S. and Ishigaki, T., 2001. Automated detection of pulmonary nodules in helical CT images based on an improved template-matching technique. *IEEE Transactions on medical imaging*, 20(7), pp.595-604.
- Lemaître, G., Martí, R., Freixenet, J., Vilanova, J.C., Walker, P.M. and Meriaudeau, F., 2015. Computer-aided detection and diagnosis for prostate cancer based on mono and multi-parametric MRI: a review. *Computers in Biology and Medicine*, 60, pp.8-31.
- Lenartowicz, A. and Loo, S.K., 2014. Use of EEG to diagnose ADHD. *Current psychiatry reports*, 16(11), p.498.
- Lenat, D. and Feigenbaum, E., 1992. On the thresholds of knowledge. *Foundations of Artificial Intelligence*, MIT Press, Cambridge, MA, pp.185-250.
- Leng, S., San Tan, R., Chai, K.T.C., Wang, C., Ghista, D. and Zhong, L., 2015. The electronic stethoscope. *Biomedical engineering online*, 14(1), p.1.
- Ledley, R.S., 1959. Digital electronic computers in biomedical science. *Science*, 130(3384), pp.1225-1234.
- Ledley, R.S. and Lusted, L.B., 1959. Reasoning foundations of medical diagnosis. *Science*, 130(3366), pp.9-21.
- Lehman, C.D., Wellman, R.D., Buist, D.S., Kerlikowske, K., Tosteson, A.N. and Miglioretti, D.L., 2015. Diagnostic accuracy of digital screening mammography with and without computer-aided detection. *JAMA internal medicine*, 175(11), pp.1828-1837.
- Li, W., Jia, F. and Hu, Q., 2015. Automatic segmentation of liver tumor in CT images with deep convolutional neural networks. *Journal of Computer and Communications*, 3(11), p.146.
- Liao, S.H., 2005. Expert system methodologies and applications—a decade review from 1995 to 2004. *Expert systems with applications*, 28(1), pp.93-103.
- Liaw, A. and Wiener, M., 2002. Classification and Regression by randomForest. *R news*, 2(3), pp.18-22.
- Lin, K.P., Chang, H.F., Chen, T.L., Lu, Y.M. and Wang, C.H., 2016. Intuitionistic fuzzy C-regression by using least squares support vector regression. *Expert Systems with Applications*, 64, pp.296-304.
- Litjens, G.J., Barentsz, J.O., Karssemeijer, N. and Huisman, H.J., 2015. Clinical evaluation of a computer-aided diagnosis system for determining cancer aggressiveness in prostate MRI. *European radiology*, 25(11), pp.3187-3199.
- Liu, H. and Motoda, H., 2012. *Feature selection for knowledge discovery and data mining* (Vol. 454). Springer Science & Business Media.
- Liu, H., Motoda, H., Setiono, R. and Zhao, Z., 2010, May. Feature selection: An ever evolving frontier in data mining. In *Feature Selection in Data Mining* (pp. 4-13).
- Lodwick, G.S., Haun, C.L., Smith, W.E., Keller, R.F. and Robertson, E.D., 1963. Computer diagnosis of primary bone tumors: a preliminary report 1. *Radiology*, 80(2), pp.273-275.
- Long, J., Shelhamer, E. and Darrell, T., 2015. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3431-3440).
- López, V., Fernández, A., García, S., Palade, V. and Herrera, F., 2013. An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. *Information Sciences*, 250, pp.113-141.
- Ludlow, A., Mohr, B., Whitmore, A., Garagnani, M., Pulvermüller, F. and Gutierrez, R., 2014. Auditory processing and sensory behaviours in children with autism spectrum disorders as revealed by mismatch negativity. *Brain and cognition*, 86, pp.55-63.
- Obermeyer, Z. and Emanuel, E.J., 2016. Predicting the future—big data, machine learning, and clinical medicine. *The New England journal of medicine*, 375(13), p.1216.

- Madhero, 2010. Phonocardiograms from normal and abnormal heart sounds. https://en.wikipedia.org/wiki/Heart_sounds Accessed July 18, 2016.
- Maldonado, S., Weber, R. and Famili, F., 2014. Feature selection for high-dimensional class-imbalanced data sets using Support Vector Machines. *Information Sciences*, 286, pp.228-246.
- Mahnke, C.B., Nowalk, A., Hofkosh, D., Zuberbuhler, J.R. and Law, Y.M., 2004. Comparison of two educational interventions on pediatric resident auscultation skills. *Pediatrics*, 113(5), pp.1331-1335.
- Mak, B., Schmitt, B.H. and Lyytinen, K., 1997. User participation in knowledge update of expert systems. *Information & Management*, 32(2), pp.55-63.
- Malladi, R., Sethian, J.A. and Vemuri, B.C., 1995. Shape modeling with front propagation: A level set approach. *IEEE transactions on pattern analysis and machine intelligence*, 17(2), pp.158-175.
- Marsousi, M., Plataniotis, K.N. and Stergiopoulos, S., 2014, August. Shape-based kidney detection and segmentation in three-dimensional abdominal ultrasound images. In *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE* (pp. 2890-2894). IEEE.
- Matsushima, K. and Frankel, H.L., 2011. Beyond focused assessment with sonography for trauma: ultrasound creep in the trauma resuscitation area and beyond. *Current opinion in critical care*, 17(6), pp.606-612.
- Martin, G.S., Mannino, D.M., Eaton, S. and Moss, M., 2003. The epidemiology of sepsis in the United States from 1979 through 2000. *New England Journal of Medicine*, 348(16), pp.1546-1554.
- Meyers, P.H., Nice Jr, C.M., Becker, H.C., Nettleton Jr, W.J., Sweeney, J.W. and Meckstroth, G.R., 1964. Automated computer analysis of radiographic images 1. *Radiology*, 83(6), pp.1029-1034.
- Micali, G. and Lacarrubba, F. eds., 2016. *Dermatoscopy in clinical practice: beyond pigmented lesions*. CRC Press.
- Mikami, R., Murao, M., Cugell, D.W., Chrétien, J., Cole, P., Meier-Sydow, J., Murphy, R.L. and Loudon, R.G., 1987. International symposium on lung sounds. Synopsis of proceedings. *CHEST Journal*, 92(2), pp.342-345.
- Miki, Y., Muramatsu, C., Hayashi, T., Zhou, X., Hara, T., Katsumata, A. and Fujita, H., 2017. Classification of teeth in cone-beam CT using deep convolutional neural network. *Computers in biology and medicine*, 80, pp.24-29.
- Miller, K.E., Capan, M., Wu, P., Jackson, E.V. and Arnold, R.C., 2015, June. Operationalizing Sepsis Alert Design and Clinical Decision Support Developing Enhanced Visual Display Models. In *Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care* (Vol. 4, No. 1, pp. 103-109). SAGE Publications.
- Miller, R.A., Pople Jr, H.E. and Myers, J.D., 1982. Internist-I, an experimental computer-based diagnostic consultant for general internal medicine. *New England Journal of Medicine*, 307(8), pp.468-476.
- Milletari, F., Navab, N. and Ahmadi, S.A., 2016, October. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 Fourth International Conference on 3D Vision (3DV)* (pp. 565-571). IEEE.
- Moeskops, P., Viergever, M.A., Mendrik, A.M., de Vries, L.S., Benders, M.J. and Išgum, I., 2016. Automatic segmentation of MR brain images with a convolutional neural network. *IEEE Transactions on Medical Imaging*, 35(5), pp.1252-1261.
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G. and Prisma Group, 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS medicine*, 6(7), p.e1000097.
- Mohr, J.C., 2000. American medical malpractice litigation in historical perspective. *Jama*, 283(13), pp.1731-1737.
- Moynihan, R., Doust, J. and Henry, D., 2012. Preventing overdiagnosis: how to stop harming the healthy. *Bmj*, 344, p.e3502.
- Morrow, M., Waters, J. and Morris, E., 2011. MRI for breast cancer screening, diagnosis, and treatment. *The Lancet*, 378(9805), pp.1804-1811.
- Murphy, R.L., Vyshedskiy, A., Power-Charnitsky, V.A., Bana, D.S., Marinelli, P.M., Wong-Tse, A. and Paciej, R., 2004. Automated lung sound analysis in patients with pneumonia. *Respiratory Care*, 49(12), pp.1490-1497.
- Musen, M.A., Middleton, B. and Greenes, R.A., 2014. Clinical decision-support systems. In *Biomedical informatics* (pp. 643-674). Springer London.

- Nahed, B.V., Babu, M.A., Smith, T.R. and Heary, R.F., 2012. Malpractice liability and defensive medicine: a national survey of neurosurgeons. *PloS one*, 7(6), p.e39237.
- Nature Index, 2017. *Institution Outputs*. URL = <https://www.natureindex.com/institution-outputs/generate/All/global/All/score> Accessed April 2017.
- Nickoloff, E.L. and Alderson, P.O., 2001. Radiation exposures to patients from CT: reality, public perception, and policy. *American Journal of Roentgenology*, 177(2), pp.285-287.
- Nieblas, C.I., Alonso, M.A., Conte, R. and Villarreal, S., 2013, August. High performance heart sound segmentation algorithm based on matching pursuit. In *Digital Signal Processing and Signal Processing Education Meeting (DSP/SPE), 2013 IEEE* (pp. 96-100). IEEE.
- Nishikawa, R.M., Schmidt, R.A., Linver, M.N., Edwards, A.V., Papaioannou, J. and Stull, M.A., 2012. Clinically missed cancer: how effectively can radiologists use computer-aided detection? *American Journal of Roentgenology*, 198(3), pp.708-716.
- Nixon, M. and Aguado, A.S., 2012. *Feature extraction and image processing for computer vision*. Academic Press.
- Ng, H.P., Ong, S.H., Foong, K.W.C., Goh, P.S. and Nowinski, W.L., 2006, March. Medical image segmentation using k-means clustering and improved watershed algorithm. In *Image Analysis and Interpretation, 2006 IEEE Southwest Symposium on* (pp. 61-65). IEEE.
- Nunes, D., Leal, A., Couceiro, R., Henriques, J., Mendes, L., Carvalho, P. and Teixeira, C., 2015, August. A low-complex multi-channel methodology for noise detection in phonocardiogram signals. In *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE* (pp. 5936-5939). IEEE.
- Nunez, P.L. and Srinivasan, R., 2006. *Electric fields of the brain: the neurophysics of EEG*. Oxford University Press, USA.
- Peng, X. and Selvachandran, G., 2017. Pythagorean fuzzy set: state of the art and future directions. *Artificial Intelligence Review*, pp. 1-55.
- Peng, X. and Dai, J., 2018. A bibliometric analysis of neutrosophic set: Two decades review from 1998 to 2017. *Artificial Intelligence Review*, pp. 1-57.
- Peng, X. and Yang, Y., 2015. Some results for Pythagorean fuzzy sets. *International Journal of Intelligent Systems*, 30(11), pp. 1133-1160.
- Patidar, S. and Pachori, R.B., 2013. Segmentation of cardiac sound signals by removing murmurs using constrained tunable-Q wavelet transform. *Biomedical Signal Processing and Control*, 8(6), pp.559-567.
- Petrack, N., Sahiner, B., Armato, S.G., Bert, A., Correale, L., Delsanto, S., Freedman, M.T., Fryd, D., Gur, D., Hadjiiski, L. and Huo, Z., 2013. Evaluation of computer-aided detection and diagnosis systems. *Medical physics*, 40(8).
- Pham, H.N.A. and Triantaphyllou, E., 2008a. The Impact of Overfitting and Overgeneralization on the Classification Accuracy in Data Mining. *Soft computing for knowledge discovery and data mining*, (Part 4), pp.391-431.
- Pham, H.N.A. and Triantaphyllou, E., 2008b. Prediction of diabetes by employing a new data mining approach which balances fitting and generalization. *Computer and Inf. Science*, pp.11-26.
- Pławiak, P., 2018. Novel methodology of cardiac health recognition based on ECG signals and evolutionary-neural system. *Expert Systems with Applications*, 92, pp.334-349.
- Podgorelec, V., Kokol, P., Stiglic, B. and Rozman, I., 2002. Decision trees: an overview and their use in medicine. *Journal of medical systems*, 26(5), pp.445-463.
- Porter, R.S., Kaplan, J.L., Homeier, B. and Beers, M., 2011. *The Merck Manual for health care professionals*. Merck and Co.
- Prostate Knowledge, 2009, Harvard Medical School + Harvard Health Publications. Accessed January 18, 2017. <http://www.harvardprostateknowledge.org>
- Razzaghi, T., Roderick, O., Safro, I. and Marko, N., 2015, July. Fast imbalanced classification of healthcare data with missing values. In *Information Fusion (Fusion), 2015 18th International Conference on* (pp. 774-781). IEEE.

- Rad, A.B., Eftestøl, T., Engan, K., Irusta, U., Kvaløy, J.T., Kramer-Johansen, J., Wik, L. and Katsaggelos, A.K., 2017. ECG-based classification of resuscitation cardiac rhythms for retrospective data analysis. *IEEE Transactions on Biomedical Engineering*, 64(10), pp.2411-2418.
- Reed, T.R., Reed, N.E. and Fritzon, P., 2004. Heart sound analysis for symptom detection and computer-aided diagnosis. *Simulation Modelling Practice and Theory*, 12(2), pp.129-146.
- Reichert, S., Gass, R., Brandt, C. and Andrès, E., 2008. Analysis of Respiratory Sounds: State of the Art. *Clinical Medicine: Circulatory, Respiratory and Pulmonary Medicine*, 2, pp.45-58.
- Ren, Z., Tang, J., Kalscheuer, T. and Maurer, H., 2017. Fast 3 - D large - scale gravity and magnetic modeling using unstructured grids and an adaptive multilevel fast multipole method. *Journal of Geophysical Research: Solid Earth*, 122(1), pp.79-109.
- Richhariya, B. and Tanveer, M., 2018. EEG signal classification using universum support vector machine. *Expert Systems with Applications*, 106, pp.169-182.
- Rifkin, M.D., Zerhouni, E.A., Gatsonis, C.A., Quint, L.E., Paushter, D.M., Epstein, J.I., Hamper, U., Walsh, P.C. and McNeil, B.J., 1990. Comparison of magnetic resonance imaging and ultrasonography in staging early prostate cancer: results of a multi-institutional cooperative trial. *New England Journal of Medicine*, 323(10), pp.621-626.
- Romera-Paredes, B. and Torr, P.H.S., 2016, October. Recurrent instance segmentation. In *European conference on computer vision* (pp. 312-329). Springer, Cham.
- Rosati, R.A., McNeer, J.F., Starmer, C.F., Mittler, B.S., Morris, J.J. and Wallace, A.G., 1975. A new information system for medical practice. *Archives of Internal Medicine*, 135(8), pp.1017-1024.
- Roy, D.L., 2003. The paediatrician and cardiac auscultation. *Paediatrics & child health*, 8(9), p.561.
- Roy, D., Sargeant, J., Gray, J., Hoyt, B., Allen, M. and Fleming, M., 2002. Helping family physicians improve their cardiac auscultation skills with an interactive CD - ROM. *Journal of Continuing Education in the Health Professions*, 22(3), pp.152-159.
- Safara, F., Doraisamy, S., Azman, A., Jantan, A. and Ramaiah, A.R.A., 2013. Multi-level basis selection of wavelet packet decomposition tree for heart sound classification. *Computers in biology and medicine*, 43(10), pp.1407-1414.
- Saito, T. and Rehmsmeier, M., 2015a. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS one*, 10(3), p.e0118432.
- Saito, T. and Rehmsmeier, M., 2015b. Classifier evaluation with imbalanced datasets - WordPress.com; Introduction to the precision-recall plot. Accessed November 21, 2016.
<https://classeval.wordpress.com/introduction/introduction-to-the-precision-recall-plot/>
- Schiff, G.D., Hasan, O., Kim, S., Abrams, R., Cosby, K., Lambert, B.L., Elstein, A.S., Hasler, S., Kabongo, M.L., Krosnjak, N. and Odwazny, R., 2009. Diagnostic error in medicine: analysis of 583 physician-reported errors. *Archives of internal medicine*, 169(20), pp.1881-1887.
- Schindera, S.T., Odedra, D., Raza, S.A., Kim, T.K., Jang, H.J., Szucs-Farkas, Z. and Rogalla, P., 2013. Iterative reconstruction algorithm for CT: can radiation dose be decreased while low-contrast detectability is preserved? *Radiology*, 269(2), pp.511-518.
- Schmidt, S.E., Holst-Hansen, C., Graff, C., Toft, E. and Struijk, J.J., 2010. Segmentation of heart sound recordings by a duration-dependent hidden Markov model. *Physiological measurement*, 31(4), p.513.
- Schmid-Saugeona, P., Guillod, J. and Thirana, J.P., 2003. Towards a computer-aided diagnosis system for pigmented skin lesions. *Computerized Medical Imaging and Graphics*, 27(1), pp.65-78.
- Schwartz, W.B., 1970. Medicine and the computer: the promise and problems of change. *New England Journal of Medicine*, 283(23), pp.1257-1264.
- Seber, G.A. and Lee, A.J., 2012. *Linear regression analysis* (Vol. 329). John Wiley & Sons.
- Seera, M. and Lim, C.P., 2014. A hybrid intelligent system for medical data classification. *Expert Systems with Applications*, 41(5), pp.2239-2249.
- Şeref, O., Razzaghi, T. and Xanthopoulos, P., 2017. Weighted relaxed support vector machines. *Annals of Operations Research*, 249(1-2), pp.235-271.

- Shan, N. and Ziarko, W., 1995. Data - based acquisition and incremental modification of classification rules. *Computational Intelligence*, 11(2), pp.357-370.
- Shin, H.C., Roth, H.R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D. and Summers, R.M., 2016. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE transactions on medical imaging*, 35(5), pp.1285-1298.
- Shortliffe, E.H. and Buchanan, B.G., 1975. A model of inexact reasoning in medicine. *Mathematical biosciences*, 23(3-4), pp.351-379.
- Smarandache, F., 1998. *Neutrosophy. Neutrosophic Probability, Set, and Logic*. Amer. Res. Press, Rehoboth, USA, 105 pages long.
- Smith, R.A., Brawley, O.W. and Wender, R.C., 2018. Screening and early detection. *The American Cancer Society's Principles of Oncology: Prevention to Survivorship*, pp.110-135.
- Snyder, S.M., Rugino, T.A., Hornig, M. and Stein, M.A., 2015. Integration of an EEG biomarker with a clinician's ADHD evaluation. *Brain and behavior*, 5(4).
- Sonka, M., Hlavac, V. and Boyle, R., 2014. *Image processing, analysis, and machine vision*. Cengage Learning.
- Sotiras, A., Davatzikos, C. and Paragios, N., 2013. Deformable medical image registration: A survey. *IEEE transactions on medical imaging*, 32(7), pp.1153-1190.
- Subhawong, T.K., Fishman, E.K., Swart, J.E., Carrino, J.A., Attar, S. and Fayad, L.M., 2010. Soft-tissue masses and mass-like conditions: what does CT add to diagnosis and management?. *American Journal of Roentgenology*, 194(6), pp.1559-1567.
- Suetens, P., 2017. *Fundamentals of medical imaging*. Cambridge university press.
- Summers, R.M., 2002. Challenges for computer-aided diagnosis for CT colonography. *Abdominal imaging*, 27(3), pp.268-274.
- Sun, W., Zheng, B. and Qian, W., 2016, March. Computer aided lung cancer diagnosis with deep learning algorithms. In *Medical imaging 2016: computer-aided diagnosis* (Vol. 9785, p. 97850Z). International Society for Optics and Photonics.
- Suykens, J.A. and Vandewalle, J., 1999. Least squares support vector machine classifiers. *Neural processing letters*, 9(3), pp.293-300.
- Szolovits, P., Patil, R.S. and Schwartz, W.B., 1988. Artificial intelligence in medical diagnosis. *Annals of internal medicine*, 108(1), pp.80-87.
- Szko, M. and Nieto, J., 2014. *Epidemiology*. Jones & Bartlett Publishers.
- Tan, P.N., Steinbach, M. and Kumar, V., 2006. *Data Mining Introduction*. By Pearson Education, Inc. London, UK.
- Tang, B.M., Eslick, G.D., Craig, J.C. and McLean, A.S., 2007. Accuracy of procalcitonin for sepsis diagnosis in critically ill patients: systematic review and meta-analysis. *The Lancet infectious diseases*, 7(3), pp.210-217.
- Tang, J., Alelyani, S. and Liu, H., 2014. Feature selection for classification: A review. *Data Classification: Algorithms and Applications*, p.37.
- Theis, F.J. and Meyer-Bäse, A., 2010. *Biomedical signal analysis: Contemporary methods and applications*. MIT Press.
- Thomaz, R.L., Carneiro, P.C. and Patrocínio, A.C., 2017, March. Feature extraction using convolutional neural network for classifying breast density in mammographic images. In *Medical Imaging 2017: Computer-Aided Diagnosis* (Vol. 10134, p. 101342M). International Society for Optics and Photonics.
- Thung, K.H., Yap, P.T. and Shen, D., 2017. Multi-stage Diagnosis of Alzheimer's Disease with Incomplete Multimodal Data via Multi-Task Deep Learning. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support* (pp. 160-168). Springer, Cham.
- Torasso, P., 1985. Knowledge based expert systems for medical diagnosis. *Statistics in medicine*, 4(3), pp.317-325.
- Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), pp.267-288.
- Times Higher Education, 2017. *World Reputation Rankings, 2016*. URL = https://www.timeshighereducation.com/world-university-rankings/2016/reputation-ranking#!/page/0/length/25/sort/by/rank/sort_order/asc/cols/stats Accessed April 2017.

- Trapeznikov, K. and Saligrama, V., 2013, April. Supervised sequential classification under budget constraints. In *Artificial Intelligence and Statistics* (pp. 581-589).
- Triantaphyllou E., 2000. Multi-criteria Decision Making Methods: A Comparative Study. Applied Optimization, Vol. 44. Springer, Boston, MA
- Triantaphyllou, E., 2010. *Data mining and knowledge discovery via logic-based methods: theory, algorithms, and applications* (Vol. 43). Springer Science & Business Media.
- Triantaphyllou, E. and Lin, C.T., 1996. Development and evaluation of five fuzzy multiattribute decision-making methods. *International Journal of Approximate Reasoning*, 14(4), pp. 281-310.
- Uğuz, H., 2012. A biomedical system based on artificial neural network and principal component analysis for diagnosis of the heart valve diseases. *Journal of medical systems*, 36(1), pp.61-72.
- U.S. News, 2017. *U.S. News Best Colleges Rankings*. URL = <https://www.usnews.com/best-colleges> Accessed April 2017.
- Vandenberg, S.G., 1960. Medical diagnosis by computer: Recent attempts and outlook for the future. *Behavioral Science*, 5(2), p.170.
- van Ginneken, B., Schaefer-Prokop, C.M. and Prokop, M., 2011. Computer-aided diagnosis: how to move from the laboratory to the clinic. *Radiology*, 261(3), pp.719-732.
- Vlachos, I.K. and Sergiadis, G.D., 2007. Intuitionistic fuzzy information—applications to pattern recognition. *Pattern Recognition Letters*, 28(2), pp.197-206.
- Vincent, C., Moorthy, K., Sarker, S.K., Chang, A. and Darzi, A.W., 2004. Systems approaches to surgical quality and safety: from concept to measurement. *Annals of surgery*, 239(4), pp.475-482.
- Wallis, M.G., 2018. How do we manage overdiagnosis/overtreatment in breast screening?. *Clinical radiology*, 73(4), pp.372-380.
- Wang, L., 2014. 'Photoacoustic tomography' in *Scholarpedia*, 9(2):10278. Accessed July 2, 2017. http://www.scholarpedia.org/article/Photo_acoustic_tomography
- Ward, M.O., Grinstein, G. and Keim, D., 2010. *Interactive data visualization: foundations, techniques, and applications*. CRC Press.
- Waterman, D.A., 1985. *A guide to expert systems*.
- Way, T.W., Hadjiiski, L.M., Sahiner, B., Chan, H.P., Cascade, P.N., Kazerooni, E.A., Bogot, N. and Zhou, C., 2006. Computer-aided diagnosis of pulmonary nodules on CT scans: segmentation and classification using 3D active contours. *Medical Physics*, 33(7), pp.2323-2337.
- Weinrauch, H. and Hetherington, A.W., 1959. Computers in medicine and biology. *Journal of the American Medical Association*, 169(3), pp.240-245.
- Weinstein, R.S., Lopez, A.M., Joseph, B.A., Erps, K.A., Holcomb, M., Barker, G.P. and Krupinski, E.A., 2014. Telemedicine, telehealth, and mobile health applications that work: opportunities and barriers. *The American journal of medicine*, 127(3), pp.183-187.
- Weiss, S.M., Kulikowski, C.A., Amarel, S. and Safir, A., 1978. A model-based method for computer-aided medical decision-making. *Artificial intelligence*, 11(1-2), pp.145-172.
- Welch, H.G., 2011. *Overdiagnosed: making people sick in the pursuit of health*. Beacon Press.
- Welch, H.G. and Black, W.C., 2010. Overdiagnosis in cancer. *Journal of the National Cancer Institute*, 102(9), pp.605-613.
- Winsberg, F., Elkin, M., Macy Jr, J., Bordaz, V. and Weymouth, W., 1967. Detection of radiographic abnormalities in mammograms by means of optical scanning and computer analysis 1. *Radiology*, 89(2), pp.211-215.
- Xia, J., Yao, J. and Wang, L.V., 2014. Photoacoustic tomography: principles and advances. *Electromagnetic waves (Cambridge, Mass.)*, 147, p.1.
- Yaffe, M.J. and Mainprize, J.G., 2011. Risk of radiation-induced breast cancer from mammographic screening. *Radiology*, 258(1), pp.98-105.
- Yale University Library, 2017. *Choosing a Journal for Publication of an Article: List of Suspicious Journals and Publishers*. URL = <https://guides.library.yale.edu/c.php?g=296124&p=1973764> Accessed April 2017.

- Yanase, J., and E. Triantaphyllou, 2019. The Seven Key Challenges for the Future of Computer-Aided Diagnosis in Medicine. *Journal of Medical Informatics*, <https://doi.org/10.1016/j.ijmedinf.2019.06.017>
- Yang, D., Xu, D., Zhou, S.K., Georgescu, B., Chen, M., Grbic, S., Metaxas, D. and Comaniciu, D., 2017, September. Automatic Liver Segmentation Using an Adversarial Image-to-Image Network. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 507-515). Springer, Cham.
- Yager, R.R., 2013. Pythagorean fuzzy subsets. In: *Proceedings of the Joint IFSAWorld Congress and NAFIPS Annual Meeting*, pp. 57–61.
- Yager, R.R., 2014. Pythagorean membership grades in multicriteria decision making. *IEEE Trans Fuzzy Syst* 22(4), pp. 958–965.
- Yager, R.R., 2016. *Properties and Applications of Pythagorean Fuzzy Sets*. Springer, Berlin.
- Yang, Y. and Pedersen, J.O., 1997, July. A comparative study on feature selection in text categorization. In *Icml* (Vol. 97, No. 412-420, p. 35).
- Yuan, Y., Giger, M.L., Li, H., Suzuki, K. and Sennett, C., 2007. A dual-stage method for lesion segmentation on digital mammograms. *Medical physics*, 34(11), pp.4180-4193.
- Zadeh, L.A., 1965. Fuzzy Sets. *Information and Control*, 8(5), pp. 338- 353.
- Zhang, Q., Eagleson, R. and Peters, T.M., 2006, April. Real-time visualization of 4D cardiac MR images using graphics processing units. In *Biomedical Imaging: Nano to Macro, 2006. 3rd IEEE International Symposium on* (pp. 343-346). IEEE.
- Zou, K.H., O'Malley, A.J. and Mauri, L., 2007. Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. *Circulation*, 115(5), pp.654-657.